AWARD NUMBER: W81XWH-15-1-0707

TITLE: The impact of ethnicity-dependent differences in breast epithelial hierarchy on tumor incidence and characteristics

PRINCIPAL INVESTIGATOR: Harikrishna Nakshatri

CONTRACTING ORGANIZATION: Indiana University School of Medicine

REPORT DATE: Dec 2019

TYPE OF REPORT: Final

PREPARED FOR: U.S. Army Medical Research and Development Command
Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for Public Release;
Distribution Unlimited

The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision unless so designated by other documentation.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. **PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.**

1. REPORT DATE Dec 2019	2. REPORT TYPE Final	3. DATES COVERED 09/30/2015 - 09/29/2019

4. TITLE AND SUBTITLE
The Impact of Ethnicity-Dependent Differences in Breast Epithelial Hierarchy on Tumor Incidence and Characteristics

5a. CONTRACT NUMBER

5b. GRANT NUMBER
W81XWH-15-1-0707

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Harikrishna Nakshatri, Brijesh Kumar and Mayuri Prasad

E-Mail:

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Indiana University School of Medicine, Indianapolis, IN 46202

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Development Command
Fort Detrick, Maryland 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT The incidence of the most aggressive triple negative breast cancer (TNBC) is significantly higher in African American (AA) than Caucasian (CA) women suggesting that the biology of the normal breasts between these two ethnic groups differ, which may contribute to altered susceptibility to tumor initiation, progression and/or metastasis. To address this hypothesis, we generated immortalized cell lines from healthy breast tissues of AA, CA and Hispanic women and characterized these cells lines for intrinsic subtypes by RNA-seq and PAM50. Based on characteristics of transformed counterparts of these cell lines, we propose that while cell type origin determines the subtype of tumor, genomic aberrations determine response to chemotherapy. We have identified an unique population of cells called PZP cells (PROCR+/ZEB1+/PDGFRalpha+) that surround the ducts and transdifferentiate into multiple cell types including fibroblasts, epithelial cells, adipocytes, and through crosstalk with epithelial cells, create a pro-inflammatory environment. PZP cells are elevated in the normal breasts of AA women and thus provide an aggressive environment for tumors to grow.

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			USAMRMC
Unclassified	Unclassified	Unclassified	Unclassified	131	19b. TELEPHONE NUMBER *(include area code)*

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39.18

TABLE OF CONTENTS

	Page
1. Introduction	4
2. Keywords	4
3. Accomplishments	4-36
4. Impact	36
5. Changes/Problems	36
6. Products	36-37
7. Participants & Other Collaborating Organizations	37
8. Special Reporting Requirements	37
9. Appendices	37

Introduction: As per American Cancer Society latest estimate (2006-2010), breast cancer mortality rate in women of African American ancestry (AA) is significantly higher compared with Caucasian (CA) women, despite lower incidence of breast cancer in AA. The incidence of young-onset and the most aggressive triple negative breast cancer (TNBC) is significantly higher in AA than CA suggesting that the biology of normal breast epithelial cells between these two ethnic groups differs, which may contribute to altered susceptibility to tumor initiation, progression and/or metastasis. Our goal is to determine the biologic basis of health disparity and to elucidate the role of AA-enriched normal breast cell population on breast cancer initiation, progression and metastasis.

Keywords: African American ancestry (AA), Caucasian ancestry (CA), SV40 T antigen, mutant Ras (H-Ras V12), stem/luminal progenitor/mature cells, PZP cells.

Accomplishments:

Specific Aim 1: To demonstrate that differences in normal breast epithelial hierarchy between African American and Caucasian women determine tumor behavior	**Timeline**	**Site 1**
Major Task 1 Generate transformed cells from breast epithelial cells of Caucasian (CA) and African American (AA) women	Months	
Subtask 1 Obtain IACUC and IRB approval	1-3 **Achieved**	Dr. Nakshatri
Subtask 2 Generate cells from breast biopsies of African American and Caucasian women	1-3 **Achieved**	Dr. Nakshatri
Subtask 3 Sort cells into different subpopulations using CD44/CD24, CD44/EpCAM and PROCR/EpCAM and immortalize the cells	3-6 **Achieved**	Dr. Nakshatri
Subtask 4 Perform transformation assays with SV40 T antigen/mutant p53 and activated H-Ras or FAB83B	6-12 **Achieved**	Dr. Nakshatri Dr. Kumar
Subtask 5 Subtype classification of transformed cells using assays such as PAM50	12-15 **Achieved**	Dr. Nakshatri
Milestone(s) Achieved: Established cell lines (at least 5 each) from breast epithelial cells of Caucasian and African American women		
Local IRB/IACUC Approval	3 **Achieved**	
Milestone Achieved: HRPO/ACURO Approval	6 **Achieved**	
Major Task 2 Characterize tumor cells for drug sensitivity, tumorigenecity, and metastasis		
Subtask 1 Determine the sensitivity of transformed cells to paclitaxel, docetaxel, doxorubicin, and cisplatin	15-17 **Achieved**	Dr. Nakshatri Dr. Kumar
Subtask 2 Conduct invasion and migration assay	17-20 **Achieved**	Dr. Nakshatri Dr. Prasad
Subtask 3 Analyze transformed cells for tumorigenecity in NSG mice- six tumor lines from CA and AA (3 each)- with and without estrogen pellets- 300 animals	15-22 **Achieved**	Dr. Nakshatri Dr. Kumar
Subtask 4 Characterize tumors and lungs by histopathology and/or PAM50	22-25 **Achieved**	Dr. Nakshatri/ IHC core
Subtask 5 Characterize transformed cells for tumor initiating capacity using limited dilution transplantation assay- three AA and three CA cell lines, three dilutions- 90 animals	22-26 **Achieved through flow cytometry characterization**	Dr. Nakshatri Dr. Kumar
Subtask 6 Characterize transformed cells for homing through intracardiac injection assay. Four cell lines per CA and AA- 80 animals	26-32 **Did not pursue as tumor characterization required additional animals and it is clear that cell of origin determines metastasis properties.**	Dr. Nakshatri Dr. Kumar
Subtask 7 Analyze brain, lungs, liver, and bones for metastasis	32-34 **Achieved**	Dr. Nakshatri/IHC core (Dr. Sandusky)
Subtask 8 Conduct statistical analysis of data from animal studies	22-36	Dr.

	Achieved	Nakshatri/Biostatistics core
Milestone(s) Achieved: Established whether transformed cells from AA and CA differ in their characteristics (tumorigenecity, invasion, metastasis etc).	36 **Achieved through characterization of PZP cells**	
Specific Aim 2: Establish that inter-individual heterogeneity within African Americans for unique cells determine susceptibility to breast cancer		
Major Task 3		
Subtask 1 Procure necessary tissues from IUSCC and Komen Normal Breast Tissue banks, prepare slides and standardize antibodies for IHC	1-8 **Created tissue microarray containing ~50 each of normal, tumor-adjacent normal and tumor of Caucasian and African American Women**	Dr. Nakshatri/IHC core (Dr. Sandusky)
Subtask 2 Perform immunohistochemical characterization of healthy normal, tumor-adjacent normal and tumor (50 samples per AA and CA) total 300	9-16 **Performed IHC for ZEB1, FOXA1, GATA3, and ER in the TMA** **Achieved**	Dr. Nakshatri/IHC core (Dr. Sandusky)
Subtask 3 Perform preliminary statistical analysis to determine the sample size required to find statistically significant differences between different ethnic groups	17-20 **Normal and tumor adjacent normal breast of African American women have statistically significantly higher ZEB1 but not ER compared to Caucasian women.**	Dr. Nakshatri/Biostatistics core (Ms. Heather Burney)
Subtask 4 Procure additional samples as required and reanalyze only those markers that are likely to be different in different ethnic groups based on preliminary analysis	21-30 **Achieved**	Dr. Nakshatri/ IHC core (Dr. Sandusky)
Subtask 5 Conduct statistical analysis of all samples	31-36 **Achieved**	Dr. Nakshatri/ Biostatistics core (Ms. Heather Burney)
Milestone(s) Achieved: Established whether breast of African American women with higher number of epithelial cells expressing stem cell-associated genes is more susceptible to breast cancer.	36 **Achieved**	Dr. Nakshatri

Specific Aim 1: To demonstrate that differences in normal breast epithelial hierarchy between African American and Caucasian women determine tumor behavior

Rationale: Despite large body of genome-wide studies in breast cancer, molecular pathways linked to ethnicity-dependent differences in breast cancer incidence and outcome are yet to be identified. One major limitation has been the non-availability of live "normal cells" from different ethnic groups, which we have overcome. We focused on epithelial plasticity as a starting point because recent studies have demonstrated validity of this concept *in vivo*. It was demonstrated that cellular plasticity contributes to regenerative capacity *in vivo* by allowing dedifferentiation of mature epithelial cells to stem cells [1]. Indeed, we recently showed elevated number of highly plastic cells in AA compared with CA women [2]. These inherent differences in plasticity should impact the behavior of tumor cells in CA and AA, despite the same oncogenic hit, which was examined in aim 1.

Isolation and immortalization of primary breast epithelial cells: Results described below have been published [3]. We have generated immortalized cell lines from six CA (KTB6, KTB34, KTB36, KTB37), nine AA (KTB8, KTB32, KTB39, KTB40, KTB41, KTB42, KTB51, KTB53, KTB55, KTB59) and three Hispanic women (KTB21, KTB22, KTB26) by overexpressing human telomerase gene (hTERT) in primary cells isolated and propagated from core breast biopsies of healthy women **(Fig. 1A)**. To ensure that cells propagated using reprogramming assay retain "stem cell like activity", we grew cells on matrigel. These immortalized cell lines formed acini of varying sizes **(Fig. 1B)**, which indicate that normal cell lines contain a subpopulation of cells with stem cell characteristic.

Ancestry mapping to determine ethnic origin of cell lines: Self-reported ethnicity does not always match with genome-driven ethnicity as determined by ancestry mapping. To ensure that KTB cell lines are representative of women of different genetic ancestry, all samples were subjected to highly discriminative ancestry informative 41-SNP (single nucleotide polymorphism) genomic analyses [4]. As shown in **Fig. 2**, self-reported white women were enriched for European ancestry markers (KTB6, KTB34, KTB36 and KTB37). The self-reported Hispanic women (KTB21, KTB22 and KTB26) displayed highly heterogeneous ancestry marker distribution. Eight self-reported African American women (KTB39, KTB8, KTB41, KTB40, KTB42, KTB53, KTB57, KTB59) had inherited >50% of African ancestry markers.

Karyotyping of primary breast epithelial cells: To ensure diploid nature of immortalized cell lines, karyotype analyses were performed. Chromosome spread were prepared from mitotic cells after arresting at metaphase of cell cycle by colcemid treatment, when chromosomes assume their most condensed conformations, which confirmed 46 chromosomes in all samples **(Fig. 3)**. Several cell lines have undergone >20 passages and remained diploid.

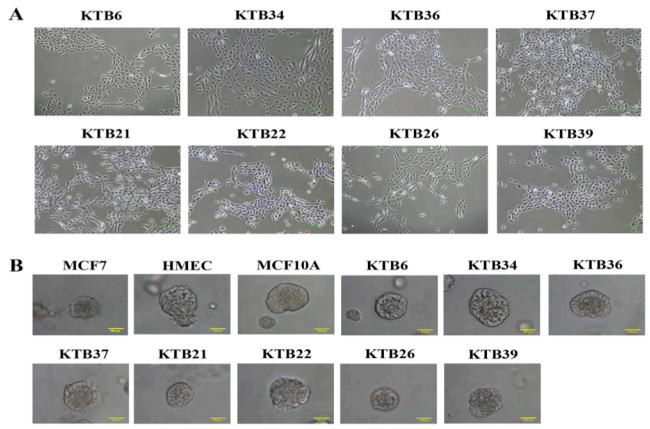


Fig. 1: (A) Phase contrast images representing the epithelial morphology of human telomerase gene (hTERT) immortalized breast epithelial cells from Caucasian (KTB6, KTB34, KTB36, KTB37), Hispanic (KTB21, KTB22, KTB26) and African American (KTB39) women. HME and MCF10A are used as control. (A) Primary cell lines from AA (KTB39), CA (KTB6, KTB34, KTB36 and KTB37) and Hispanic (KTB21, KTB22 and KTB26) women. (B) Immortalized cell lines generate acini on matrigel.

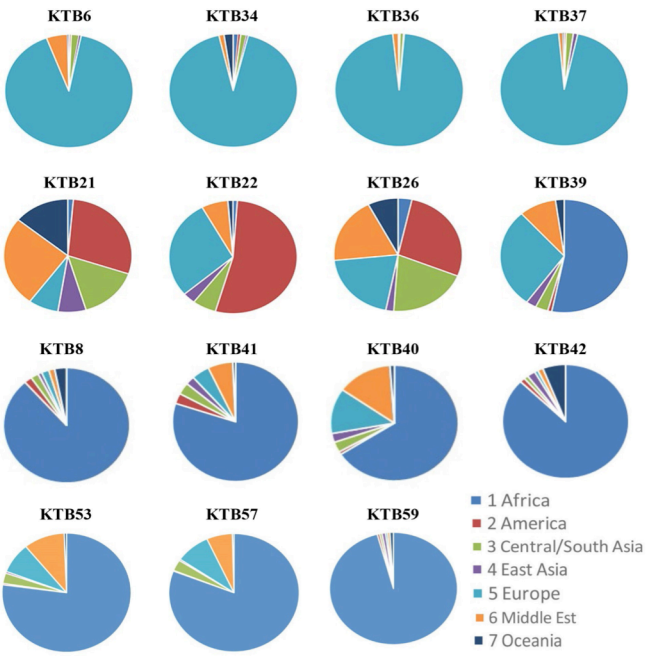


Fig. 2: Genetic ancestry mapping of Caucasian (KTB6, KTB34, KTB36, KTB37), Hispanic (KTB21, KTB22, KTB26) and African American (KTB39, KTB8, KTB41, KTB40, KTB42, KTB53, KTB57, KTB59) women based on the genotype analysis using a panel of 41-SNP along with a Bayesian clustering method.

Molecular subtype classification of KTB cell lines: To determine whether gene expression patterns in immortalized KTB cell lines differ from the commonly used normal cell lines MCF10A and HMECs, we performed RNA-seq analyses. A cell line derived from a BRCA2 mutant carrier 1505-10B was also included in the analyses. Principle component analyses (PCA) revealed distinct differences in gene expression pattern between KTB cell lines compared to HMEC, MCF10A, and the luminal breast tumor cell line MCF7 **(Fig. 4A)**. Independent PCA analyses of KTB cell lines showed significant differences between cell lines, further confirming inter-individual variations in characteristics of "normal" breast cell lines **(Fig. 4B)**.

Unsupervised clustering analyses separated KTB cell lines from HMEC and MCF10A with the exception of KTB40 and KTB42, which clustered with HMEC and MCF10A **(Fig. 2C)**. Previous studies have shown enrichment of basal cell gene expression pattern in HMEC and MCF10A but the KTB cell lines displayed a distinct gene expression pattern compared to HMEC and MCF10A cell lines (4). We performed ingenuity pathway analyses of 733 genes expressed at higher level in KTB luminal cell lines compared to HMEC, MCF10A, KTB40, and KTB42. Canonical pathways typically enriched in luminal cells including ErbB and PAK signaling were

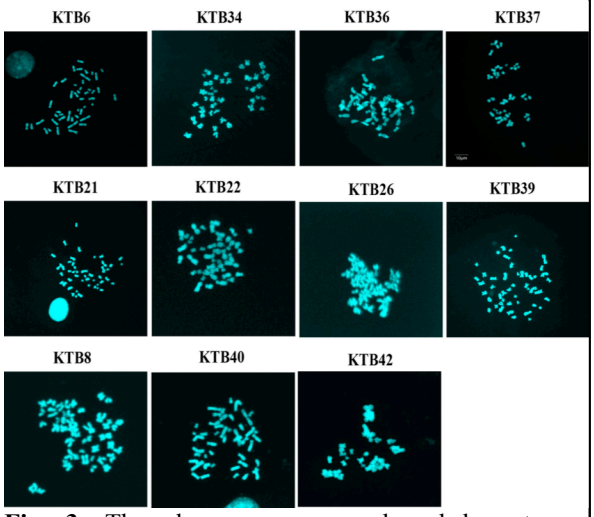


Fig. 3: The chromosome spread and karyotype analysis of immortalized cell lines to confirm human origin and diploid nature of cell lins.

elevated in KTB cell lines [3]. Encouraged by these results, we next determined whether KTB cell lines represent "normal" counterparts of intrinsic subtypes of breast cancer. Intrinsic subtypes have been predicted to originate from normal counterparts but it has never been demonstrated experimentally [5]. Using PAM50 model in genefu, we calculated the subtype probability score for each cell line. KTB34 represented luminal A, KTB21, KTB22, KTB36 represented basal and KTB26, KTB37, KTB39 and 1505-10B (BRCA2 mutant) represented normal-like subtypes **(Fig. 4D and 4E)**. These results document the presence of immortalized cells with gene expression pattern similar to intrinsic subtypes of breast cancer.

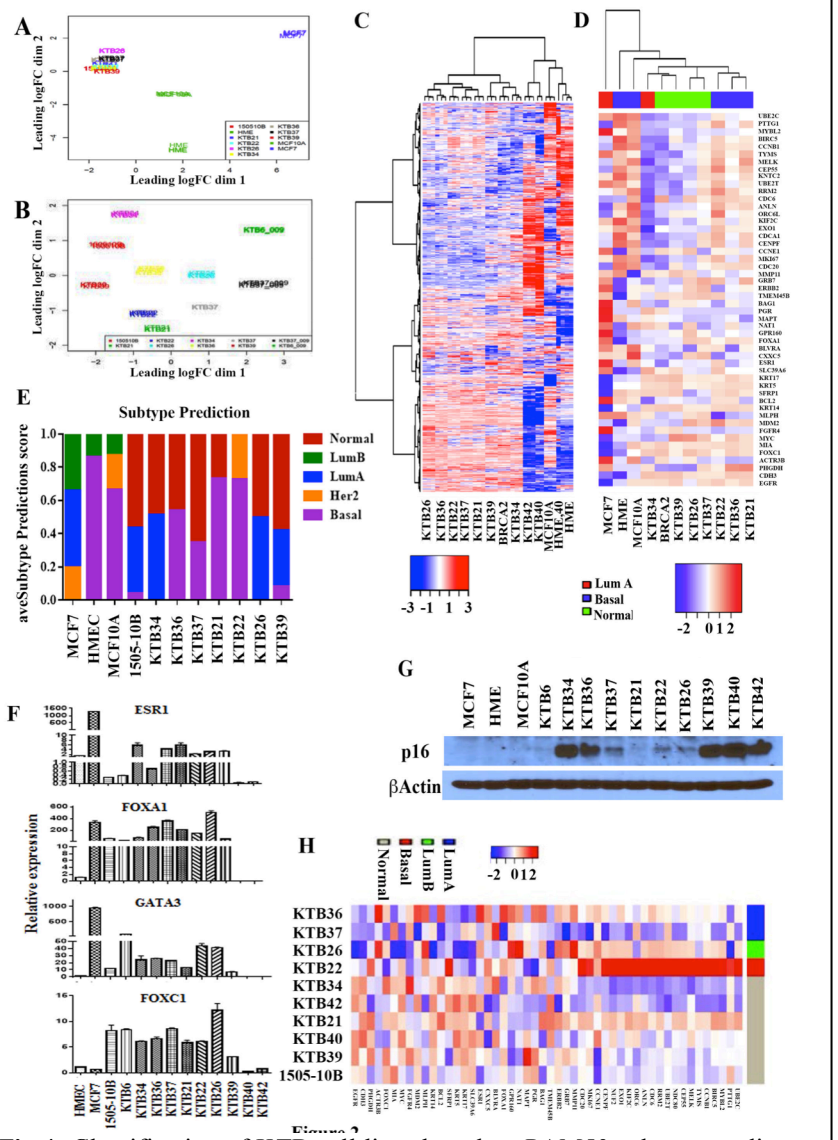


Fig.4: Classification of KTB cell lines based on PAM50 subtype predictor. A) PCA analyses of RNA-seq data of all immortalized cell lines including HME, MCF10A and MCF7. 1505-10B is a BRCA2 mutant cell line. B) PCA analyses of RNA-seq data of KTB cell lines show heterogeneity in gene expression. C) Unsupervised clustering of RNA-seq data shows distinct gene expression pattern in KTB cell lines except KTB40 and KTB42 compared to MCF10A and HME. KTB40 and KTB42 clustered with HME and MCF10A. D) Classification of KTB cell lines based on PAM50 subtype classifier. E) Subtype prediction scores of individual cell lines. F) Expression levels of luminal enriched *ESR1, FOXA1, GATA3*, and basal-enriched *FOXC1* in different KTB cell lines. G) p16(INK4) expression in different cell lines. H) Classification of primary cells prior to immortalization based on PAM50 subtype classifier.

To further characterize KTB cell lines for luminal and basal-enriched gene expression, we measured the expression levels of *ESR1*, *FOXA1*, *GATA3*, and *FOXC1*. While *ESR1*, *FOXA1*, and *GATA3* are expressed predominantly in luminal cells and form a hormone-responsive lineage-specific transcription factor network, *FOXC1* expression identifies cells with basal cell-enriched gene expression [6, 7]. Indeed, all four of these transcription factors were expressed at variable levels in KTB cell lines and to our knowledge, this is the first report to document immortalized breast epithelial cell lines expressing *ESR1*, *GATA3*, and *FOXA1* **(Fig. 4F)**. We also noted another difference in KTB cell lines and general observation about immortalization of breast epithelial cells. Immortalization is often associated with the loss of cell cycle inhibitor and senescence promoter p16(INK4) [8]. However, RNA-seq results showed significant expression of p16(INK4) transcripts in KTB cell lines. Western blotting also confirmed expression of p16(INK4) in several of KTB cell lines **(Fig. 4G)**. Thus, immortalized KTB cell lines are unique in their characteristics compared to available cell lines including the expression of luminal-enriched transcription factors and p16(INK4).

To rule out the possibility that intrinsic subtype phenotype is acquired during immortalization, we performed RNA-seq analyses of primary cells prior to immortalization. PAM50 classification identified primary cells with luminal A, luminal B, basal and normal-like features **(Fig. 4H)**. It does appear that immortalization selects for specific subtypes as intrinsic subtype classification prior and after immortalization did not overlap. In particular, KTB40 and KTB42 prior to immortalization clustered as "normal-like" subtype. However, we do note a caveat in our analyses, as RNA-seq of primary and immortalized cells was done at different times due to volume of samples and these comparisons have limitations. Nonetheless, our results clearly show that cells derived from the normal breast prior and after immortalization have gene expression patterns overlapping the gene expression patterns in intrinsic subtypes of breast cancer.

Keratin expression pattern identifies distinct subpopulation of cells within immortalized KTB cell lines: Keratin 14 (KRT14) is typically expressed at higher levels in basal cells, whereas keratin 19 (KRT19) is expressed in luminal cells. Cells that express both KRT14 and KRT19 (KRT14+/KRT19+) are considered to have luminal progenitor properties [9]. We used immunofluorescence to determine expression pattern of these two keratins. Consistent with the results of PAM50 classification, the luminal A cell line KTB34 was strongly positive for KRT19, similar to MCF7, although low level KRT14 expression was seen in this cell line, whereas the basal cell line KTB22 was strongly positive for KRT14 **(Fig. 5A)**. The normal-like cell line KTB39 was KRT14+/KRT19+.

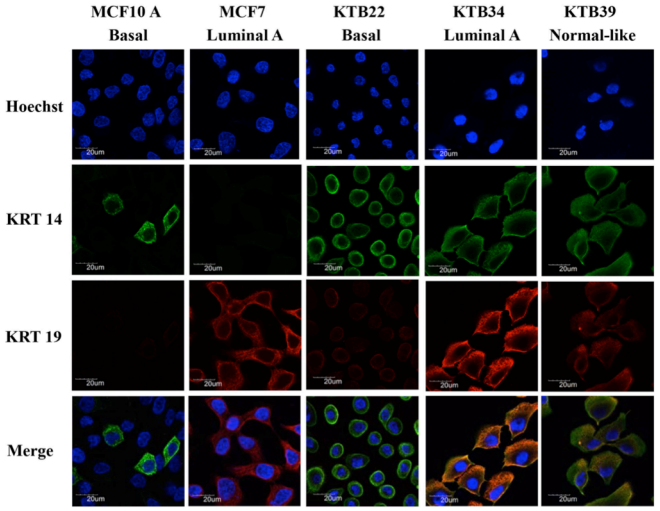


Fig. 5: Luminal and basal cell-enriched features of KTB cell lines. Representative data are presented. KRT14 and KRT19 staining patterns of different cell lines. MCF10A and MCF7 were used as controls.

Phenotypic heterogeneity in immortalized and non-immortalized cell lines: In order to confirm that the immortalized cell lines show the inter-individual variability as previously shown for the primary breast epithelial cells [2], we subjected KTB cell lines prior to and after immortalization to phenotypic analyses using additional markers that define stem/basal, luminal-progenitor, luminal-differentiated, luminal, and basal cells [10]. These analyses enabled us to document enormous phenotypic heterogeneity in

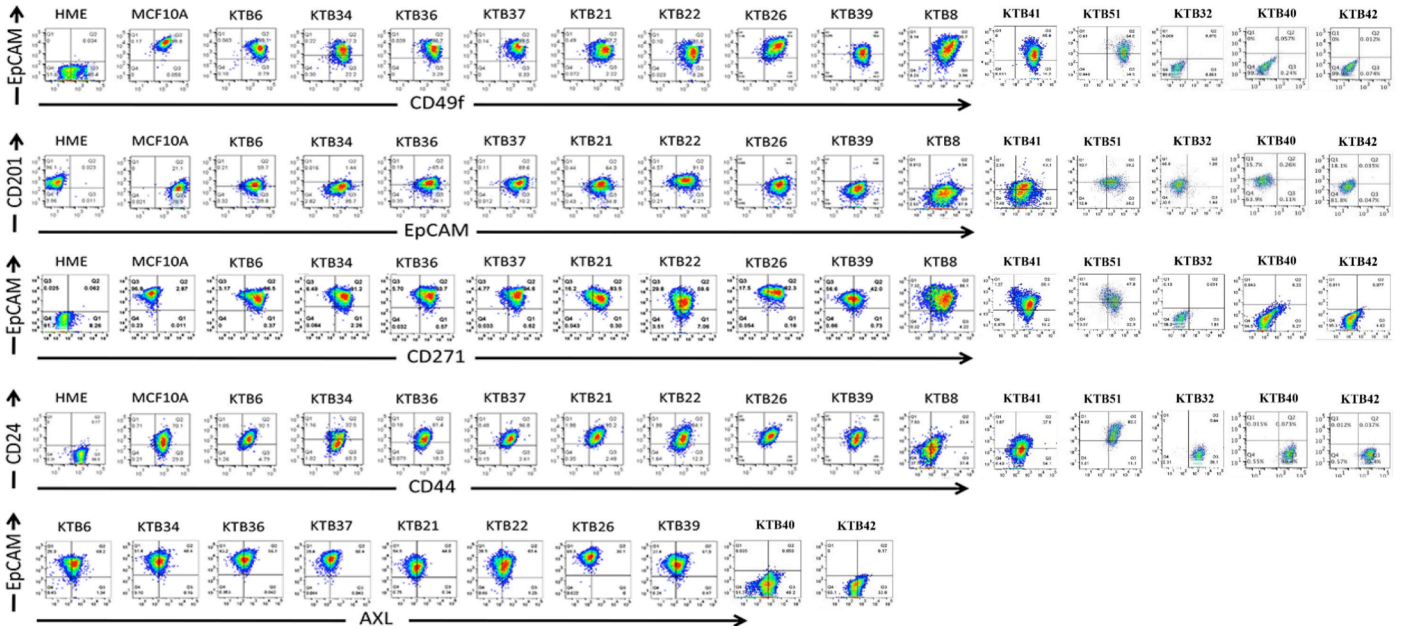


Fig. 6: Stem, progenitor and differentiated cell state of immortalized cell lines. Cells were stained with indicated antibodies and flow cytometry was used to identify subpopulation of cells. HME and MCF10A were used as control. KTB6, 34, 36, 37 (CA), 21, 22, 26 (HA), 39, 8, 41, 51, 32, 40, 42 (AA).

breast immortalized cell lines [KTB6, KTB34, KTB36, KTB37 from CA; KTB21, KTB22, KTB26 from Hispanic; and KTB8, KTB39, KTB41, KTB51, KTB32, KTB40, KTB42, KTB53, KTB55, KTB59 from AA] derived from healthy individuals.

CD49f and EpCAM staining pattern: CD49f^{high}/EpCAM^{low}, CD49f^{hig}h/EpCAM^{medium}, and CD49f^{low}/EpCAM^{high} cells are described as breast stem, luminal progenitor, and mature/differentiated cells, respectively [10]. Cells from each ethnic group contained a very low number of stem cells. Notably, AA-derived cell lines showed comparatively high stem cell populations. Inter-individual variability was noted in progenitor and mature cells **(Fig. 6)**.

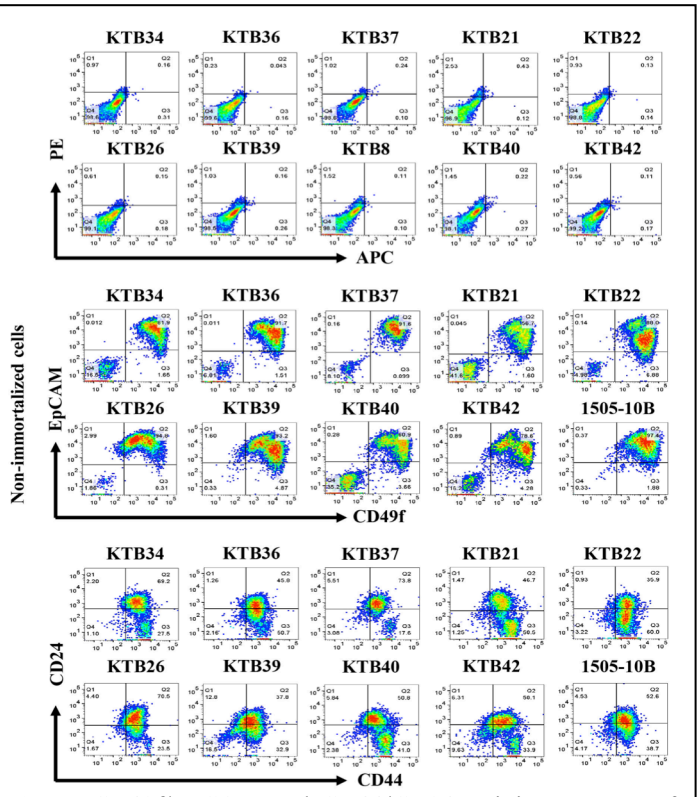


Fig. 7: CD49f/EpCAM and CD44/CD24 staining patterns of primary KTB cells. Flow cytometry patterns of various primary cell lines stained with CD49f/EpCAM and CD44/CD24, isotype controls are also shown.

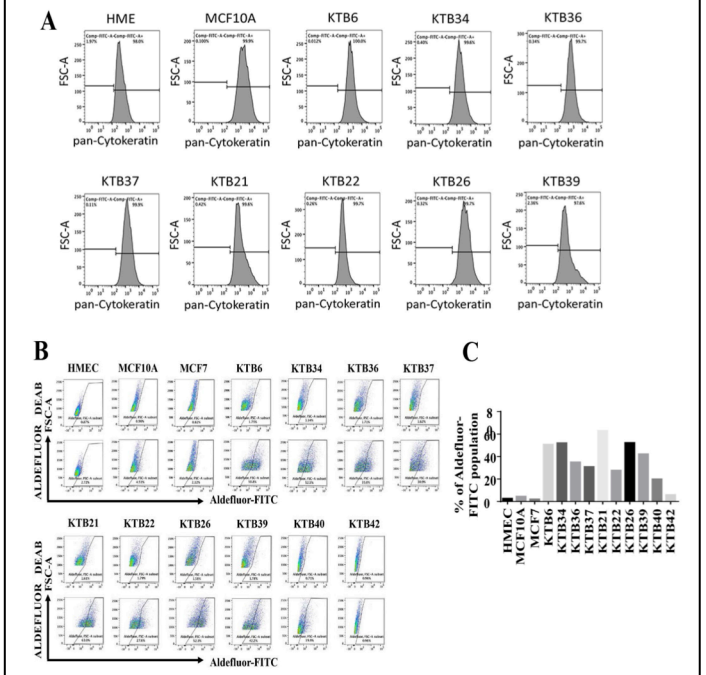


Fig. 8: (A) Pan-cytokeratin antibody was used as a marker to characterize the epithelial nature of immortalized cell lines. (B &C) ALDEFLUOR positive cells from normal breast epithelium have stem cell properties. Primary breast epithelial cells of Caucasian, Hispanic & African American women show heterogeneity in stem/progenitor cell state. HME, MCF10A & MCF7 are used as control.

CD201 and EpCAM staining pattern: CD201+ (PROCR) cells are considered to represent multi-potent stem cells in the mouse mammary gland [11]. During immortalization, we consistently noted that CD201+/EpCAM- cells of AA women immortalized quickly and dominated the culture. Using differential trypsinization and flow cytometry, we created immortalized cell lines lacking CD201+/EpCAM- subpopulation of cells from AA women (KTB8 and 39). Four CD201+/EpCAM- cell lines have also been created. We have renamed these cells as PZP cells (PROCR+/ZEB1+/PDGFRα+) because of expression of the additional markers. Among the cell lines described above, we observed remarkable variability in CD201+/EpCAM+ and CD201-/EpCAM+ subpopulation levels **(Fig. 6)**.

CD271 and EpCAM staining pattern: $CD271^+$ tumor cells are the basal cells with CSC activity and represent a minor portion of cells in luminal tumors [12]. The number of $CD271^+/EpCAM^-$, $CD271^+/EpCAM^+$, and $CD271^-/EpCAM^+$ cells varied significantly between individuals **(Fig 6)**. For example, KTB39 had highest number of $CD271^-/EpCAM^+$ cells, possibly representing differentiated luminal cells.

CD44/CD24 staining pattern: CD44 and CD24 are the "original" markers used to characterize cancer stem cells (CSCs) in breast cancer [13]. However, we noted remarkable inter-individual variability in the ratio between $CD44^+/CD24^-$ and $CD44^+/CD24^+$ normal cells, possibly reflecting inherent inter-individual differences in the rate of differentiation **(Fig 6)**. KTB34 from CA, KTB39 and KTB8 from AA contained a higher subpopulation of $CD44^+/CD24^-$ cells.

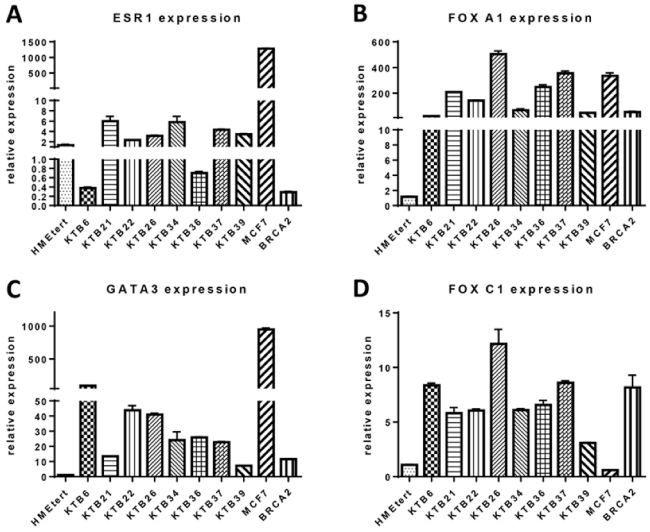


Fig. 9: Expression analysis of ESR1, FOXA1, GATA3 and FOXC1 genes in immortalized cell lines using qRT-PCR. HME and MCF7 were used as control.

AXL and EpCAM staining pattern: AXL is a receptor tyrosine kinase (RTK) that promotes survival, stimulates cell proliferation and/or inhibits apoptosis in response to targeted therapies [14]. It also drives epithelial to mesenchymal transition (EMT) and confers resistance to PI3K inhibitors [15]. To determine whether cells from AA, CA and Hispanic differ in AXL-positivity, we stained cells for AXL and EpCAM. We observed a remarkable inter-

individual variability in the ratio between AXL/EpCAM normal cells **(Fig 6)**. These results raise the possibility of inherent differences in AXL+ positive tumor cells with differing EMT potential and response to therapy.

Primary cells prior to immortalization showed inter-individual variation in the levels of all three populations of cells **(Fig 7)**. CD44/CD24 staining pattern also confirmed inter-individual heterogeneity. After immortalization, this variation decreased considerably with the majority of cell lines displaying luminal progenitor phenotype (CD49f+/EpCAM+) **(**Compare staining patterns between **Fig 6 and Fig 7)**.

Pan-cytokeratin and ALDEFLUOR+ staining pattern: To further confirm the epithelial nature of immortalized cell lines, cells were stained with pan-cytokeratin antibody. Only epithelial cells stained positive with pan-cytokeratin confirming the epithelial nature of the KTB6, KTB34, KTB36, KTB37, KTB21, KTB22, KTB26 and KTB39 **(Fig. 8A)**. Aldehyde dehydrogenase (ALDH) expressing cells are considered as breast CSCs enriched in luminal/HER2+ breast cancers [16, 17]. In breast cancer, CD44+/CD24- cells correspond to highly invasive and slowly proliferating mesenchymal cells, whereas ALDH+ cells are rapidly proliferating epithelial cells [18]. In the normal breast, ALDH expression is elevated in luminal progenitor (ALDH1a3) and luminal-restricted cells (ALDH5a1) [19]. ERα-negative/ALDH1+ stem/progenitor cells are believed to generate ERα-positive cells of luminal lineage [20]. There was considerable inter-individual variation in the number of ALDH+ cells (as measured by ALDEFLUOR assay) suggesting that luminal to non-luminal cell ratio vary between individuals, which may ultimately determine the number of ERα-positive cells in the breast **(Fig. 8B, C)**.

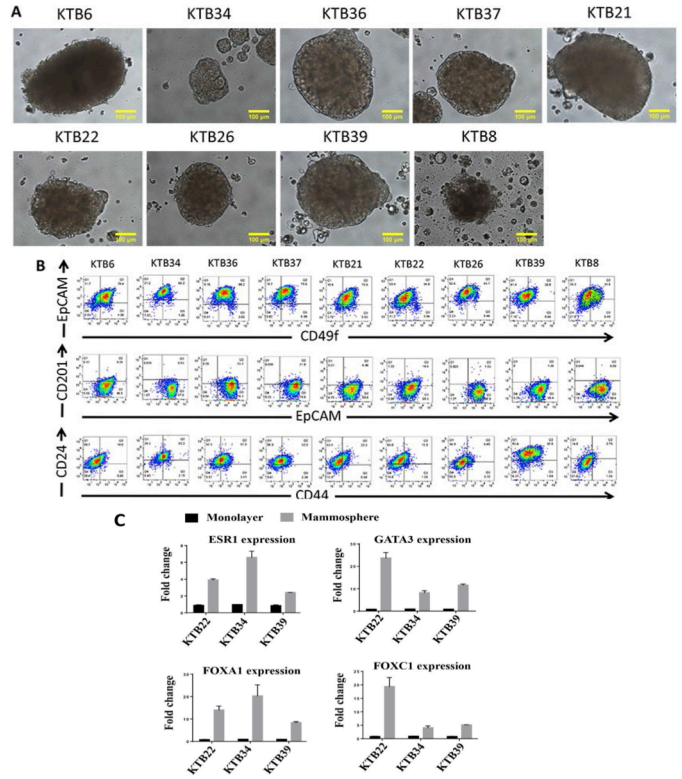


Fig. 10: (A) KTB cell lines form mammospheres of variable sizes. (B) Cells in mammospheres show different levels of differentiation based on CD44/CD24, CD49f/EpCAM and CD201/EpCAM staining pattern. (C) Cells in mammospheres show elevated levels of stem/basal (*FOXC1*) and luminal differentiation markers (*ESR1, FOXA1*, and *GATA3*).

Expression of *ESR1, FOXA1, GATA3* and *FOXC1* gene in immortalized cell lines: Breast epithelial cells that express ESR1, FOXA1 and GATA3 are considered to be hormonally responsive luminal cells, whereas FOXC1 is expressed predominantly in basal cells [6, 7]. RNA-seq showed differential expression of these genes in our cell lines. We confirmed differential expression of *ESR1, FOXA1, GATA3* and *FOXC1* by qRT-PCRs. Expression of ESR1 was high in KTB21, KTB22, KTB26, KTB34, KTB37 and KTB39 as compared to KTB6, KTB36 and 1505-10B (BRCA2) **(Fig. 9A)**. Expression of FOXA1 was variable in cell lines with comparatively higher in KTB26 and KTB37 **(Fig. 9B)**. GATA3 expression was higher in KTB6, moderate in KTB22, KTB26, KTB34, KTB36, KTB37, and lower in KTB21, KTB39, and 1505-10B cells **(Fig. 9C)**. FOXC1 expression was higher in KTB6, KTB26, KTB37, 1505-10B, moderate in KTB21, KTB22, KTB34, KTB36, and lower in KTB39 cells **(Fig. 9D)**. These results indicate that our cell lines contain heterogeneous population of cells with luminal, luminal progenitor and basal cell type-characteristics. We also note that our cell lines differ from the most commonly used control "normal" cell lines such as MCF10A and HMECs, which are predominantly basal cell lines.

Immortalized cell lines show inter-individual variability in stemness: To determine whether KTB cell lines show inter-individual variability in stemness, we performed mammosphere assay in all cell lines. KTB cell lines showed variable size of spheres indicating differences in stemness property of each cell line **(Fig. 10A)**. Mammospheres were subjected to flow cytometry for CD49f/EpCAM, CD201/EpCAM and CD44/CD24 expression to distinguish stem, luminal progenitor, basal and mature/differentiated cells. This assay suggested inter-individual differences in differentiation capacity **(Fig. 10B)**. For example, CD44 and CD24 staining of mammospheres showed inter-individual variability in the ratio between CD44+/CD24−, CD44−/CD24+ and CD44+/CD24+ cells. Notably, more CD44-/CD24+ cells, which were never observed under 2D culture of primary cells before immortalization or after immortalization **(Fig. 6)**, were found in mammosphere cultures of most of the cell lines **(Fig. 10B)**. Similar increase in CD49f-/EpCAM+ cells, representing mature/differentiated cells, was noted in mammosphere cultures of few cell lines **(Fig. 10B)**. We further confirmed inter-individual differences in differentiation under mammosphere condition by comparing the expression levels of *ESR1, FOXA1, GATA3,* and *FOXC1* under 2D and mammosphere conditions. The expression levels of *ESR1, FOXA1*, and *GATA3* increased substantially in cells from mammosphere cultures compared to 2D cultures **(Fig. 10C)**. Thus, stemness and differentiation capabilities show inter-individual variability and cell lines/assay systems described here are useful for assessment at individual level.

PROCR (CD201)+/EpCAM- cells enriched in African American women correspond to naturally occurring ZEB-1 positive cells with epithelial to mesenchymal transition (EMT) phenotype: KTB40 and KTB42 cell lines distinguished themselves from the

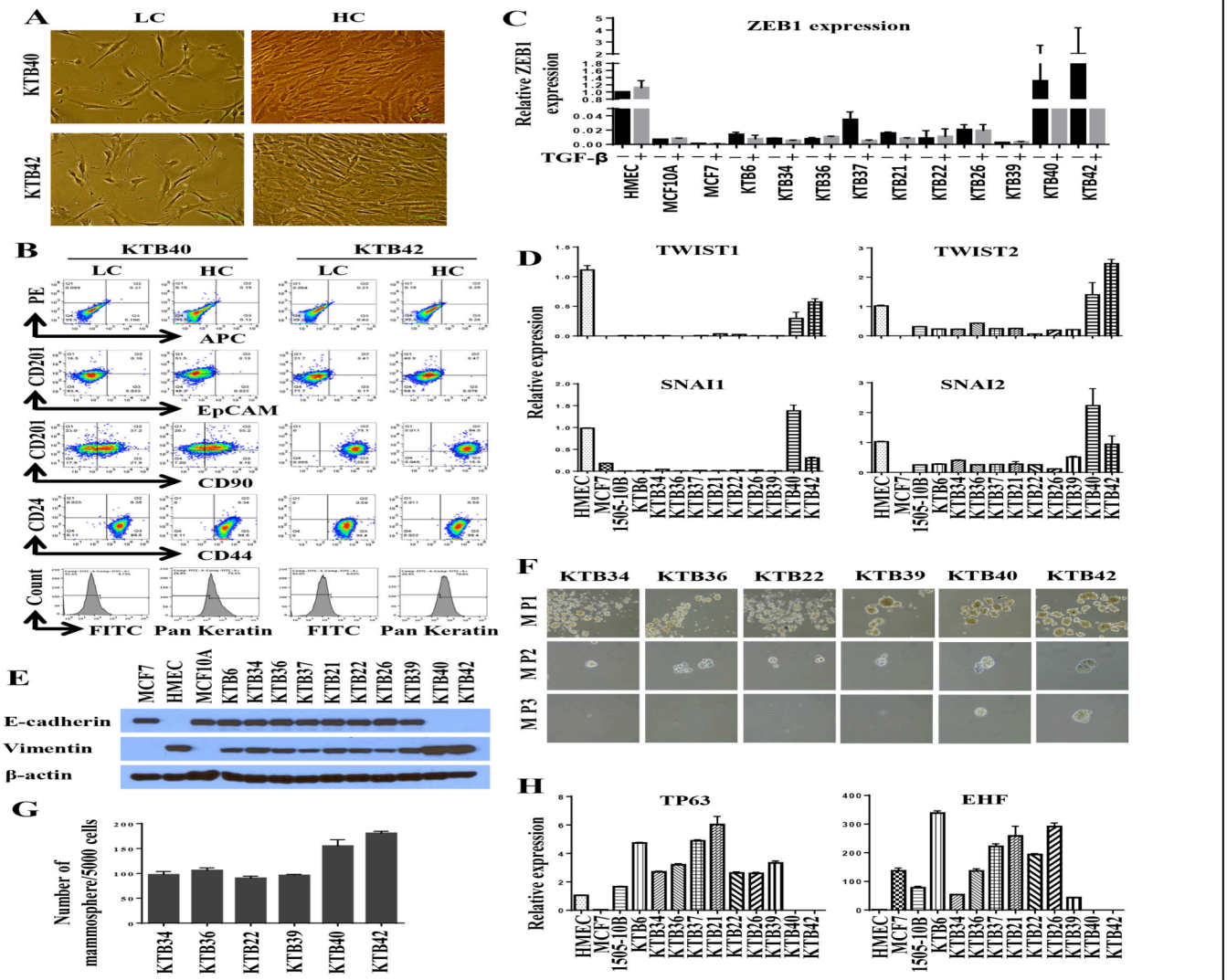


Fig. 11: CD201^{+}/EpCAM^{-} cells express pan-keratin but display features of EMT. (**A)** Confluence-dependent changes in the morphology of CD201^{+}/EpCAM^{-} cells. (**B)** CD24, CD44, CD90, CD201, EpCAM, and pan-keratin staining pattern of cell lines. (**C)** *ZEB1* expression levels in various KTB cell lines with and without TGFβ1 treatment. *ZEB1* expression in KTB40 and KTB42 was significantly higher than in other KTB cell lines ($P <$ 0.0001). (**D)** *TWIST1, TWIST2, SNAI1,* and *SNAI2* expression levels in various KTB cell lines. A *TWIST1, TWIST2, SNAI1,* and *SNAI2* expression in KTB40 and KTB42 was significantly higher than in other KTB cell lines. (**E)** E-Cadherin and vimentin expression levels in various KTB cell lines. (**F)** Self-renewal capacity of various cell lines as measured by mammosphere assay. Primary mammospheres of KTB40 and KTB42 but not other KTB cell lines formed tertiary mammospheres. MP1, mammosphere passage 1; MP2 mammosphere passage 2; MP3, mammosphere passage 3. (**G)** Mammosphere-forming efficiency of various KTB cell lines. KTB40 and KTB42 cell lines were more efficient than other cell lines in generating mammospheres. (**H)** *TP63* and *EHF* expression levels in various KTB cell lines.

rest of the cell lines with respect to gene expression patterns and cell surface marker profiles. These cell lines were derived from African American women. During our efforts to immortalize breast epithelial cells from African American women, we consistently observed rapid immortalization of cells with EMT features, which overtook colonies with epithelial phenotype. We placed additional efforts to characterize these cell lines. These cells at low density displayed epithelial morphology but acquired fibroblastic phenotype with increasing density **(Fig. 11A)**. These cell lines contained heterogeneous population of cells with CD201+/EpCAM-, CD201-/EpCAM-, CD201+/CD90-, CD201+/CD90+, CD201-/CD90+, and CD44+/CD24- cell surface marker profiles **(Fig. 11B)**. The majority of cells stained positive for pan-keratin antibody **(Fig. 11B)**. We included CD90 in this analysis because CD90 is expressed in MaSC, fibroblasts, adipocytes, and mesenchymal stem cells [21]. Unlike other KTB cell lines described above, these cells expressed EMT and stemness-associated transcription factor *ZEB1, TWIST1, TWIST2, SNAI1* and *SNAI2* **(Fig. 11C and 11D)** [22, 23]. KTB40 and KTB42 did not demonstrate E-Cadherin positivity in western blotting while the remaining cell lines expressed significant levels of E-cadherin **(Fig. 11E)**. Vimentin expression was significantly higher in KTB40 and KTB42 cell lines compared to other KTB cell lines.

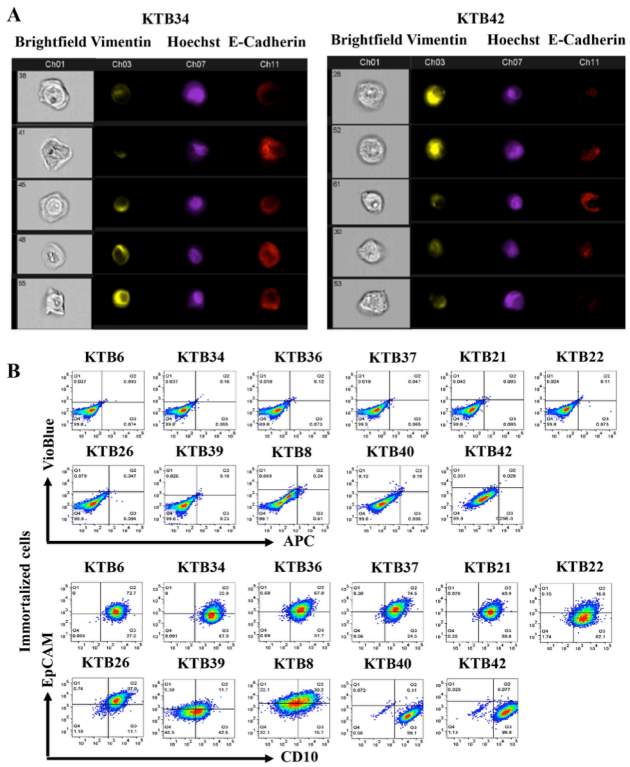


Fig. 12: (A) Vimentin, E-cadherin and Hoechst staining patterns of KTB34 and KTB42 using flow cytometry imaging. (B) CD10/EpCAM flow cytometry staining in immortalized KTB cell lines to separate the luminal from myoepithelial cell types and isotype controls are also shown.

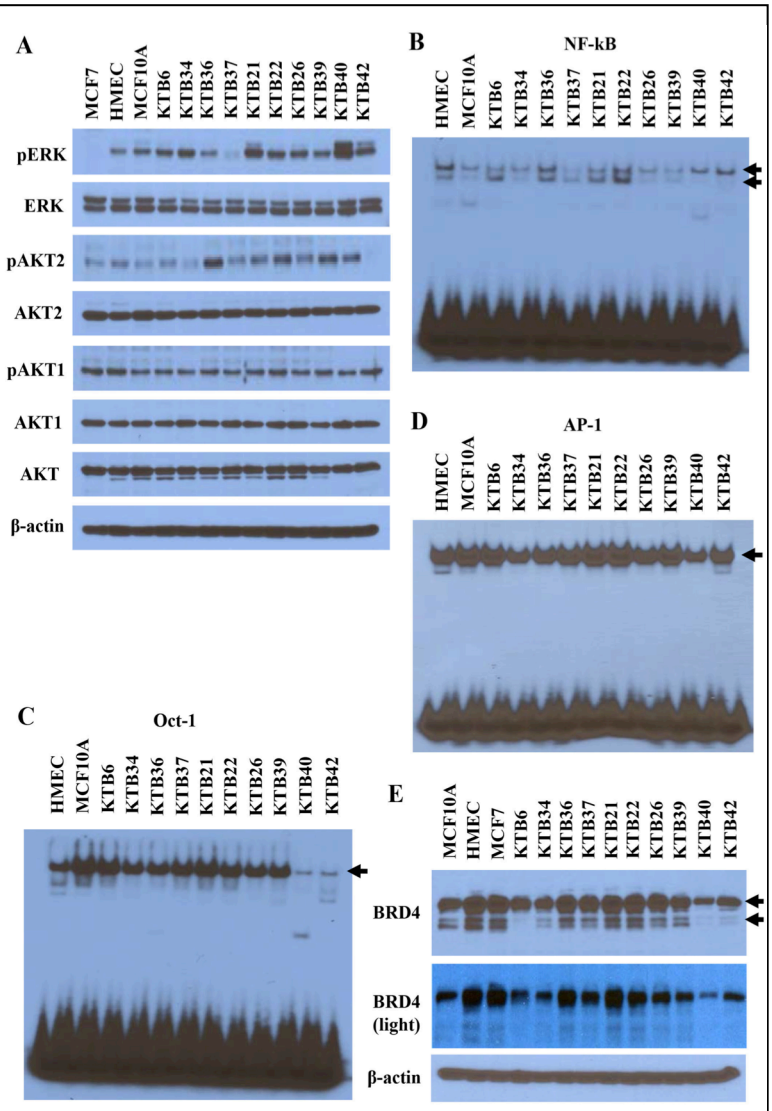


Fig. 13: KTB cell lines show differences in basal activity of several signaling molecules. (A) pERK and pAKT2 but not pAKT1 show cell type specific differences. (B) NF-kB DNA binding activity in various cell lines. (C) OCT-1 DNA binding activity in various cell lines. (D) AP-1 DNA binding activity. (E) BRD4, involved in epigenetic gene regulation and a drug target, shows enormous variability in expression. Both long and short isoforms are shown.

Lack of E-cadherin but elevated vimentin provided further evidence for EMT features of KTB40 and KTB42 cell lines. Since EMT features partially overlap with stemness [24], serial dilution mammosphere assay was performed to determine whether KTB40 and KTB42 differ in stemness relative to other KTB cell lines. KTB40 and KTB42 displayed higher mammosphere forming ability than other KTB cell lines and only these two cell lines were able to form tertiary mammospheres **(Fig. 11F and 11G)**.

It was puzzling to detect low levels of vimentin expression in our E-cadherin positive KTB cell lines. This could be due to the presence of distinct subpopulation of cells expressing either E-cadherin or vimentin. Alternatively, there could be a population of cells expressing both proteins. Recent studies support the concept of epithelial/mesenchymal or luminal/basal hybrid cells playing a role in organogenesis and tumor progression [25, 26]. To distinguish between these possibilities, we measured the expression levels of luminal progenitor cell enriched EHF and basal cell enriched TP63 transcription factors [27]. While KTB40 and KTB42 did not express these two genes, the remaining KTB lines expressed variable levels of both *EHF* and *TP63* **(Fig. 11H)**. The most commonly used HMEC cells expressed only TP63. To confirm the presence of hybrid cells, we double stained cells for E-cadherin and vimentin and single cells were analyzed by imaging cytometry. A significant number of cells in KTB34 cell line were double positive. Interestingly, although western blotting did not detect E-Cadherin expression, a small fraction of KTB42 expressed E-cadherin **(Fig. 12A)**.

Since our cell lines were derived from core biopsies of healthy women, there is possibility that a fraction of cells in our cell lines display myoepithelial features. CD10+/EpCAM- cells are considered myoepithelial cells although a recent study showed CD10 positivity of cancer associated fibroblasts [28, 29]. While KTB40 and KTB42 were CD10+/EpCAM-, the remaining cell lines showed variable levels of CD10 positivity without any distinct subpopulation with CD10+/EpCAM- phenotype **(Fig. 12B)**.

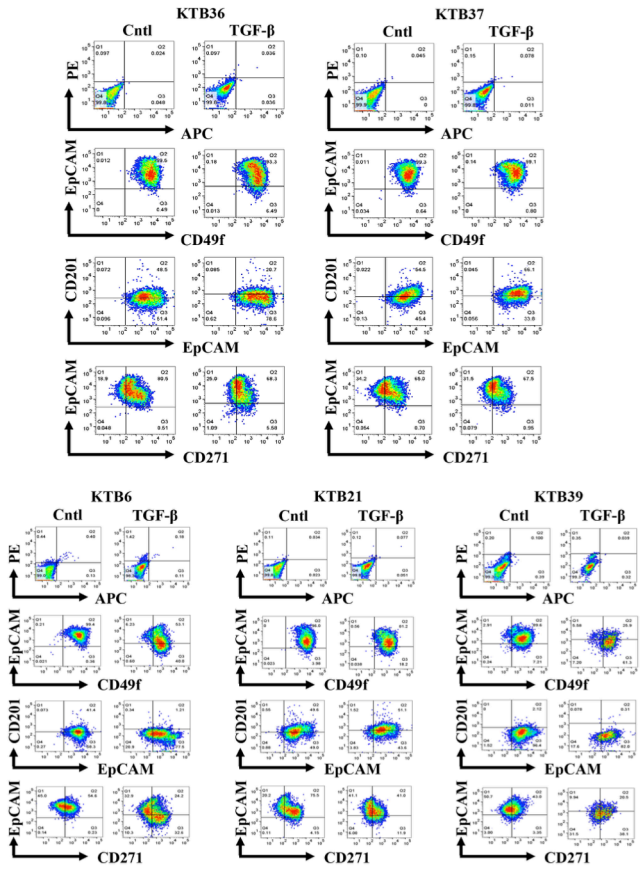


Fig. 14: The effect of TGFβ1 on stem/differentiated cell associated cell surface marker expression. The effect of TGFβ1 is cell type specific. The intensity of CD49f, CD44 and CD24 staining differed between untreated and TGFβ1 treated cells.

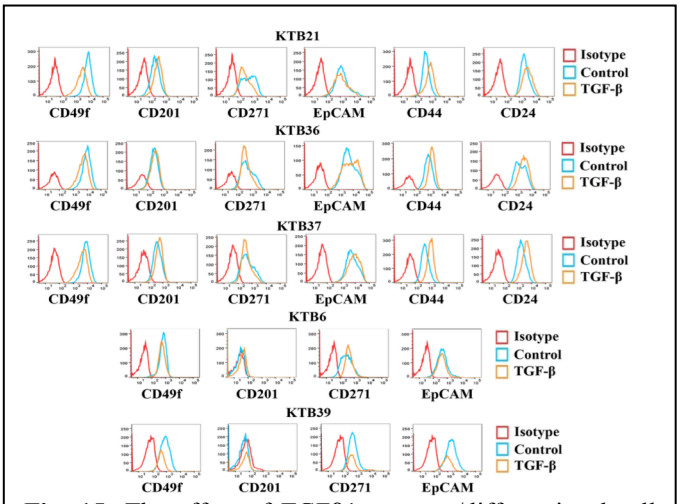


Fig. 15: The effect of TGFβ1 on stem/differentiated cell associated cell surface marker expression. The intensity of CD49f, CD44 and CD24 staining differed between untreated and TGFβ1 treated cells.

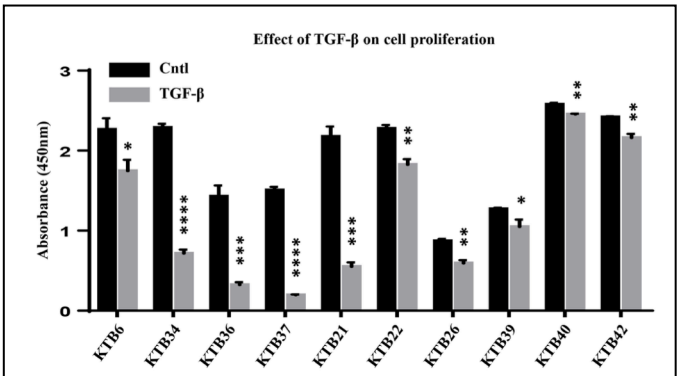


Fig. 16: The effect of TGFβ1 on proliferation of KTB cell lines. KTB cell lines show variable sensitivity to TGFβ1 in cell proliferation assay. The effects indicated by stars are statistically significant but the extent of effects varied between cell lines (*p<0.05, **p<0.004, ***p<0.0007, ****p<0.0001.

Inter-individual differences in basal activity of select signaling networks: Since both transcriptome and phenotypic analyses documented inter-individual heterogeneity in cell lines, we next examined whether basal signaling events were also different in these cell lines. We measured phospho-ERK, phospho-AKT1, phospho-AKT2, OCT-1, AP-1 and NF-κB DNA binding activity because intrinsic subtypes are known to differ in activity of signaling molecules [30]. Phospho-ERK levels varied between cell lines with KTB40 and KTB42 containing highest levels of both phospho-ERK1 and the remaining cell lines showing variable levels of phospho-ERK2 but very low levels of phospho-ERK1 **(Fig. 13A)**. Phospho-AKT2 levels varied between cell lines, while phospho-AKT1 levels were nearly similar between cell lines **(Fig. 13A)**. NF-κB DNA binding activity also showed cell line variability with KTB40 and KTB42 showing predominant p65:p50 DNA binding activity (**Fig. 13B**). OCT-1 and AP-1 DNA binding activity was least variable among KTB cell lines with the exception of very little OCT-1 DNA binding activity in KTB40 and KTB42 cell lines **(Fig. 13C and D)**. These results suggest that few of the reported activation of signaling pathways in cancer could be due to higher baseline activity in cells from which cancer originated instead of acquired changes during transformation.

BET bromodomain (BRD) proteins have recently been identified as major regulators of oncogenic transcription factors and BRD4 among them has been targeted therapeutically [31]. Two isoforms of BRD4 with opposing functions in cancer progression have been described; a long isoform with tumor suppressor activity and a short isoform with pro-metastatic functions [32]. KTB cell lines expressed different levels of both long and short forms of BRD4 **(Fig. 13E)**, which suggest that tumors originating from cells expressing short BRD4 isoform are predisposed for metastatic progression.

The effect of transforming growth factor-β (TGF-β) on stem/progenitor/mature cell hierarchy of KTB cell lines: TGF-β regulates various biological processes, including cell proliferation, extracellular matrix (ECM) synthesis, angiogenesis, immune response, apoptosis, and differentiation [33]. TGF-β also inhibits proliferation of ERα+ normal breast epithelial cells and it is believed that loss of this TGF-β-mediated growth inhibition is essential for the initiation of ERα+ breast cancer [34]. We examined the effect of TGF-β1 on characteristics of the cells by analyzing the levels of CD49f, CD201, CD271 and EpCAM markers with and without TGF-β1 treatment. TGF-β1-treated KTB6, KTB21, KTB36, KTB37 and KTB39 cells contained higher proportion of cells with basal/stem

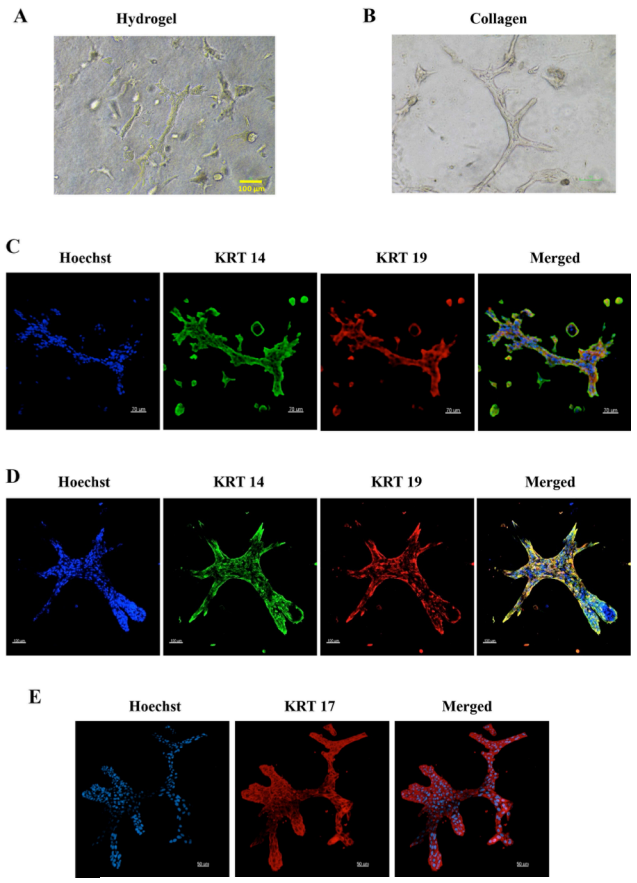
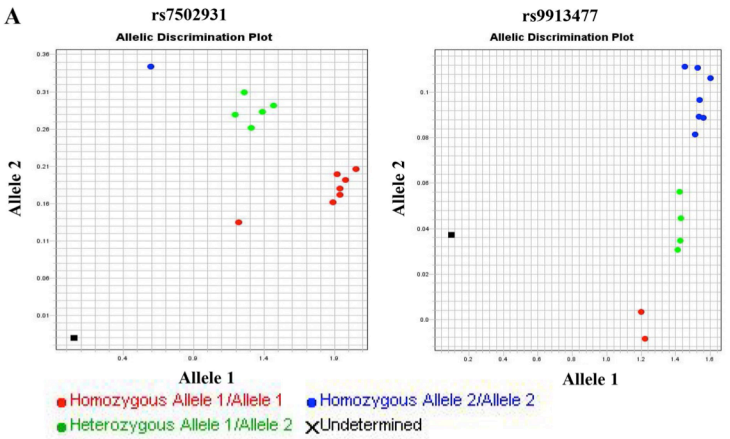


Fig. 17: KTB cell lines form ductal structures under 3D culture conditions. (**A**) Phase contrast images of hydrogel cultures. **(B)** Phase contrast images of collagen cultures. (**C**) KRT14 and KRT19 staining patterns of cells in hydrogel cultures. (**D**) KRT14 and KRT19 staining patterns of collagen cultures. (**E**) KRT17 staining patterns of collagen cultures.

cell characteristics compared to untreated cells as evident from higher number of CD49f+/EpCAM- and CD271+/EpCAM− subpopulations **(Fig. 14)**. Also, TGF-β1 had cell line-specific effects on CD49f, CD201, CD271 and EpCAM expression **(Fig. 15)**. Overall, these results document inter-individual variability in response to TGF-β1.



SNP	Homo/Heterozygous	Alleles	Cell lines
rs7502931	Homozygous	Allele 1/Allele 1	KTB26, KTB34, KTB39, KTB41, KTB32, KTB40, KTB42
		Allele 2/Allele 2	KTB22
	Heterozygous	Allele 1/Allele 2	KTB21, KTB6, KTB36, KTB37, KTB51
rs9913477	Homozygous	Allele 1/Allele 1	KTB21, KTB22
		Allele 2/Allele 2	KTB26, KTB34, KTB39, KTB41, KTB51, KTB32, KTB40
	Heterozygous	Allele 1/Allele 2	KTB6, KTB36, KTB37, KTB42

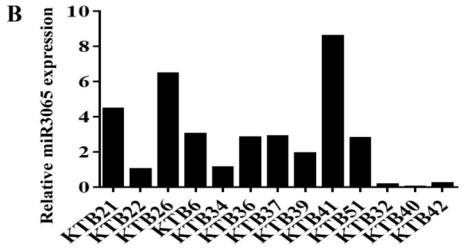


Fig. 18: (A) Allelic discrimination plots for rs7502931 and rs9913477 SNP revealing genotype plotting on the graph. Two representative plots showing distribution of alleles in KTB cell lines. These assays show clear separation of signals derived from allele 1 and allele 2. Allele 1 is shown in red and allele 2 is shown in blue. Green represents a mixture of the two alleles (allele 1/allele 2. Table shows the genotype of different KTB cell lines. (B) miR3065 expression in KTB cell lines.

The effect of transforming growth factor-β on cell proliferation of KTB cell lines: To determine the effect of TGFβ on cell proliferation, BrDU-incorporation-ELISA cell proliferation assay was performed. TGFβ1 inhibited proliferation but response varied between cell lines with highest inhibition of KTB34, KTB36, KTB37, KTB21, KTB26 and moderate inhibition of KTB6, KTB22, and KTB39 **(Fig. 16A)**. Note that KTB40 and 42, cell lines derived from CD201+/EpCAM- cells, were least sensitive to TGFβ1-mediated growth inhibition. However, there was no correlation between intrinsic subtype classification of these cell lines and the level of response to TGFβ1 **(Fig. 16)**. Overall, these results document inter-individual variability in response to TGFβ1, which includes growth inhibition, dedifferentiation or lack of response.

KTB cell lines generate ductal-like structure in 3D *in vitro* breast development assays: There is growing interest in developing organoid culture system to evaluate growth characteristics and individualized drug screening [35]. We used hydrogel and collagen gel 3D culture models to determine whether KTB cell lines could create structures similar to those found in normal breast [36, 37]. In both hydrogel and collagen scaffolds assays, KTB cell lines formed a complex branched tubular structure when cultured for 10-12 days **(Fig. 17A and 17B)**. These ductal structures had some resemblance to the interlobular ducts that form through the process of ductal initiation and branching **(Fig. 17A and 17B)**. While the outer layer of these structures contained either KRT14+ or KRT14+/KRT19+ cells, the inner layer contained few cells that stained only for KRT19+ cells **(Fig. 17C and 17D)**. We further confirmed that cells in these structures are epithelial by staining for KRT17, which is expressed in all epithelial cell types including mature luminal cells or KRT14+/KRT19+ luminal progenitor cells [38] **(Fig. 17E)**. Collectively, resources generated in this awar and assays utilized to characterize individual cell lines highlight the need to use a panel of "normal" cells from multiple sources as controls in signaling studies.

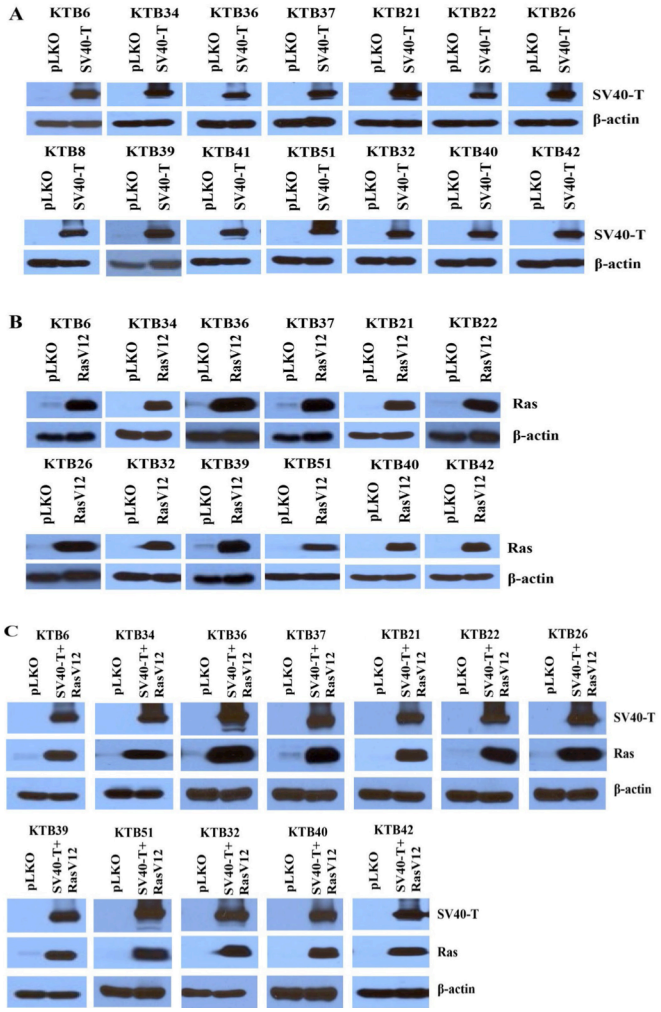


Fig. 19: (A) Western blotting was used to detect overexpression of SV40-T antigen in KTB6, 34, 36, 37, 21, 22, 26, 8, 39, 41, 51, 32, 40 and 42. (B) Western blotting was used to detect overexpression of mutant Ras in KTB6, 34, 36, 37, 21, 22, 26, 32, 39, 51, 40 and 42. (C) Western blotting was used to detect overexpression of SV40-T antigen and mutant Ras in double transformed KTB6, 34, 36, 37, 21, 22, 26, 39, 51, 32, 40 and 42. pLKO was used as a control cell line. β–actin was used as an internal control.

Immortalized cell lines carry single nucleotide polymorphisms (SNPs) linked to breast cancer susceptibility: Single nucleotide polymorphism (SNP) is the simplest form of DNA variation among individuals. These simple changes can be of transition or transversion type and they occur throughout the genome at a frequency of about one in 1,000 bp. They may be responsible for the diversity among individuals, genome evolution, the most common familial traits such as curly hair, interindividual differences in drug response, and complex and common diseases such as cancer, diabetes, obesity, hypertension, and psychiatric disorders. A recent study described SNPs in the genomic region coding for miR-3065 in African American women with breast cancer [39]. Four SNPs (rs9913477, rs1428882938, rs28585511, and rs7502931) reside in this genomic region and it includes both the predicted primary transcript of the noncoding miRNA gene miR3065 and the first intron of the gene for brain specific angiogenesis inhibitor 1-associated protein 2 (BAIAP2). To identify the presence of these SNPs in KTB cell lines, we performed genotyping. Genotyping with rs7502931 showed homozygous of allele 1 in KTB26, KTB34, KTB39, KTB41, KTB32, KTB40, KTB42, homozygous of allele 2 in KTB22 and heterozygous alleles in KTB21, KTB6, KTB36, KTB37, and KTB51 **(Fig. 18A)**. Genotyping with rs9913477 showed homozygous of allele 1 in KTB21, KTB22, homozygous of allele 2 in KTB26, KTB34, KTB39, KTB41, KTB51, KTB32, KTB40 and heterozygous alleles in KTB6, KTB36, KTB37, and KTB42 **(Fig. 18A)**. Furthermore, we measured miR3065 expression in all KTB cell lines. KTB21, KTB26 and KTB41 showed higher expression level, whereas KTB6, KTB36, KTB37, KTB39 and KTB51 showed moderate expression level. Expression level in KTB22 and KTB34 was lowest, while KTB32, KTB40 and KTB42 did not express miR-3065 **(Fig. 18B)**. Most importantly, SNP distribution pattern did not show any correlation with miR-3065 expression suggesting that the link between SNPs in this region and breast cancer risk it not due to altered miR-3065 expression.

Transformation of immortalized cell lines: Mutant Ras is one of the potent oncogenes used to transform breast epithelial cells in vitro. Although initially considered not a relevant oncogene in breast cancer, recent studies have clearly shown the role of Ras oncogene in endocrine resistance and metastasis of luminal breast cancer [40]. H-RasG12V overexpression often activates senescence program and once the senescence barrier is lost, transformation can be achieved. SV40T antigen overexpression results in inactivation of two tumor suppressor genes retinoblastoma and p53 [41]. Inactivation of retinoblastoma is observed in luminal B breast cancer, whereas p53 loss/mutation is common in triple negative breast cancers [42, 43]. Since KTB cell lines differed from HMECs, KTB cell lines from AA, CA and Hispanic women were transformed with SV40-T antigen, mutant HRasG12V (referred HRasG12V now onwards) and in combination of both SV40-T antigen and mutant HRasG12V using retrovirus and lentivirus, since this combination is the most effective in breast epithelial cell transformation [44]. Western blotting were used to detect the overexpression of SV40-T antigen in KTB6, KTB34, KTB36, KTB37, KTB21, KTB22, KTB26, KTB8, KTB39, KTB41, KTB51, KTB32, KTB40 and KTB42 cell lines **(Fig. 19A),** mutant HRasG12V in KTB6, KTB34, KTB36, KTB37, KTB21, KTB22, KTB26, KTB39, KTB51, KTB32, KTB40 and KTB42 cell lines **(Fig. 19B)** and combination of both SV40-T and mutant HRasG12V in KTB6, KTB34, KTB36, KTB37, KTB21, KTB22, KTB26, KTB39, KTB51, KTB32, KTB40 and KTB42 cell lines **(Fig. 19C)** to ensure that transformation is oncogene driven but not spontaneous. Phase contrast images of few of the transformed cell lines (KTB34, KTB6, KTB22, KTB39, KTB32, KTB40 and KTB42) are shown in **Fig. 20**. To determine the effect of SV40-T antigen, mutant HRasG12V, and combinations of both on cell proliferation, BrDU-incorporation-ELISA cell proliferation assay was performed. SV40-T antigen promoted cell proliferation with variable growth rate between cell lines, while effect of mutant HRasG12V on proliferation was modest (KTB34 and KTB6 both luminal) or none (KTB39 and KTB22, normal- and basal-like cell lines, respectively) **(Fig. 21)**.

Characterization of transformed cells for stem/progenitor/differentiated cell hierarchy: SV40-T antigen and mutant HRasG12V overexpressing cells were first analyzed by flow cytometry using various markers (CD49f, PROCR (CD201), CD271, CD44, CD24 and EpCAM) to determine the subpopulation of cells enriched in CA, AA or Hispanic after transformation.

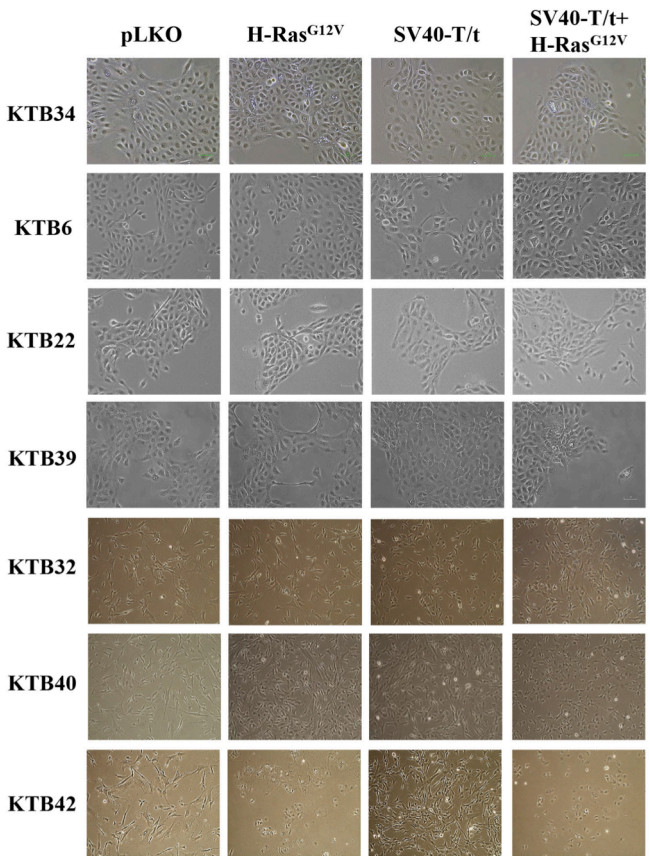


Fig. 20: Phase contrast images showing morphology of KTB34, KTB6, KTB22, KTB39, KTB32, KTB40 and KTB42 transformed with HRasG12V, SV40-T antigen and SV40-T+HRasG12V.

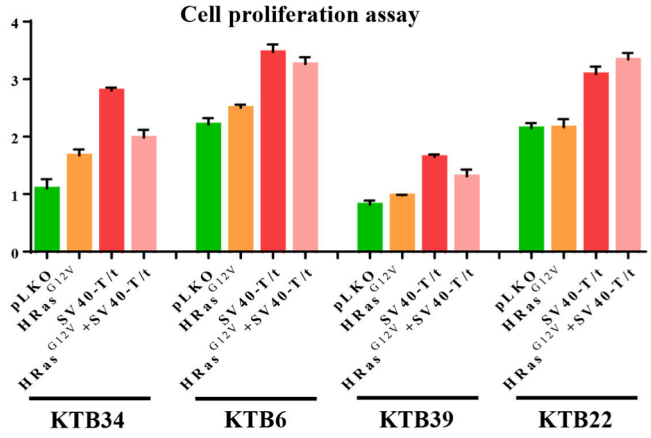


Fig. 21: Cell proliferation rates of KTB34, KTB6, KTB39 and KTB22 transformed cells as measured by BrDU-incorporation ELISA cell proliferation assay.

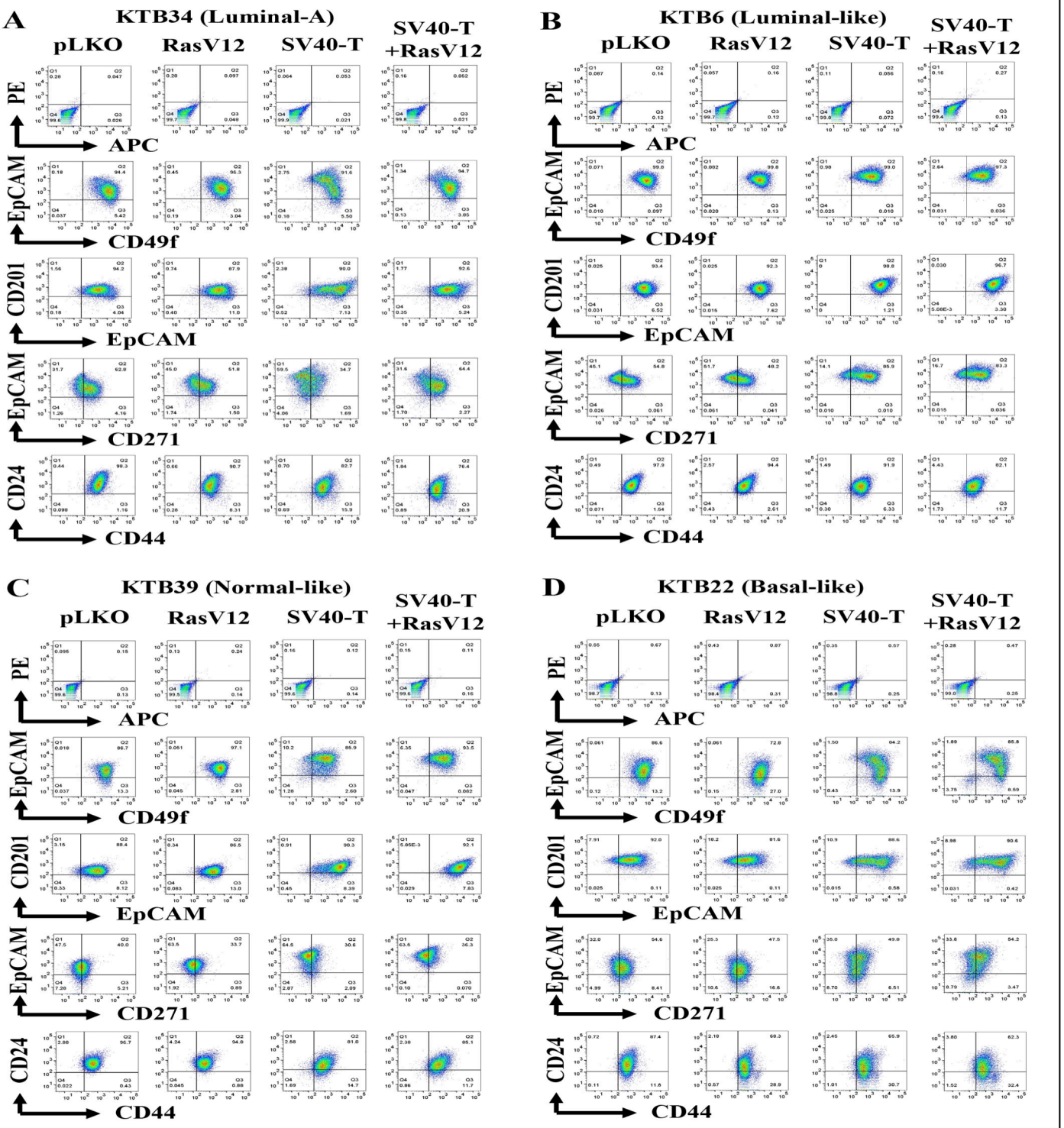


Fig. 22: Stem, progenitor and differentiated/mature cell state of transformed cell lines. (A) Phenotype of KTB34 cells with pLKO, HRasG12V, SV40-T antigen and SV40-T+HRasG12V. (B) Phenotype of KTB6 cells pLKO, mutant HRasG12V, SV40-T antigen and SV40-T+HRasG12V. (C) Phenotype of KTB39 cells with pLKO, HRasG12V, SV40-T antigen and SV40-T+HRasG12V. (D) Phenotype of KTB22 cells with pLKO, HRasG12V, SV40-T antigen and SV40-T antigen+HRasG12V. Cells were stained with indicated antibodies and flow cytometry was used to identify subpopulation of cells.

Effect of SV40-T antigen on cell surface marker profiles: SV40-T antigen overexpression had cell line specific effects on subpopulation of cells. For example, SV40-T antigen altered CD49f/EpCAM staining pattern with distinct subpopulations in KTB34, KTB39 and KTB22 but not KTB6 cell line **(Fig. 22A, B, C, D)**. Similar cell line specific changes in stem/progenitor/mature cell hierarchy upon SV40-T antigen overexpression were observed when cells were analyzed for CD201/EpCAM, CD271/EpCAM and CD44/CD24 expression patterns **(Fig. 22A, B, C, D).** All cell lines showed elevated EpCAM staining intensity upon SV40 T-antigen overexpression. Thus, SV 40-T antigen can influence stem/progenitor/differentiation hierarchy of transformed cells.

Effect of mutant HRasG12V on cell surface marker profiles: Unlike SV40 T-antigen, HRasG12V overexpression had minimum effects on phenotype of cell lines **(Fig. 22A, B, C, D)**. Notably, KTB22 showed a modest increase in the levels of CD49f+/EpCAM-, CD44+/CD24-, CD201+/EpCAM- and CD271+/EpCAM- subpopulations **(Fig. 22D)**.

Effect of overexpression of both SV40-T antigen and HRasG12V on cell surface markers: The phenotype of double transformed cells closely resembled that of SV 40-T antigen transformed cells with an increase in EpCAM staining intensity and individual specific effects on CD271+ cells **(Fig. 22A, B, C, D)**. A unique subpopulation was observed in KTB22 double transformed cells i.e. $CD49f^{low}/EpCAM^{low}$ and CD44+/CD24- population **(Fig. 22D)**. Taken together, these results indicate that transformation with SV40-T antigen and mutant HRasG12V alters the stem/progenitor/mature cell hierarchy on an individual basis.

Effect of SV40-T antigen overexpression on CD201+/EpCAM- (PZP) cells: To examine the effect of SV40-T antigen on CD201+/EpCAM- cells (KTB32, KTB40 and KTB42), we overexpressed the SV40-T antigen in KTB32, KTB40 and KTB42 cells. Overexpression of SV40-T antigen increased the levels of CD49f+/EpCAM- and CD201+/CD90+ cells **(Fig. 23A, B, C)**. CD90 and CD10 markers are used to determine the rare endogenous pluripotent somatic stem cells and myoepithelial cells (45).

Effect of HRasG12V on CD201+/EpCAM- cells: To examine the effect of HRasG12V on CD201+/EpCAM- cell lines (PZP cell lines), we overexpressed the mutant HRasG12V in KTB32, KTB40 and KTB42 cells. HRasG12V expressing cells showed morphological changes with increased cell growth **(Fig. 20)** and elevated levels of CD49f+/EpCAM-, CD201+/CD90-, CD201+/CD10+ and CD201+/EpCAM+ cells on an individual basis **(Fig. 23)**.

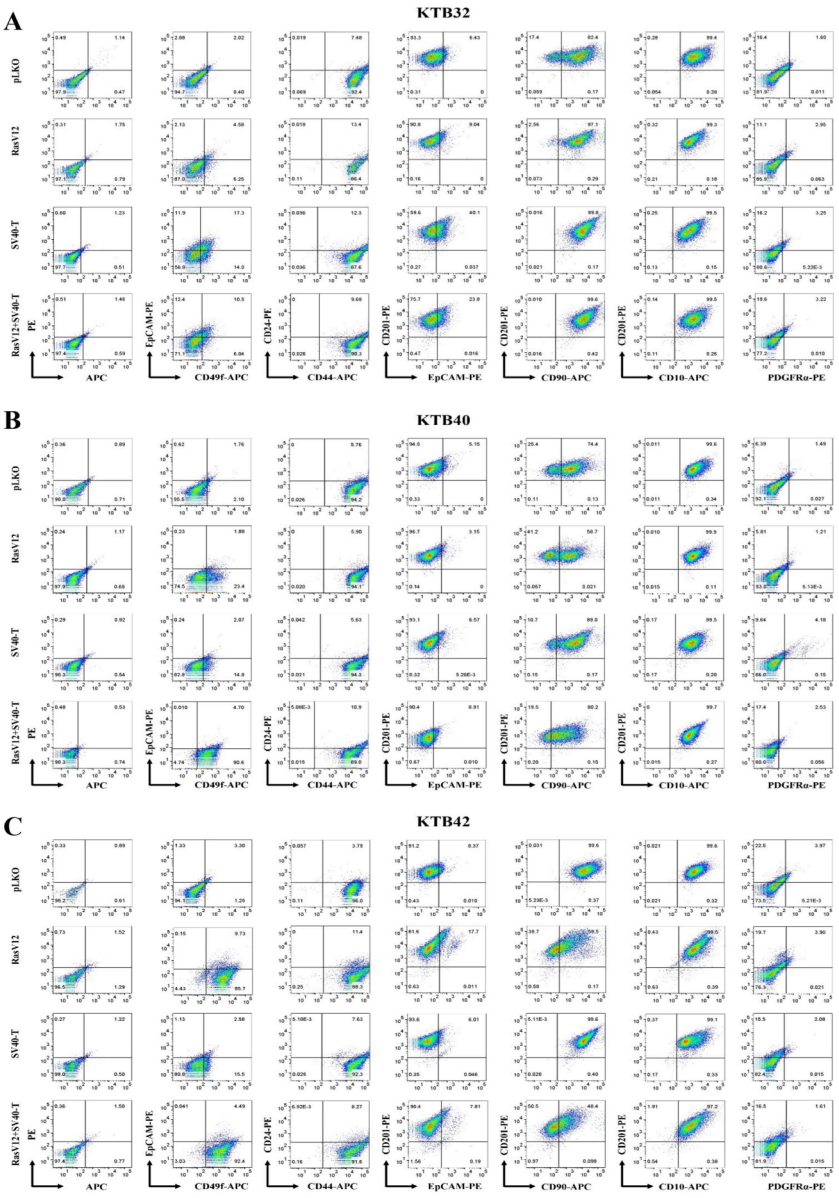


Fig. 23: KTB32 (A), KTB40 (B) and KTB42 (C) transformed cells were stained with indicated antibodies and flow cytometry was used to identify subpopulation of cells.

Effect of SV40-T antigen and HRasG12V double overexpression on CD201+/EpCAM- (PZP) cells: The phenotype of double transformed cells closely resembled that of HRasG12V transformed cells **(Fig. 23)**. Thus, unlike in EpCAM+ luminal, normal-like and basal cell lines, SV40-T antigen failed to alter stem/progenitor/differentiated cell hierarchy of PZP cells. Collectively, the cell-of-origin and oncogenic mutations determine differentiation capabilities of transformed cells.

Role of SV40-T antigen and HRasG12V in basal activity of signaling networks: To extend our observations on cell-of-origin dependent variability in basal and transformation-mediated signaling changes, we measured OCT-1, AP-1 and NF-κB DNA binding activity in immortalized and transformed cell lines. While overexpression of HRasG12V had minimum effect in NF-κB DNA binding activity, SV40-T antigen increased NF-κB **(Fig. 24).** The normal-like cell line from African American women displayed highest induction. We did not observe an effect of oncogenes on AP-1 and Oct-1 binding activity. These results suggest that signaling pathway activation in transformed cells is dependent on both cell-of-origin and the oncogenic event.

Stemness properties of transformed cell lines: To determine whether transformed KTB cell lines show variability in stemness property, we performed mammosphere assay in KTB34 and 39 cells transformed with HRasG12V. KTB34 and KTB39 transformed cell lines showed larger mammospheres compared to pLKO control, representing increased stemness properties **(Fig. 25A)**. TGFβ1 increased the size of mammospheres from pLKO but not transformed cells **(Fig. 25A)**. Mammospheres were subjected to flow cytometry for CD49f/EpCAM, CD201/EpCAM, CD271/EpCAM and CD44/CD24 expression to distinguish stem, luminal progenitor, basal and mature/differentiated cells. Notably, more CD44-/CD24+ and CD49f-/EpCAM+ cells, which were never observed under 2D culture of primary cells, were found in mammosphere cultures of the cell lines **(Fig. 25B),** representing mature/differentiated cells. Mammospheres of transformed KTB34 but not KTB39 cells showed more CD49f+/EpCAM- and CD49f+/EpCAM+ subpopulations and reduced CD49f-/EpCAM+ subpopulation compared to pLKO control. Inter-individual differences were also noted between KTB34 and KTB39. As noted above in PAM50 classification, KTB34 is a luminal cell line, whereas KTB39 is a basal cell line. In mammospheres, KTB34 but not KTB39 (both control and transformed) contained CD44-/CD24+ population. By contrast, KTB39 but not KTB34 mammospheres contained CD201+/EpCAM- cells **(Fig. 25B).** We further examined the effect of TGFβ1 on characteristics of transformed cells by analyzing the levels of CD49f, CD201, CD271 and EpCAM markers. TGFβ1-treated pLKO and HRasG12V transformed cells contained higher proportion of cells with stem cell characteristics compared to untreated cells as evident from higher number of CD49f+/EpCAM- and CD44+/CD24- subpopulations **(Fig. 25B)**. We have also performed mammosphere assay in CD201+/EpCAM- KTB42 cells transformed with mutant HRasG12V. KTB42 transformed cell lines showed larger mammospheres compared to pLKO control, representing increased stemness properties **(Fig. 26A)**. Mammospheres of KTB42 transformed cells showed more CD49f+/EpCAM-, CD201+/EpCAM-, CD201+/CD90+, CD201+/CD90- and CD201+/CD10+ subpopulations of cells compared to pLKO control **(Fig. 26B).** Overall, these results document HRasG12V has cell type-specific effects on stemness properties.

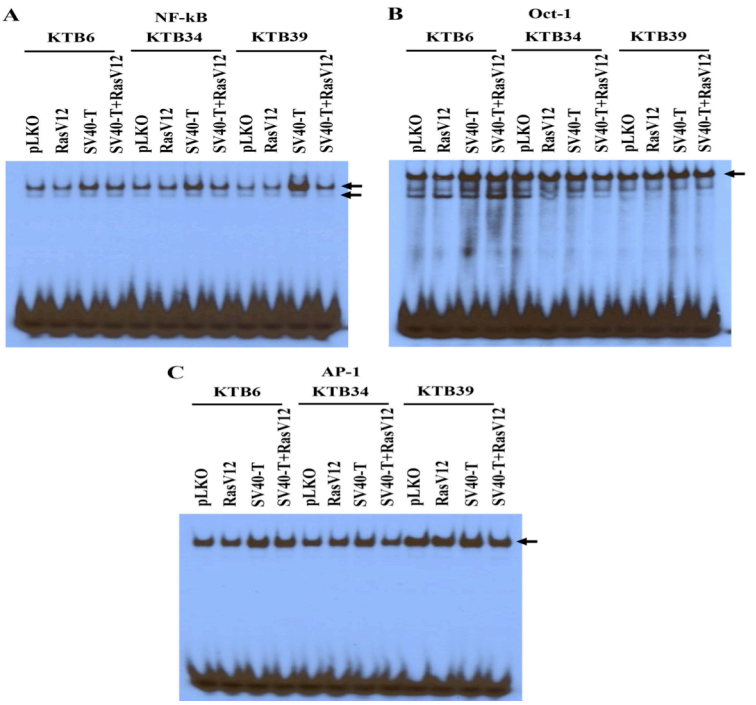


Fig. 24: DNA binding activity of various transcription factors in transformed cell lines. (A) NF-κB DNA binding activity, (B) OCT-1 DNA binding activity, and (C) AP-1 DNA binding activity.

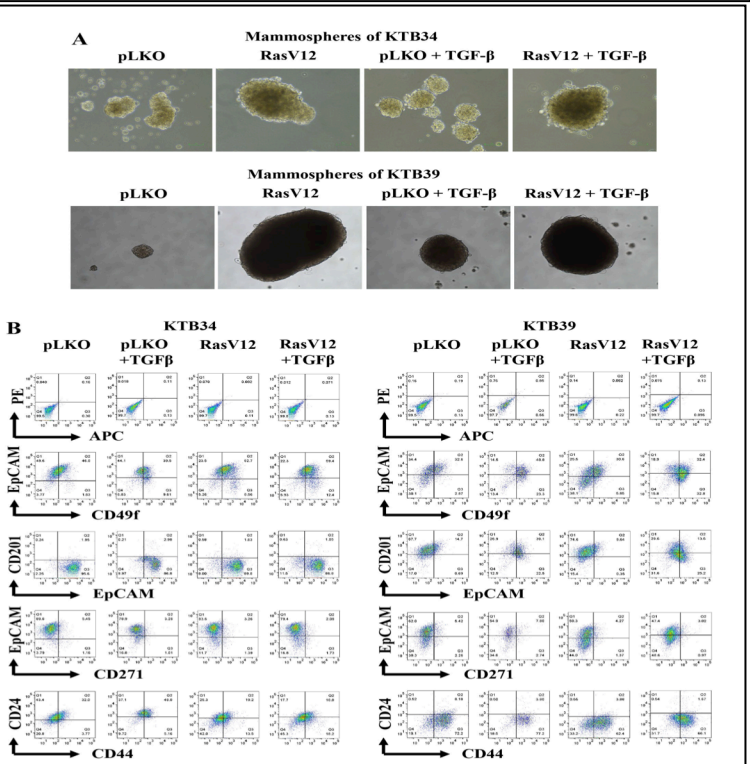


Fig. 25: (A) HRasG12V transformed KTB34 and KTB39 cell lines formed mammospheres, which were larger than mammospheres generated by pLKO cells. (B) Transformed cells in mammospheres show different levels of stem/progenitor/mature cells based on CD49f/EpCAM and CD44/CD24 staining pattern. Mammospheres from pLKO cells were used as controls.

Tumorigenecity of transformed cells: To determine whether KTB cell lines expressing SV40-T antigen and mutant HRasG12V form tumors, we injected 5 X 10^6 transformed cells with 50% matrigel in 100 microliter HBSS into the mammary fat pad/subcutaneous of 6-7 week old female NSG (NOD/SCID/IL2Rgnull) mice. Tumor was resected and analyzed by H&E staining and expression of estrogen receptor alpha (ERα), GATA3, FOXA1, CK5/6, CK8, CK14 and CK19 using immunohistochemistry. We also created cell lines from the half of tumors to characterize the phenotypic cellular heterogeneity using various cell surface markers. We used human specific antibody against Na+/K+ ATPase CD298 (ATP1B3) cell surface marker to sort the CD298-enriched human tumor cells populations from surrounding mouse cells [45]. Notably, SV40-T antigen alone expressing cells were not tumorigenic. Among HRasG12V alone transformed cell lines, only KTB22 developed tumors suggesting that HRasG12V alone can transform depending on the cell type.

Immunohistochemistry study of tumors: To determine whether tumors express luminal or basal markers, we performed immunohistochemistry for the expression of ERα, GATA3, FOXA1, CK5/6, CK14 and CK19. ERα-FOXA1-GATA3 transcription factor network generally expressed in hormonally responsive luminal cells [7]. The luminal cells express cytokeratin 19 (CK19), while basal cell types express cytokeratin 5/6 (CK5/6) and cytokeratin 14 (CK14), and cells expressing both CK14 and CK19 show luminal progenitor phenotype [9].

H&E staining of four KTB34-SV40T+HRasG12V tumors showed poorly differentiated carcinoma, moderately differentiated squamous cell carcinoma and well-differentiated squamous cell carcinoma **(Fig. 27)**. KTB6-SV40T+HHRasG12V tumor showed undifferentiated carcinoma **(Fig. 27)**. Five KTB39-SV40T+HRasG12V tumors showed anaplastic carcinoma with squamous cell features and undifferentiated carcinoma, poorly differentiated squamous cell carcinoma and anaplastic squamous sarcoma **(Fig. 27)**. KTB22-SV40T+HRasG12V tumors showed squamous cell carcinoma, anaplastic, adenosquamous and adenoid cystic carcinomas **(Fig. 27)**. KTB22-HRasG12V tumor showed undifferentiated carcinoma **(Fig. 28)**.

We next examined the ERα, GATA3 and FOXA1 expressions in all tumors by immunohistochemistry. Staining of KTB34-SV40T+HRasG12V – derived tumors showed ERα-/GATA3-/FOXA1+, ERα-/GATA3+/ FOXA1-, ERα-/GATA3+/FOXA1+ and ERα-/GATA3+/FOXA1- staining patterns **(Fig. 27).** KTB6-SV40-T+HHRasG12V cell derived tumors were ERα-/GATA3+/FOXA1- and ERα-/GATA3-/FOXA1- **(Fig. 27)**. All tumors derived from KTB39-SV40T+HRasG12V cells were ERα-/GATA3-/FOXA1- **(Fig. 27)**. KTB22-SV40-T+HRasG12V cell-derived tumors were ERα-/GATA3+/ FOXA1+ and ERα-/GATA3-/ FOXA1+ **(Fig. 27).** KTB22-HRasG12V cell-derived tumor was also ERα-/GATA3-/FOXA1- **(Fig. 28)**.

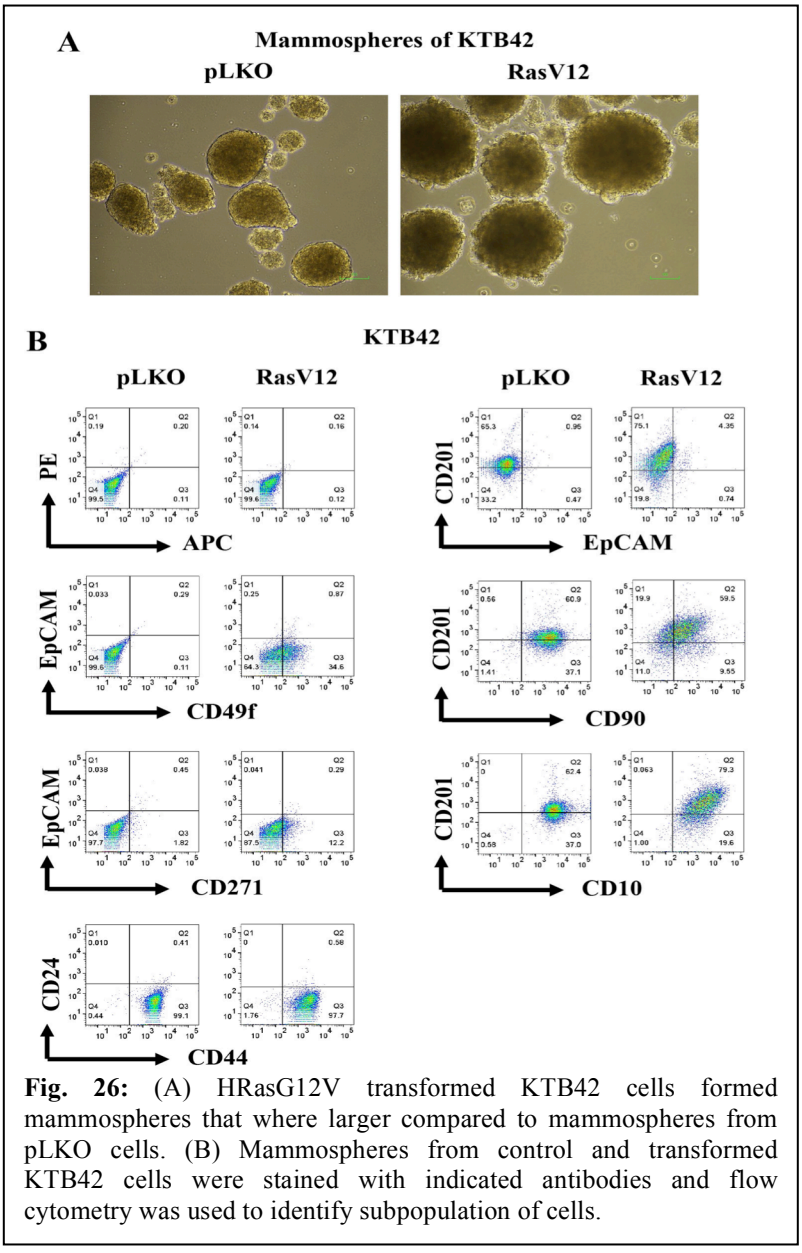


Fig. 26: (A) HRasG12V transformed KTB42 cells formed mammospheres that where larger compared to mammospheres from pLKO cells. (B) Mammospheres from control and transformed KTB42 cells were stained with indicated antibodies and flow cytometry was used to identify subpopulation of cells.

We examined lungs of tumor-bearing mice to identify metastasis. All KTB34- SV40T+HRasG12V tumors, KTB6-SV40-T+HHRasG12V tumor, KTB22-SV40-T+HRasG12V tumor and KTB22-HRasG12V tumor showed metastasis to lung with few ERα-/GATA3+/FOXA1+ cells, whereas none of the KTB39-SV40T+HRasG12V tumors showed lung metastasis **(Fig. 28 and 29)**.

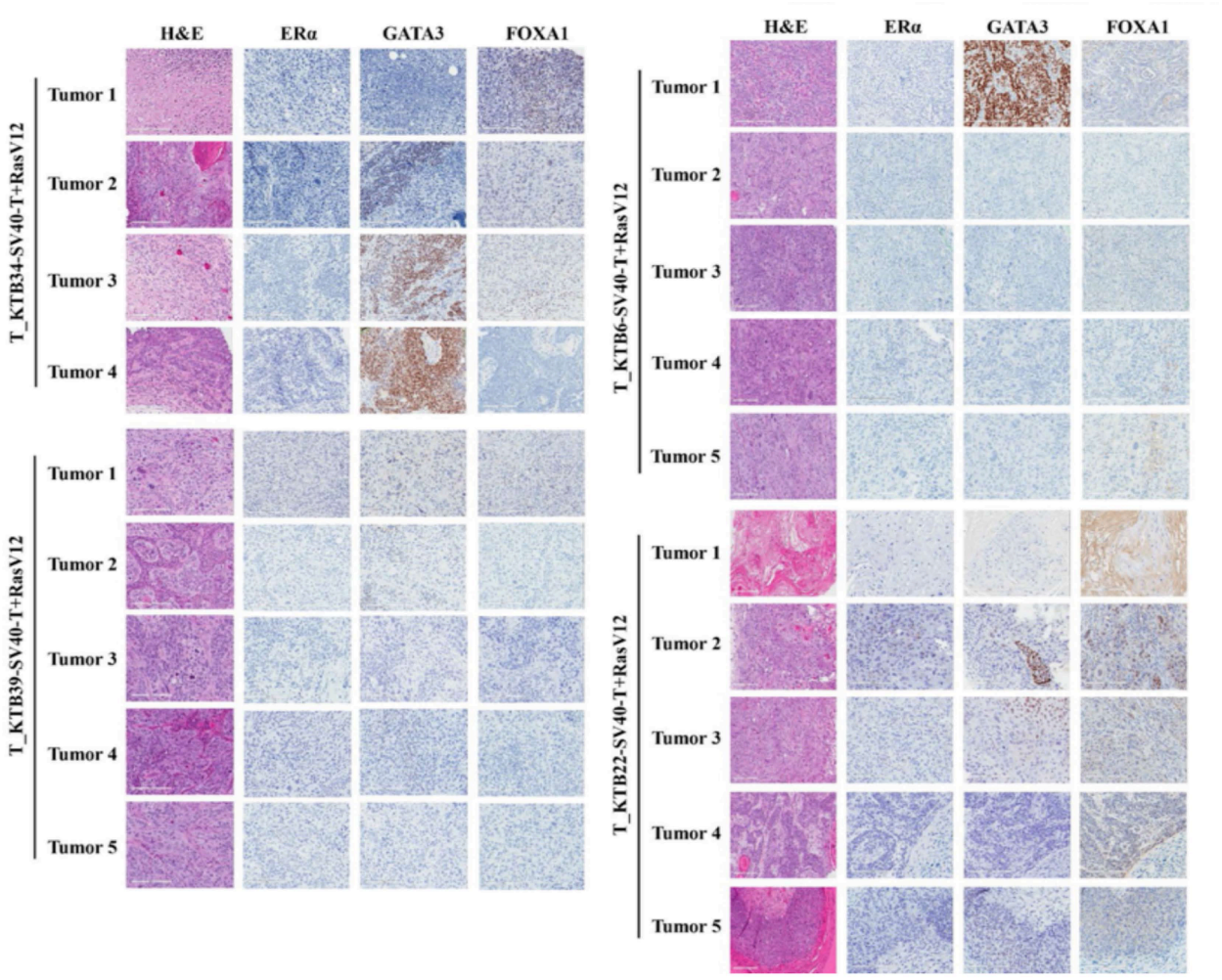


Fig. 27: Immunohistochemistry analyses of luminal markers ERα, GATA3 and FOXA1 in tumors developed from the KTB34-SV40-T+HRasG12V, KTB6-SV40-T+HRasG12V, KTB39-SV40-T+HRasG12V and KTB22-SV40-T transformed cells. H&E staining shows the type of tumors. Representative immunohistochemistry (IHC) data are shown.

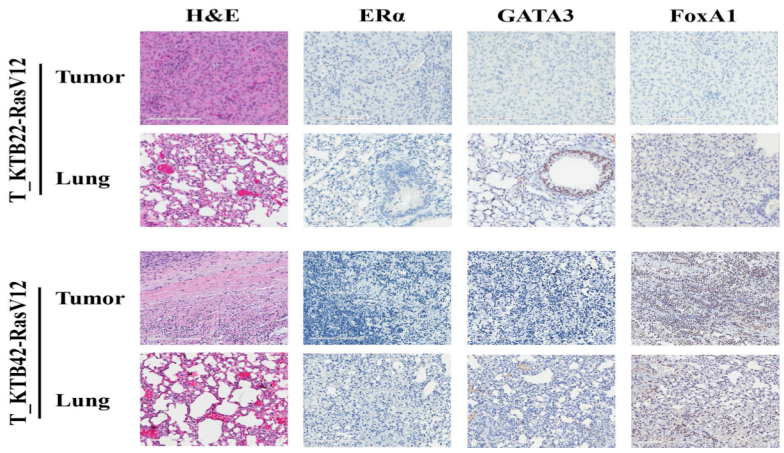


Fig. 28: Immunohistochemistry analyses of luminal markers ERα, GATA3 and FOXA1 in tumors and lungs developed from the KTB22-HRasG12V and KTB42-HRasG12V transformed cells. H&E staining shows the type of tumors.

We examined the CK5/6, CK14 and CK19 expressions in all tumors and lungs by immunohistochemistry. Kertain expression profiles in KTB34-derived tumors were CK5/6-/CK14-/CK19-, CK5/6+/CK14+/CK19-, CK5/6-/CK14+/CK19- and CK5/6-/CK14+/CK19- **(Fig. 30)**. KTB6-SV40T+HRasG12V tumors showed CK5/6-/CK14+/CK19- and few tumors with small population of CK19+ expression **(Fig. 30)**. All KTB39-SV40T+HRasG12V tumors were CK5/6+/CK14+/CK19- **(Fig. 30)**. KTB22-SV40T+HRasG12V tumors showed CK5/6+/CK14+/CK19- and CK5/6+/CK14+/CK19+ expression **(Fig. 30)**. KTB22-HRasG12V-derived tumor was CK5/6-/CK8-/CK14+/CK19- **(Fig. 31)**. While CK14 expression in tumors that were positive was uniform, CK5/6 expression in the same tumor was mosaic, suggesting heterogeneity in those tumors. All the tumors derived from KTB34, KTB6 and KTB39 double transformed cells were enriched for EGFR expression, while KTB22 double transformed cells formed both EGFR+ and EGFR- tumors **(Fig. 30)**. Taken together, immunohistochemistry analysis of ERα, GATA3, FOXA1, CK5/6 and CK14 again confirmed the inter-tumor heterogeneity, which could be due to differences in cell-of-origin of tumors.

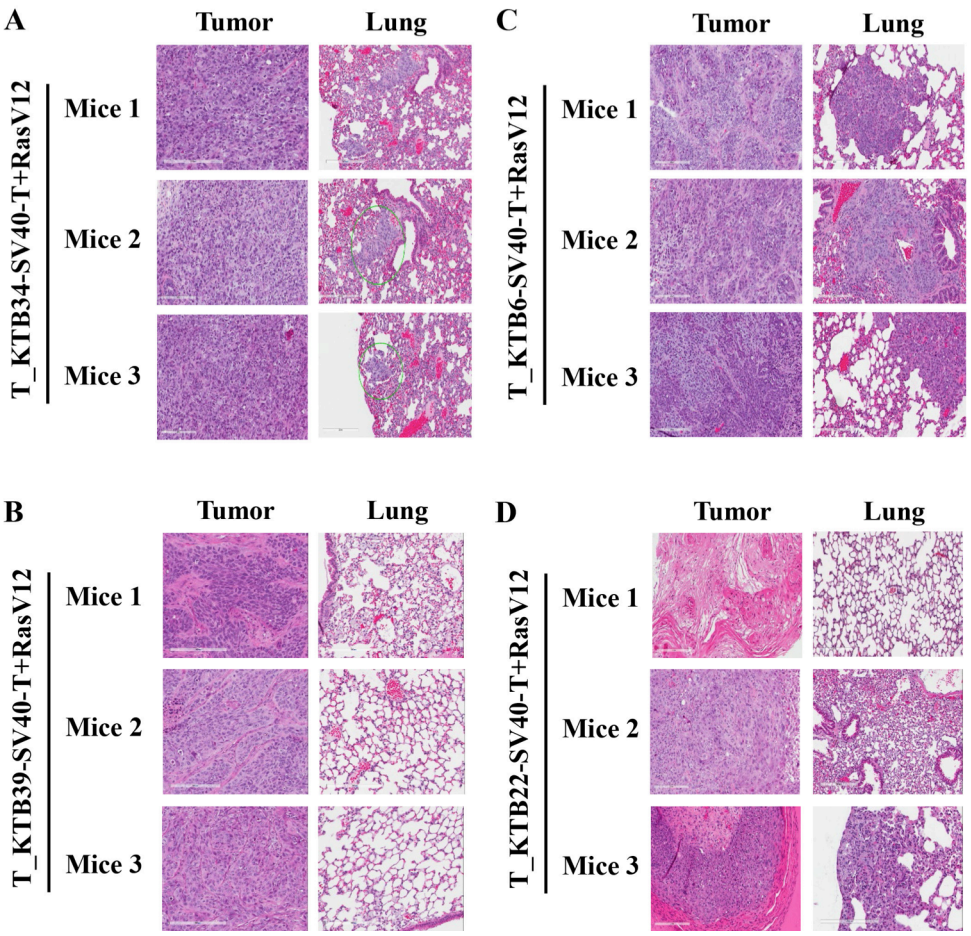


Fig. 29: Tumor developed from KTB34-SV40-T+HRasG12V, KTB6-SV40-T+HRasG12V and KTB22-SV40-T+HRasG12V but not KTB39-SV40-T+HRasG12V transformed cell lines metastasized to lungs. Representative immunohistochemistry (IHC) data are shown.

Tumorigenecity of CD201+/EpCAM- (PZP) KTB42 transformed cells: CD201+/EpCAM- KTB42 cell line transformed with mutant HRasG12V developed tumors in NSG mice. KTB42-HRasG12V cell-derived tumor was ERα-/GATA3-/FOXA1+ **(Fig. 28)** and showed lung metastasis, which exhibited few ERα-/GATA3+/ FOXA1+ cells in the lung **(Fig. 28)**. KTB42-HRasG12V-derived tumor was CK5/6-/CK8-/CK14-/CK19- **(Fig. 28)**. KTB42 cell line transformed with both mutant HRasG12V and SV40-T antigen also developed tumors in NSG mice. KTB42-HRasG12V+SV40-T cell-derived tumor was pleomorphic anaplastic sarcoma with ERα-/GATA3-/FOXA1- expression **(Fig. 32A).** KTB42-HRasG12V-derived tumor was CK5/6-/CK14-/CK19- **(Fig. 32B)** and showed lung metastasis **(Fig. 32C)**.

Tumorigenecity of KTB34 and KTB39 cells after transformation with mutant PIK3CA (PIK3CA^{H1047R}) and SV40-T antigen: PIK3CA mutations is one of the major route of PI3K signal activation in breast cancer and has been shown to correlate with expression of estrogen and progesterone receptors (ER/PR), ERBB2, and lymph node metastasis [43, 46]. The normal-like KTB39 cell line transformed with mutant PIK3CA alone or in combination with SV40-T antigen did not form tumor. Luminal cell line KTB34 cells transformed with mutant PIK3CA alone did not form tumor, but in combination with SV40-T antigen formed the tumors. Staining of KTB34-SV40T+ PIK3CA^{H1047R} – derived tumors showed both ERα-/GATA3+/FOXA1+ and ERα-/GATA3-/ FOXA1- subpopulations **(Fig. 33A).** Keratin expression profiles in KTB34-SV40T+ PIK3CA^{H1047R} - derived tumors were CK5/6-

/CK14+/CK19+ with small population of CK19+ expression and all the tumors were enriched for EGFR expression **(Fig. 33B)**. KTB34-SV40T+ PIK3CA^{H1047R} cell - derived tumors were adenocarcinoma and poorly differentiated squamous cell carcinoma **(Fig. 33C)** suggesting that the cell-of-origin but not oncogenes determine tumor phenotypes**.**

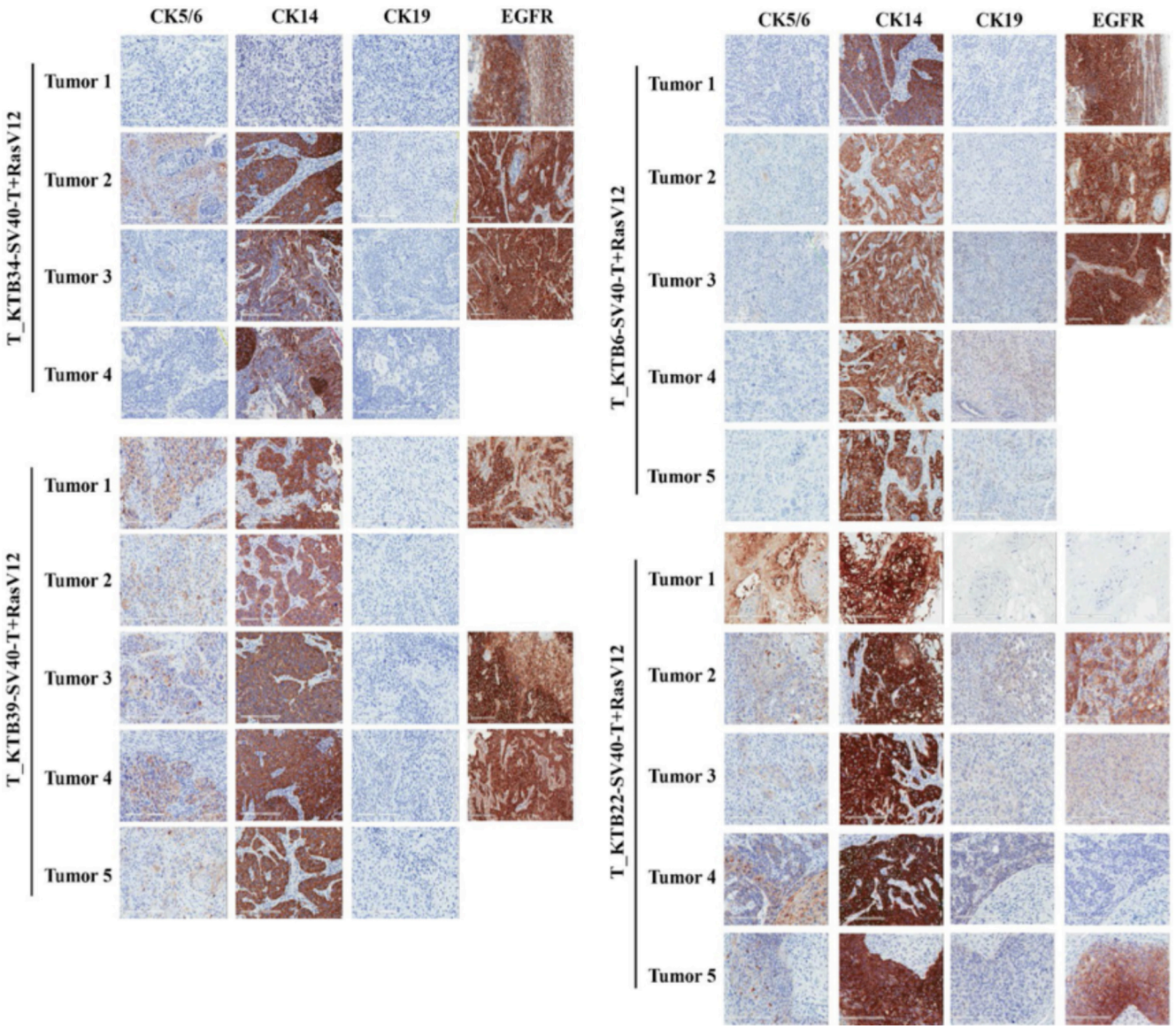


Fig. 30: Immunohistochemistry analyses of distinct cytokeratins CK5/6, CK14, CK19 and EGFR in tumors developed from KTB34-SV40-T+HRasG12V, KTB6-SV40-T+HRasG12V, KTB39-SV40-T+HRasG12V and KTB22-SV40-T transformed cells. Basal marker: CK5/6 and CK14, Luminal marker: CK19. Representative immunohistochemistry (IHC) data are shown.

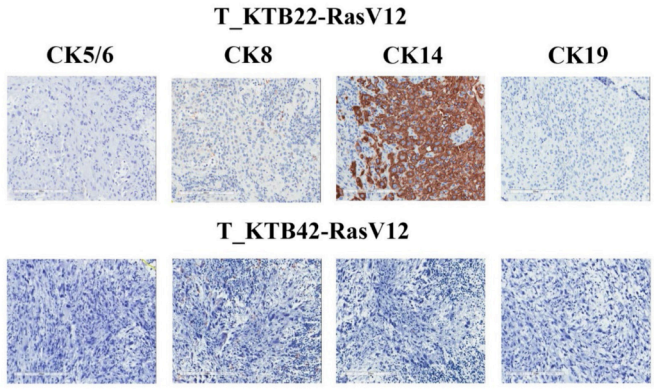


Fig. 31: Immunohistochemistry analyses of cytokeratins CK5/6, CK14, CK8 and CK19 in tumors developed from the KTB22-HRasG12V and KTB42-HRasG12V transformed cells. Basal marker: CK5/6 and CK14, Luminal marker: CK8 and CK19.

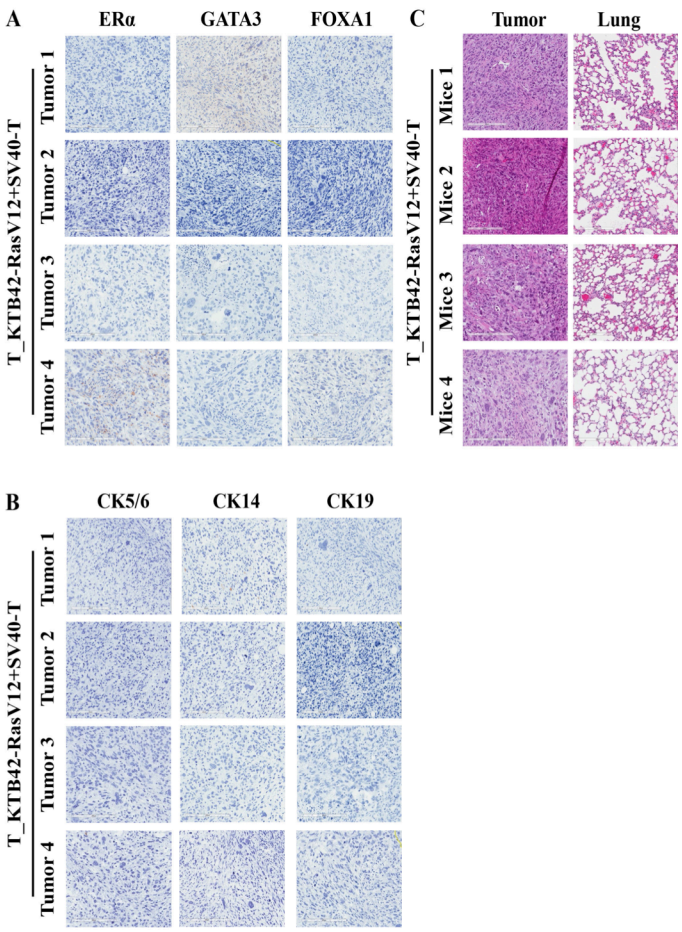


Fig. 32: (A) Immunohistochemistry analyses of luminal markers ERα, GATA3 and FOXA1 in tumors developed from the KTB42-HRasG12V+SV40-T transformed cells. H&E staining shows the type of tumors. (B) Immunohistochemistry analyses of distinct cytokeratins CK5/6, CK14 and CK19 in tumors developed from KTB42-HRasG12V+SV40-T transformed cells. Basal marker: CK5/6 and CK14, Luminal marker: CK19. (C) H&E staining shows the type of tumors and metastasis to lung. Representative immunohistochemistry (IHC) data are shown.

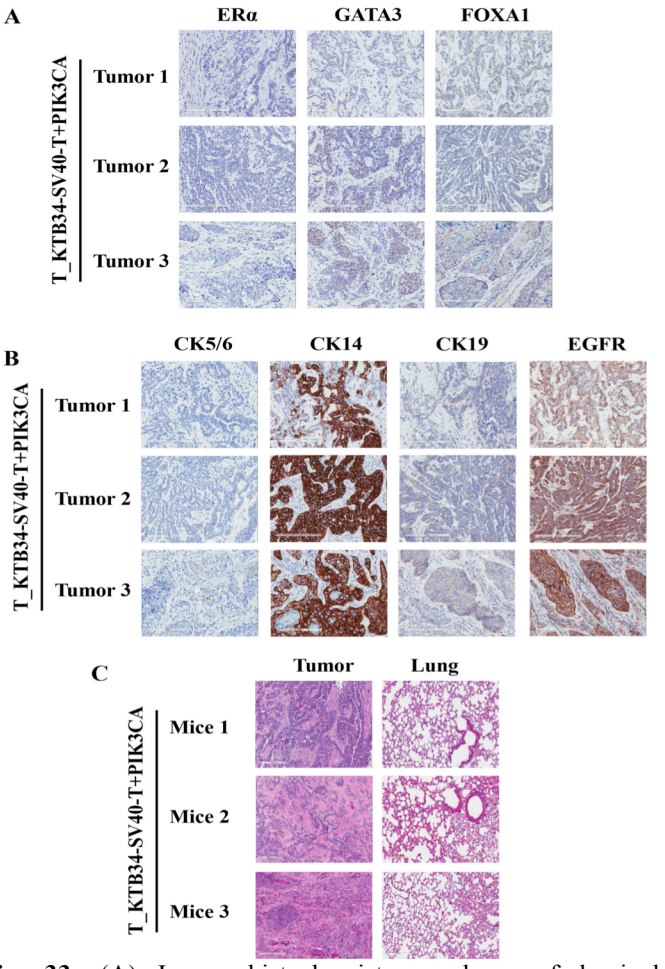


Fig. 33: (A) Immunohistochemistry analyses of luminal markers ERα, GATA3 and FOXA1 in tumors developed from the KTB34-SV40-T+PIK3CA transformed cells. H&E staining shows the type of tumors. (B) Immunohistochemistry analyses of distinct cytokeratins CK5/6, CK14, CK19 and EGFR in tumors developed from the KTB34-SV40-T+PIK3CA transformed cells. Basal marker: CK5/6 and CK14, Luminal marker: CK19. (C) H&E staining shows the type of tumors and metastasis to lung.

Characterization of cell lines derived from tumors:

KTB34-SV40T+HRasG12V transformed cells (Caucasian origin; luminal A type cell line): Since each cell lines generated distinct subtypes of tumors, we next wanted to characterize properties of tumors derived from single transformed cell. For this purpose, we first established cell lines from tumors. Cell lines established from tumors of KTB34 transformed cells with SV40-T antigen and HRasG12V showed two distinct phenotypes, such as epithelial and senescence-like flat and enlarged cell morphology **(Fig. 34A)**. We examined the expression of SV40-T antigen and Ras proteins in tumor-derived cell lines using western blotting. Compared to pLKO control, tumor-derived cell lines showed the expression of SV40-T antigen and higher levels of Ras **(Fig. 34B)**. We further characterized these tumor cells by flow cytometry using various markers (CD49f, CD201, CD271, CD44, CD24 and EpCAM) to determine the enriched subpopulation of cells after transformation. Staining of double transformed KTB34 tumor cells showed two distinct subpopulations of CD49f+/EpCAM- and CD49f+/EpCAM+; CD44+/CD24- and CD44+/CD24+; CD201+/EpCAM- and CD201+/EpCAM+; and CD271+/low/EpCAM- and CD271+/low/EpCAM+ cells **(Fig. 34C)**. These results indicate that there is selection of transformed cells *in vivo* as the majority of transformed cells gained CD201 expression or CD201+ transformed cells were selected *in vivo*. Please note that transformed KTB34 cell line *in vitro* contained very low number of CD201+/EpCAM- cells **(Fig. 22)**. We also noted inter-tumor phenotypic heterogeneity **(Fig. 34C)**. Thus, despite the same oncogenic mutations, there is cell type dependent variability in phenotype of transformed cells.

We used soft agar assay as a surrogate to further confirm the transformed phenotype of tumor cells driven by SV40-T antigen and HRasG12V. As expected, cells from tumor 1, tumor 2 and tumor 3 of KTB34-SV40T+HRasG12V formed colonies on soft agar **(Fig.**

35A). We next established 2D cultures of these soft agar clones by transferring soft agar colonies to tissue culture plates to obtain single cell-derived transformed cells. Tumor cells derived from soft agar showed morphological changes with aggressive phenotype compared to parental cells **(Fig. 35B)**. Western blotting was used to determine the expression of SV40-T antigen and mutant Ras in these cells. Each clone from tumor 1, tumor 2 and tumor 3 of KTB34-SV40T+HRasG12V transformed cells showed expression of SV40-T antigen and mutant Ras at variable levels. Clone 3 of tumor 1, clones 1, 2, 4, 5 of tumor 2 and clones 1-6 of tumor 3 showed very strong expression of mutant Ras **(Fig. 35C)**. To further characterize these clones for stem/progenitor/mature cell phenotypes, we performed flow cytometry with CD49f, EpCAM, CD44, and CD24 surface markers. All the clones derived from tumor 1, tumor 2 and tumor 3 showed predominantly CD49f$^{+/low}$/EpCAM- and CD44+/CD24- subpopulation of cells representing cancer stem cells **(Fig. 36).** Clone 6 of tumor 1, clones 3, 6 of tumor 2 and clones 2, 4, 5 of tumor 3 also contained CD49f+/EpCAM+ and CD44+/CD24+ subpopulation of cells, which suggest that these clones have the capacity to undergo some level of differentiation into cells with luminal progenitor phenotype **(Fig. 36)**.

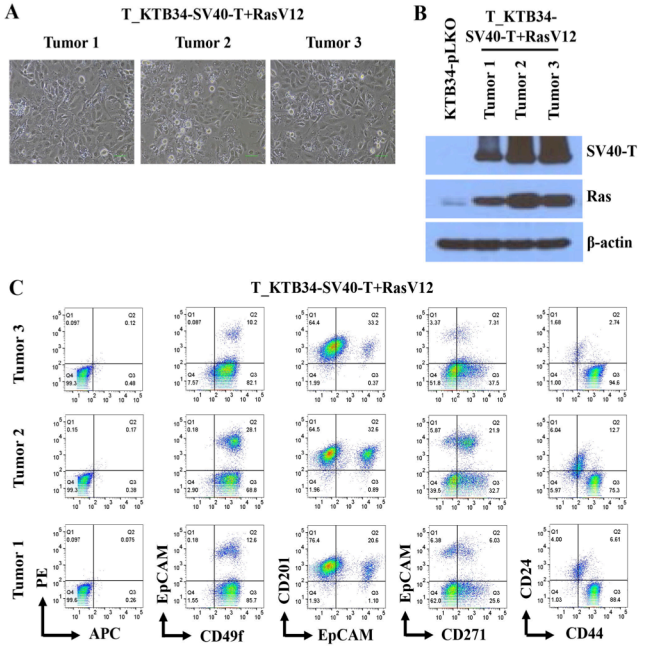


Fig. 34: (A) Phase contrast images showing the morphology of KTB34-SV40-T+HRasG12V tumor-derived cells. (B) Western blotting was used to detect the expression of SV40-T antigen and mutant Ras in KTB34-SV40-T+HRasG12V tumors. pLKO was used as a vehicle control. β–actin was used as an internal control. (C) KTB34-SV40-T+HRasG12V tumors were stained with indicated antibodies and flow cytometry was used to identify subpopulation of cells.

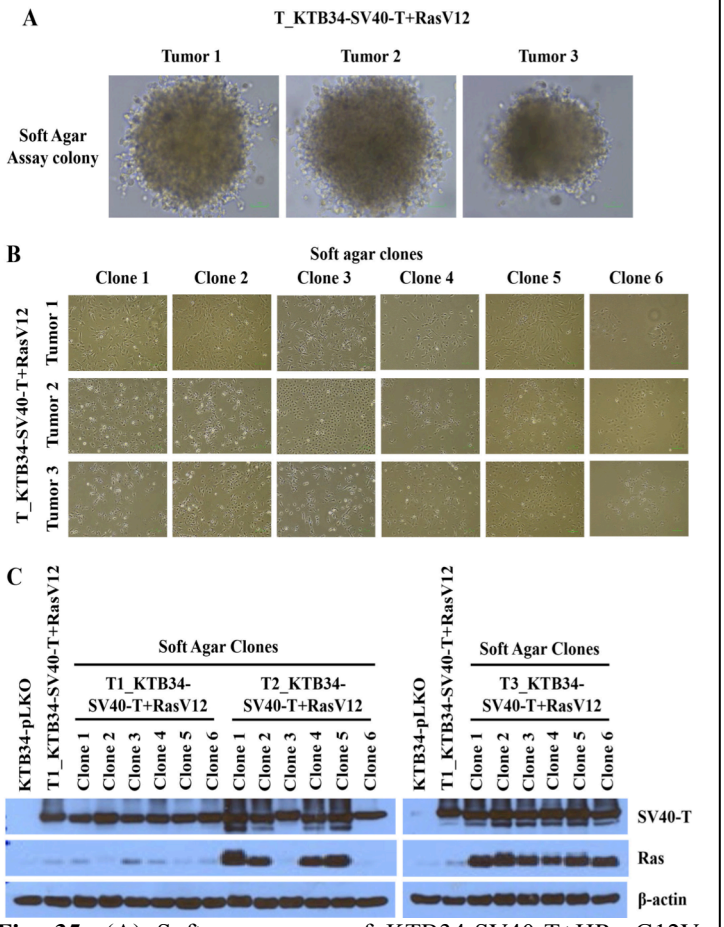


Fig. 35: (A) Soft agar assay of KTB34-SV40-T+HRasG12V tumor-derived cells. (B) Phase contrast images representing the morphology of cells derived from soft agar clones of KTB34-SV40-T+HRasG12V tumors. (C) Western blotting was used to detect the expression of SV40-T antigen and mutant Ras in cells derived from soft agar clones of KTB34-SV40-T+HRasG12V tumors. pLKO was used as a control. β–actin was used as an internal control.

KTB6-SV40T+HRasG12V transformed cells (Caucasian origin, PAM50 characterization not done): Cell lines established from tumor of KTB6 transformed cells with SV40-T antigen and HRasGV12 showed similarity to KTB34-derived cell lines **(Fig. 36A)**. Tumor cells were sorted using JAM-1/CD321 and EpCAM antibodies to separate the epithelial (JAM-1+/EpCAM+) and nonepithelial (JAM-1-/EpCAM-) subpopulations of cells and KTB6-pLKO cells were used to represent the control epithelial cells **(Fig. 36B)**. We examined the expression of SV40-T antigen and Ras proteins in tumor derived JAM-1+/EpCAM+ and JAM-1-/EpCAM- subpopulations of cells using western blotting and compared them with parental pLKO control cells. Both subpopulations of cells showed the higher expressions of SV40-T antigen and Ras **(Fig. 36C)**. We further characterized the tumor cells by flow cytometry using various markers (CD49f, CD201, CD271, CD44, CD24 and EpCAM) to determine the enriched subpopulation of cells after transformation. Staining of double transformed KTB6 tumor cells showed two distinct subpopulations with each of the maker combinations: CD49f$^{+/low}$/EpCAM- and CD49f+/EpCAM+; CD44+/CD24- and CD44+/CD24+; CD201+/EpCAM- and CD201$^{+/low}$/EpCAM+; and CD271^{low}/EpCAM- and CD271^{low}/EpCAM+ **(Fig. 36D)**. We further analyzed the JAM-1+/EpCAM+ and JAM-1-/EpCAM- subpopulations of cells. JAM-1+/EpCAM+ cells showed predominant CD49f+/EpCAM+, CD44+/CD24+/low, CD201+/EpCAM+, CD201-/EpCAM+ and CD271$^{+/low}$/EpCAM+ subpopulations, whereas JAM-1-/EpCAM- cells showed predominant CD49f$^{+/low}$/EpCAM-, CD44+/CD24-,

CD201+/EpCAM- and CD271$^{-/low}$/EpCAM- **(Fig. 36E)**. Thus, it appears that JAM-1-/EpCAM- cells are enriched for cancer cells with stem cell marker profiles.

We used soft agar assay to further confirm the transformed phenotype of tumor cells derived from SV40-T antigen and HRasGV12 overexpression. As expected, KTB6-SV40T+ HRasG12V transformed tumor cells formed colonies **(Fig. 37A)**. We next established 2D

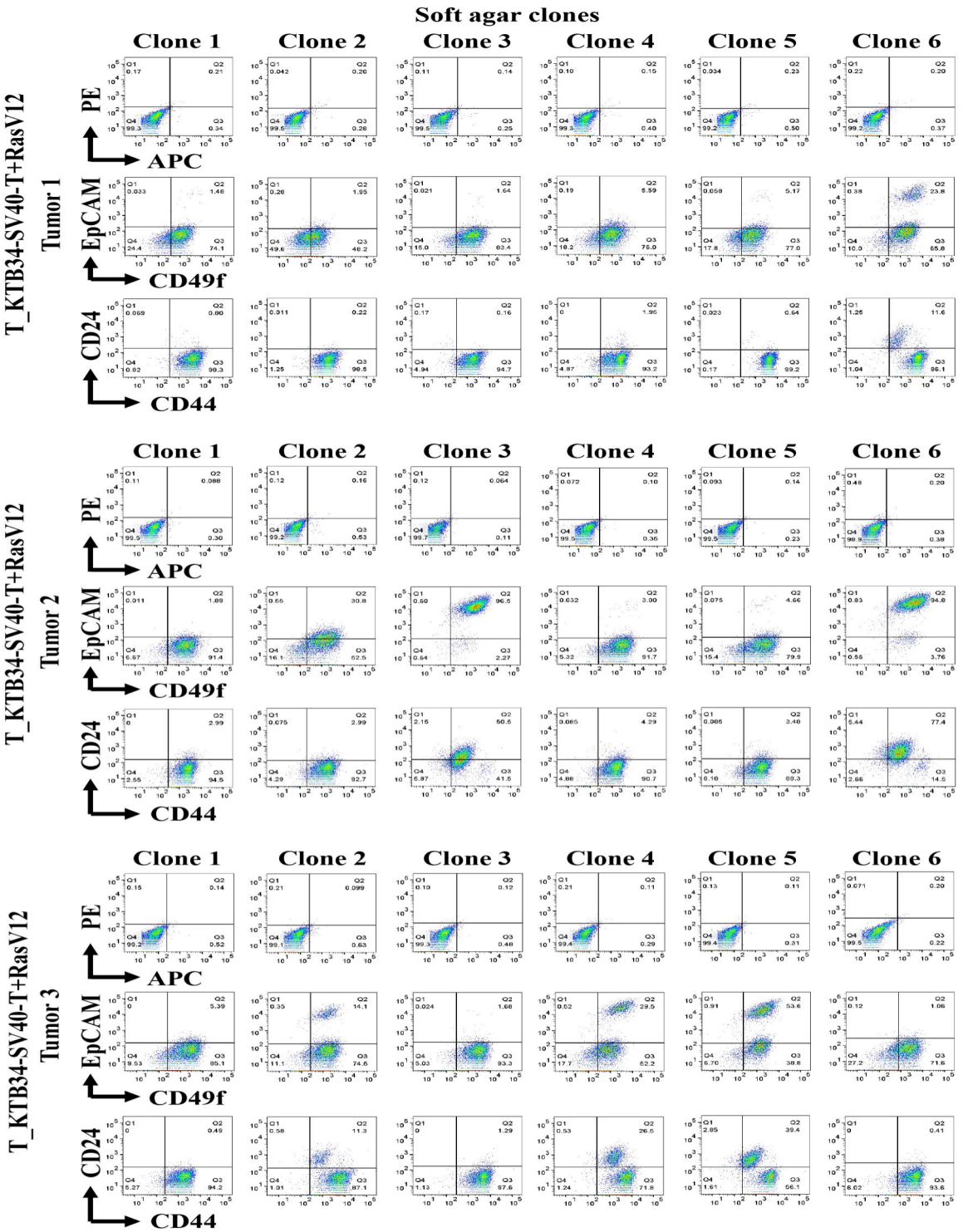


Fig. 36: Cells derived from soft agar clones of KTB34-SV40-T+HRasG12V tumors were stained with indicated antibodies and flow cytometry was used to identify subpopulation of cells.

cultures of these soft agar clones by transferring soft agar colonies to tissue culture plates. Tumor cells derived from soft agar showed morphological changes with aggressive phenotype compared to parental cells **(Fig. 37B)**. Western blotting was used to identify the expression of SV40-T antigen and mutant Ras in these cells. All clones from KTB6-SV40T+HRasGV12 tumor cells showed expression of SV40-T antigen and mutant Ras at variable levels **(Fig. 37C)**. To further characterize these clones for stem/progenitor/mature cell phenotypes, we performed flow cytometry with CD49f, EpCAM, CD44 and CD24 surface markers. All the clones showed predominant CD49f$^{+/low}$ /EpCAM- and CD44+/CD24- subpopulation of cells **(Fig. 37D)**. In general, it appears that

soft agar assay selects for transformed cells that are enriched for stem cell markers such as CD49f, CD201, and CD44 but lower levels of EpCAM. Only six out of 27 colonies examined contained a fraction of cells that expressed EpCAM.

KTB39-SV40T+HRasG12V transformed cells (African American Origin, normal-like cell line): We established cell lines from tumors of KTB39 transformed cells with SV40-T antigen and mutant HRasG12V and examined the expression of SV40-T antigen and Ras proteins using western blotting. These tumor derived cells expressed SV40-T antigen and Ras at variable levels **(Fig. 38A)**. We further characterized these tumor cells by flow cytometry using various markers (CD49f, CD201, CD271, CD44, CD24 and EpCAM) to determine the enriched subpopulation of cells after transformation. Staining of double transformed KTB39 tumor cells showed two distinct subpopulations of CD49f$^{+/low}$/EpCAM- and CD49f+/EpCAM+; CD44+/CD24- and CD44+/CD24+; CD201$^{+/low}$/EpCAM- and CD201$^{+/low}$/EpCAM+; and CD271$^{+/low}$/EpCAM- and CD271$^{+/low}$/EpCAM+ **(Fig. 38B)**. Two tumors showed differences in staining pattern and their staining pattern differed from the same cells maintained *in vitro* **(See Fig. 22)**. These results once again suggest that specific cell types are selected for growth *in vivo*. In soft agar assay, KTB39-SV40T+HRasG12V tumor cells formed smaller colonies **(Fig. 38C)** compared to KTB34-SV40T+HRasG12V and KTB6-SV40T+HRasG12V tumor cells.

KTB22-SV40T+HRasG12V transformed cells (basal-like cell line from Hispanic women): Cell lines established from tumors of KTB22 transformed cells with SV40-T antigen and HRasG12V showed distinct epithelial phenotypes **(Fig. 39A)**. We further characterized these tumor cells by flow cytometry using various markers (CD49f, CD201, CD271, CD44, CD24 and EpCAM) to determine the enriched subpopulation of cells after transformation. Staining of double transformed KTB22 tumor cells showed CD49f-/EpCAM+ and CD49f+/EpCAM+; CD44+/CD24+; CD201-/EpCAM+ and CD201+/EpCAM+; and CD271-/EpCAM+ and CD271+/EpCAM+ cells **(Fig. 39B)**. Four tumors showed differences in staining pattern and their staining pattern differed from the same cells maintained *in vitro* **(See Fig. 22)**. These results once again suggest that specific cell types are selected for growth *in vivo*.

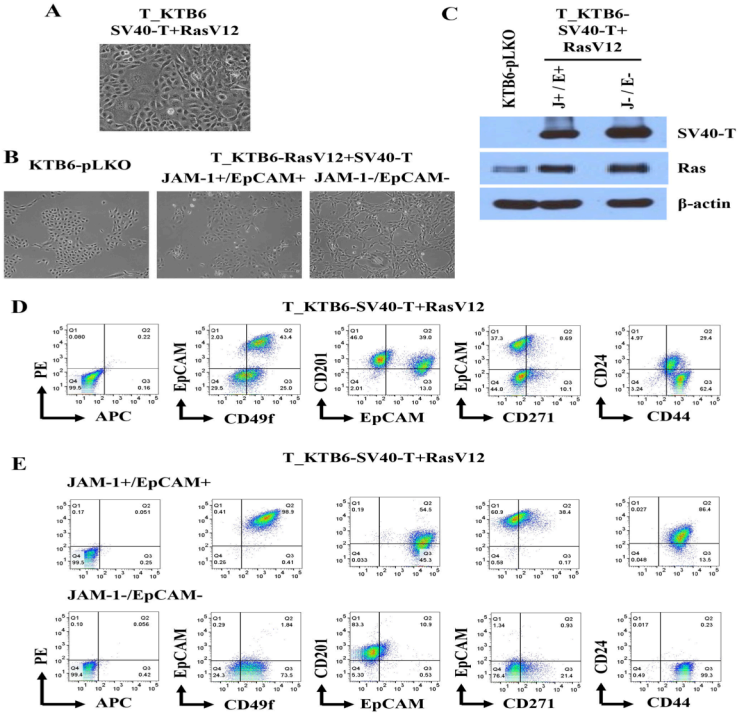


Fig. 36: (A) Phase contrast images showing the morphology of KTB6-SV40-T+HRasG12V tumor-derived cells. (B) Western blotting was used to detect the expression of SV40-T antigen and mutant Ras in KTB6-SV40-T+HRasG12V tumor. pLKO was used as a control. β–actin was used as an internal control. (C) Phase contrast images showing the morphology of JAM-1+/EpCAM+ and JAM-1-/EpCAM- subpopulations of KTB6-SV40-T+HRasG12V tumor. (D) KTB6-SV40-T+HRasG12V tumor-derived cells were stained with indicated antibodies and flow cytometry was used to identify subpopulation of cells. (E) JAM-1+/EpCAM+ and JAM-1-/EpCAM- subpopulations of KTB6-SV40-T+HRasG12V tumor cells were stained with indicated antibodies and flow cytometry was used to identify subpopulation of cells.

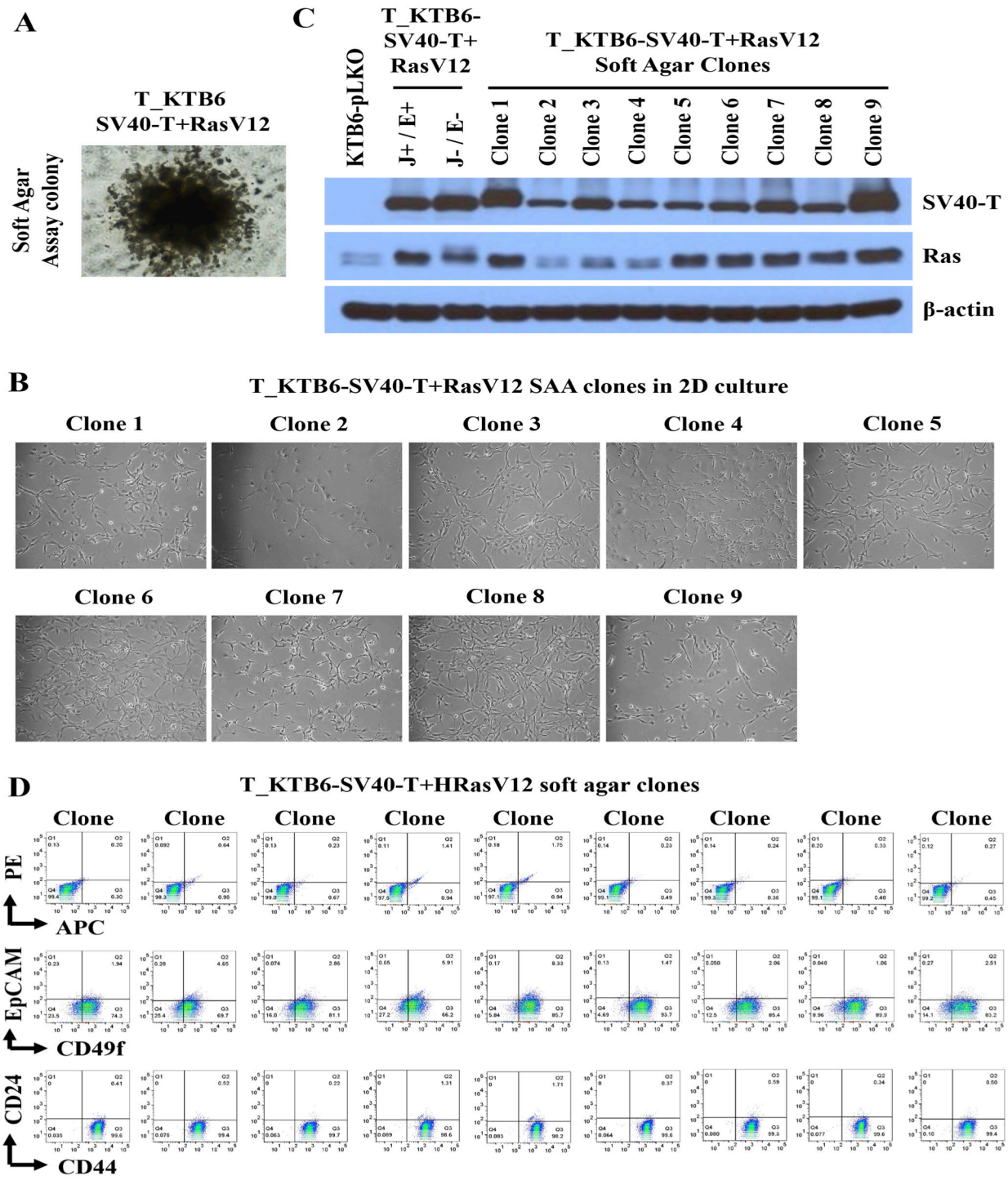


Fig. 37: (A) Soft agar assay of KTB6-SV40-T+HRasG12V tumor cells. (B) Phase contrast images showing the morphology of cells derived from soft agar clones of KTB6-SV40-T+HRasG12V tumors. (C) Western blotting was used to detect the expression of SV40-T antigen and mutant Ras in cells derived from soft agar clones of KTB6-SV40-T+HRasG12V tumors. pLKO was used as a vehicle control. β–actin was used as an internal control. (D) Cells derived from soft agar clones of KTB6-SV40-T+HRasG12V tumors were stained with indicated antibodies and flow cytometry was used to identify subpopulation of cells.

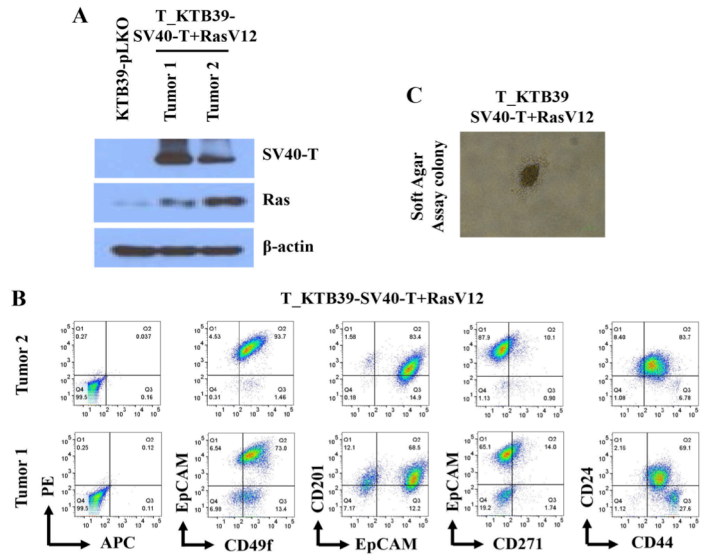


Fig. 38: (A) Western blotting was used to detect the expression of SV40-T antigen and mutant Ras in KTB39-SV40-T+HRasG12V tumors. pLKO was used as a control. β–actin was used as an internal control. (B) KTB39-SV40-T+HRasG12V tumors were stained with indicated antibodies and flow cytometry was used to identify subpopulation of cells. (C) Soft agar assay of KTB39-SV40-T+HRasG12V tumor cells.

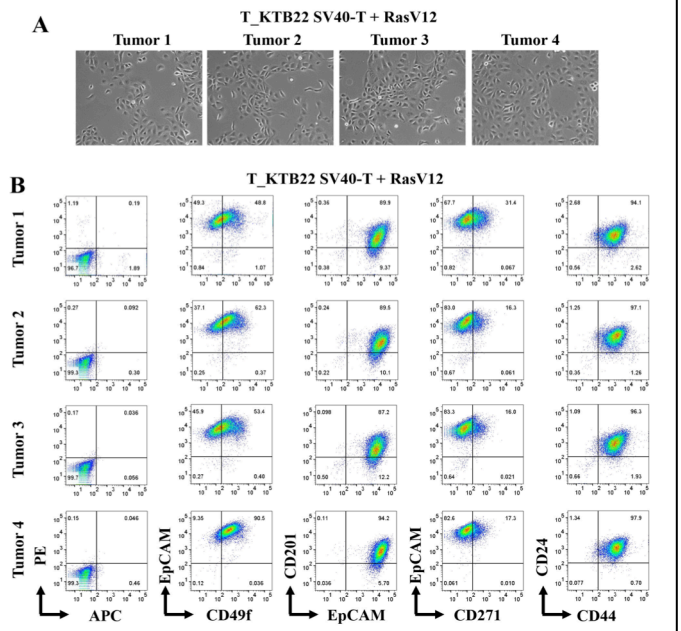


Fig. 39: (A) Phase contrast images showing the morphology of KTB22-SV40-T+HRasG12V tumor cell lines. (B) KTB22-SV40-T+HRasG12V tumor cells were stained with indicated antibodies and flow cytometry was used to identify subpopulation of cells.

HRasG12V transformed KTB22 cells: Ras transformed KTB22 developed tumor and cell line generated from the tumor showed senescence-like flat and enlarged cell morphology **(Fig. 40A).** We examined the expression of mutant Ras protein using western blotting and this tumor derived cells showed slightly higher expressions of Ras **(Fig. 40B)**. We further characterized these tumor cells by flow cytometry using various markers (CD49f, CD201, CD271, CD44, CD24 and EpCAM) to determine the enriched subpopulation of cells after transformation **(Fig. 40C)**. The majority of cells were CD49f$^{+/low}$/EpCAM- and CD44+/CD24-. In soft agar assay, KTB22-HRasG12V tumor cells formed colonies **(Fig. 40D)**.

HRasG12V transformed KTB42 cells: Cell line was generated from tumor developed from mutant Ras transformed KTB42 cells and Ras expression was confirmed by Western blotting **(Fig. 40E)**. Tumor derived cell line is enriched for stem cell markers such as CD49f, CD44 and CD201 **(Fig. 40F)** and formed colonies on soft agar **(Fig. 40G)**.

KTB34-SV40T+PIK3CA^{H1047R} transformed cells (Caucasian origin; luminal A type cell line): Cell lines established from tumors of KTB34 transformed cells with SV40-T antigen and PIK3CA^{H1047R} showed distinct epithelial phenotypes and these tumor cells were characterized by flow cytometry using various markers (CD49f, CD201, CD271, CD44, CD24 and EpCAM) to determine the enriched subpopulation of cells after transformation. Tumor derived cells displayed CD49f+/EpCAM+ and CD49f-/EpCAM+; CD44+/CD24+; CD201+/EpCAM+ and CD201-/EpCAM+; and CD271-/EpCAM+ and CD271+/EpCAM+ phenotypes **(Fig. 41)**. Tumor cells derived from KTB34-SV40T+PIK3CA^{H1047R} transformed cells differed from the tumor cells derived

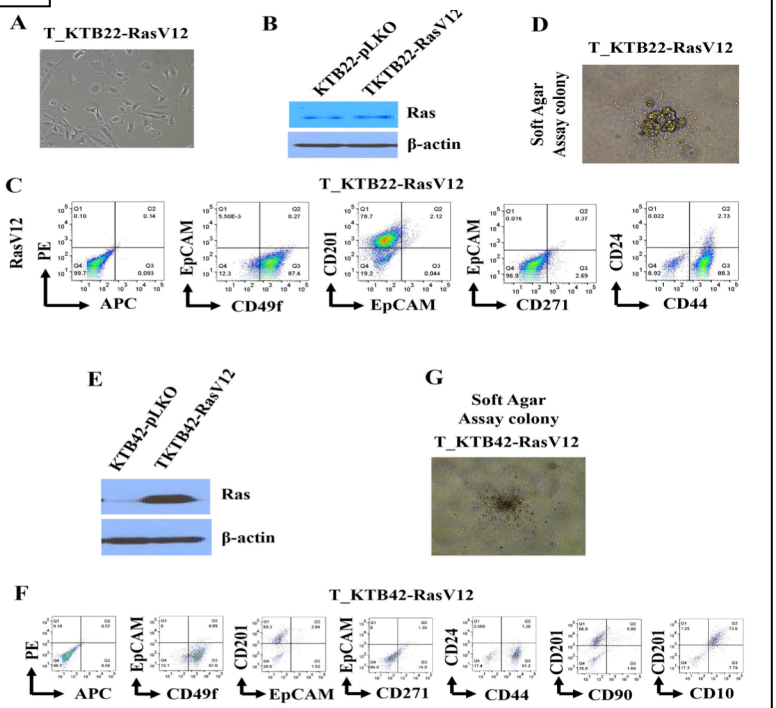


Fig. 40: (A) Phase contrast images showing the morphology of KTB22-HRasG12V tumor. (B) Western blotting was used to detect the expression of Ras in KTB22-HRasG12V tumor cells. pLKO was used as a control. β–actin was used as an internal control. (C) KTB22-HRasG12V tumor were stained with indicated antibodies and flow cytometry was used to identify subpopulation of cells. (D) Soft agar assay of KTB39-SV40-T+HRasG12V tumor-derived cells. (E) Western blotting was used to detect the expression of Ras in KTB42-HRasG12V tumor cells. pLKO was used as a control. β–actin was used as an internal control. (F) KTB22-HRasG12V tumor were stained with indicated antibodies and flow cytometry was used to identify subpopulation of cells. (G) Soft agar assay of KTB39-SV40-T+HRasG12V tumor cells.

from KTB34-SV40T+HRasG12V transformed cells in stem/progenitor/mature cell hierarchy **(See Fig. 34)**. These results suggest that while cell-of-origin determines tumor subtype (adenocarcinomas versus squamous carcinomas), oncogenes determine stem/progenitor/mature cell hierarchy.

Tumorigenicity of soft agar clones derived from KTB34-SV40T+HRasG12V tumor cells:
To determine whether soft agar clones (Prominent CD49f+/EpCAM- and CD49f+/EpCAM+ subpopulations) derived from KTB34-SV40-T+HRasG12V tumor cells form tumors, we injected $5x10^6$ cells with 50% matrigel in 100 microliter HBSS into the mammary fat pad/subcutaneous of 6-7 week old female NSG (NOD/SCID/IL2Rgnull) mice. Tumors were formed, resected and analyzed by H&E staining, which showed lung metastasis in both case **(Fig. 42)**.

Characterization of tumor and metastatic cells derived from tumor and lung metastasis: Tumor and metastatic cells created from the KTB34-SV40-T+HRasG12V and KTB6-SV40-T+HRasG12V tumor and lung were characterized by flow cytometry using various markers (CD49f, CD201, CD271, CD44, CD24 and EpCAM) to determine the enriched subpopulation of cells. KTB34-SV40-T+HRasG12V tumor and metastatic cells showed similar staining patterns **(Fig. 43A)**. Similar results were obtained with KTB6-SV40-T+HRasG12V tumor and metastatic cells **(Fig. 43B)**.

Single cell analyses of paired immortalized and transformed cells: Since the same luminal cell line transformed with the same oncogene produced four different tumor types, whereas basal and normal-like cell lines produced mostly squamous carcinomas, we hypothesized that different population of cells within a cell line are susceptible for transformation. To test this possibility, we

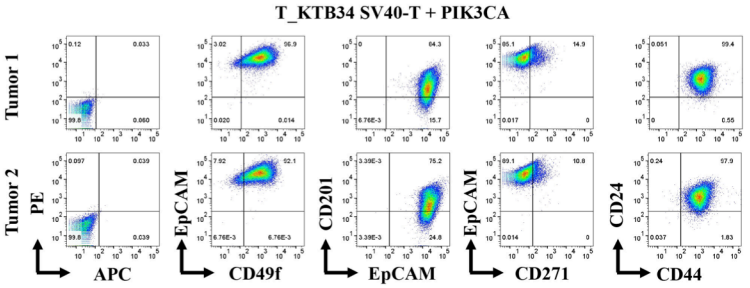


Fig. 41: KTB34-SV40-T+PIK3CA tumor-derived cells were stained with indicated antibodies and flow cytometry was used to identify subpopulation of cells.

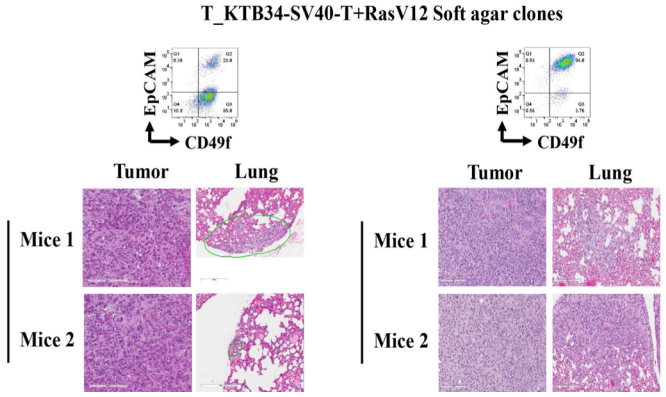


Fig. 42: Tumor developed from the soft agar clones of T_KTB34-SV40-T+HRasG12V tumor cells (Prominent CD49f+/EpCAM- and CD49f+/EpCAM+ subpopulations) metastasized to lungs. Representative immunohistochemistry (IHC) data are shown.

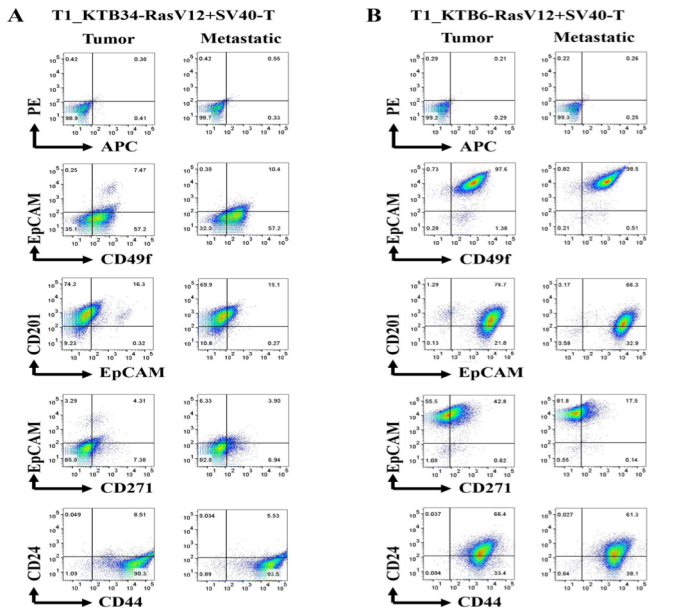


Fig. 43: Tumor and metastatic cells created from the KTB34-SV40-T+HRasG12V and KTB6-SV40-T+HRasG12V tumor and lungs were characterized by staining with indicated antibodies and flow cytometry was used to identify subpopulation of cells.

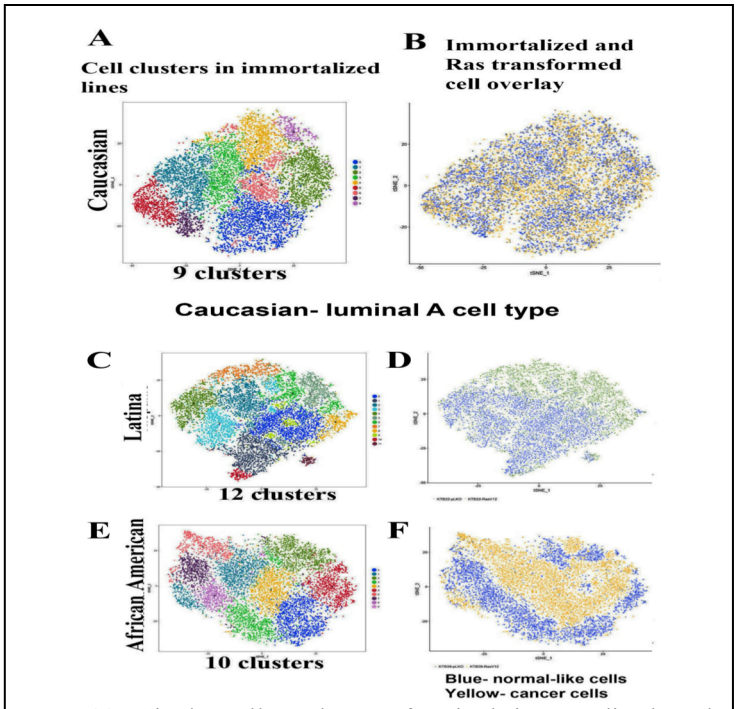


Fig. 44: Single cell analyses of paired immortalized and transformed cells showed remarkable variability in transformation-prone cell population between cell lines.

performed single cell analyses of paired immortalized and transformed cells. We used only HRasG12V transformed cells to avoid any ambiguity due to differential expression of Ras and SV40-T antigen. This paired single cell analyses of immortalized and transformed cells showed remarkable variability in transformation-prone cell population between cell lines. While almost all the nine subpopulations of cells in the luminal A cell line were prone for transformation, select subpopulation of cells within basal (4 out of 12) and normal-like (5 out of 10) immortalized cell lines were prone for transformation (**Fig. 44)**.

DNA damage response of transformed and tumor derived cells: To begin to investigate the role of cell-of-origin and oncogenes on response to therapy, we determined the double stranded DNA break (DSB) repair capacity of immortalized and transformed cells upon treatment with chemotherapy. Flow cytometry was used to identify induction of an established marker for unrepaired DSBs γ-H2AX in response after 48 hours of treatment with doxorubicin, paclitaxel and cisplatin. The type of oncogene appears to influence DSB repair, as doxorubicin increased γ-H2AX levels more efficiently in SV40 T-antigen overexpressing cells compared to HRasG12V overexpressing cells **(Fig. 45A).** Similar effects were observed with variable response in case of paclitaxel treatment **(Fig. 45B)**.

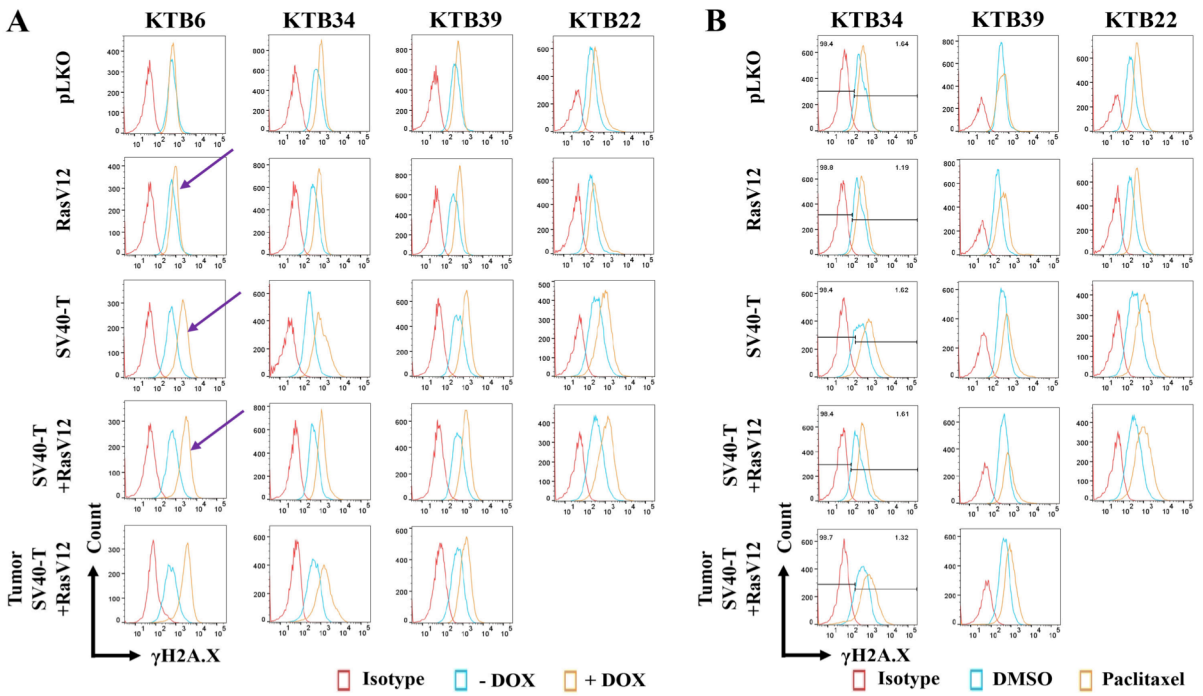


Fig. 45: DNA DSB repair of KTB6, KTB34, KTB39 and KTB22 transformed cells were examined after treatment with doxorubicin (A) and paclitaxel (B) by flow cytometry using pγ-H2AX.

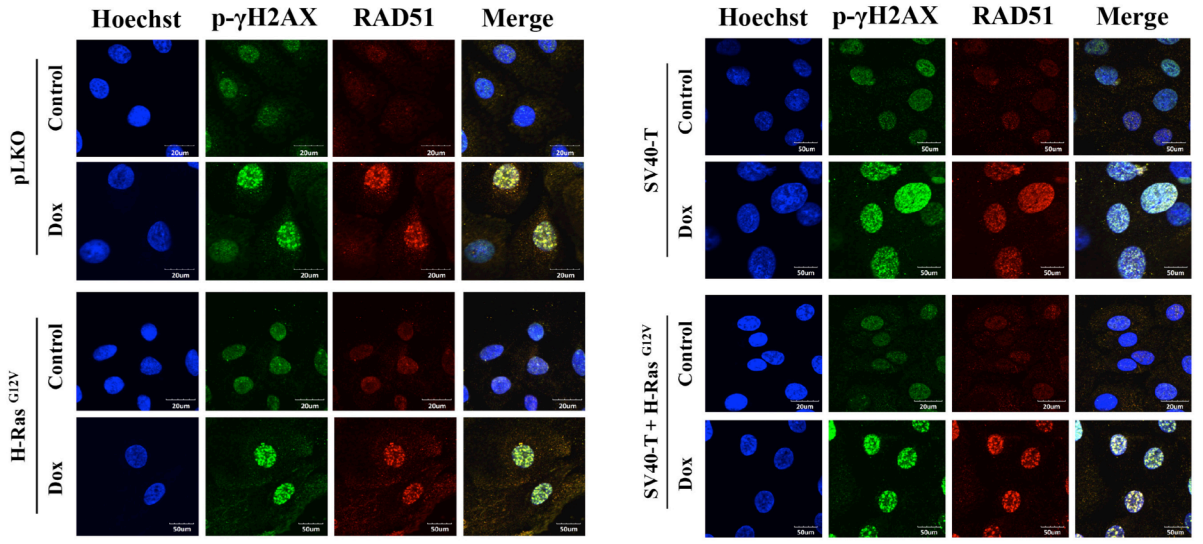


Fig. 46: SV40-T/t antigen induced signals disrupt DNA repair. Immunofluorescence staining with pγ-H2AX and RAD51 antibodies was used to measure response to chemotherapeutic drug doxorubicin.

Similar results were obtained with cisplatin (data not shown). Thus, while cell type determines the type of tumors, the oncogenes determine DNA repair capacity of transformed cells in response to chemotherapy-induced DNA damage.

To further investigate oncogene-dependent changes in double-strand DNA breaks (DSBs) repair pathways, we performed co-localization studies of γ-H2AX and Rad51. Co-localization of these two molecules indicates active repair of DSBs. Doxorubicin-treated pLKO and H-RasG12V transformed cells showed strong γ-H2AX and Rad51 co-localization and distinct foci, whereas signals in SV40-T antigen transformed cells were much more diffuse **(Fig. 46)**. Taken together, these results indicate that while DNA repair pathways are relatively unaffected in pLKO and HRasG12V-transformed cells, SV40-T/t antigen induced signals disrupt DNA repair.

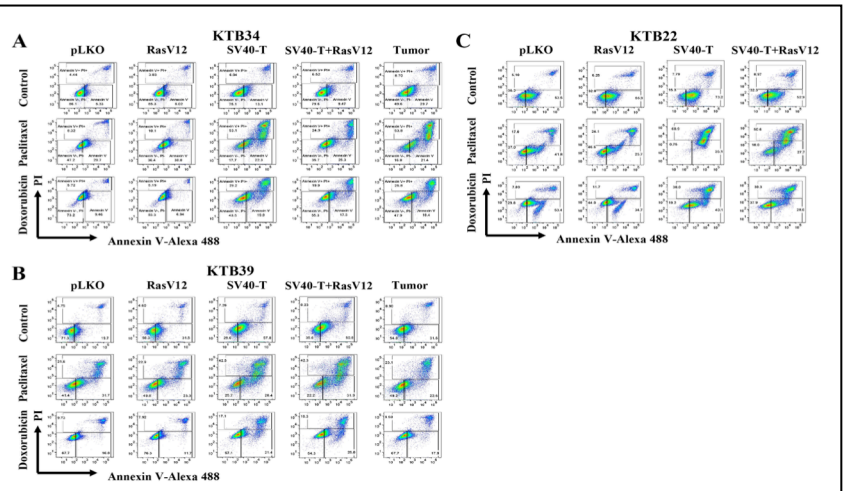


Fig. 47: Annexin V staining was used to measure response to chemotherapeutic drugs doxorubicin and paclitaxel in (A) KTB34 transformed cells, (B) KTB39 transformed cells and (C) and KTB22 transformed cells.

Therapy-induced apoptosis is oncogene-dependent: To elucidate the role of oncogenes in chemotherapy-induced apoptosis, annexin V staining was used to measure response to chemotherapeutic drug doxorubicin, paclitaxel, and cisplatin in KTB34, KTB39 and KTB22 transformed and tumor cells. SV40-T antigens transformed compared to immortalized or Ras-transformed cells were impaired in repairing doxorubicin-induced DNA damage, irrespective of cell-of-origin of transformed cells, and thus were more sensitive to doxorubicin and paclitaxel-induced apoptosis **(Fig. 47A, B, C)**.

Specific Aim 2: Establish that inter-individual heterogeneity within African Americans for unique cells determine susceptibility to breast cancer

ZEB1+ cells are enriched in the normal breasts of African Ancestry (AA) compared to European Ancestry (EA) women: Results presented below are published in two articles and a patent application has been filed [47, 48]. In the mouse mammary gland, PROCR+/EpCAM- cells are suggested to function as multi-potent stem cells [11]. In our previous study focused on evaluating ethnicity-dependent differences in the normal breast, we observed specific enrichment of PROCR+/EpCAM- cells in cultured normal breast epithelial cells from biopsies of healthy AA women compared to EA women [2]. These cells are enriched for the expression of stemness-related transcription factor ZEB1 and have enhanced Wnt pathway activity compared to PROCR±/EpCAM+ cells [2]. ZEB1 has recently been demonstrated to limit onco-suppressive p53-driven DNA damage response in stem cells and thus increase the stem cells, intrinsic susceptibility to malignant transformation [28]. ZEB1+ cells co-express the methionine sulfoxidase reductase (MSRB3), which protects against DNA damage [28]. These observations raised the possibility that PROCR+/ZEB1+ cells are naturally present at higher levels in the normal breasts of AA women and failure to consider natural variation in gene expression pattern, influenced at least partially by genetic ancestry, could have an impact on developing cancer-specific biomarkers. Measuring PROCR itself in the breast tissue is complicated because there are four haplotypes of PROCR due to SNPs and only one among them is a cell surface protein [49]. Since ZEB1 expression is enriched in PROCR+/EpCAM- cells, we used ZEB1 as a surrogate marker for PROCR+/EpCAM- cells in un-manipulated breast tissues.

Table 1: Differences in expression levels of ER, ZEB1 and FOXA1 in KTB-normal between women of African American and European ancestry

	African American				European Ancestry				Two-sided Wilcoxon Test p-value
Variable Name	N	Median	Minimum	Maximum	N	Median	Minimum	Maximum	
ER Positivity	38	0.009837	0.001769	0.085101	39	0.010387	0.000000	0.064659	0.8345
ER H Score	38	2.165487	0.315269	21.777591	38	2.265046	0.103586	17.165665	0.7514
ZEB1 Positivity	38	0.004324	0.000316	0.025044	41	0.001224	0.000221	0.028532	<0.0001**
ZEB H Score	38	0.931903	0.045633	5.926867	41	0.157922	0.026349	3.927299	<0.0001**
FOXA1 Positivity	42	0.037941	0.010844	0.147725	47	0.021856	0.007987	0.171964	0.0033**
FOXA1 H Score	42	5.108708	1.414022	20.066698	47	3.126083	1.033697	23.666012	0.0031**

A tissue microarray (TMA) comprising breast tissues of ancestry-mapped 100 age-matched healthy women from the Komen Tissue

Bank (KTB) at Indiana University and normal tissues adjacent to tumors (NATs)-tumors pairs from 100 women (300 samples total) was analyzed for the levels of ZEB1, an oncogenic transcription factor that is central to cell fate, mature luminal cell enriched estrogen receptor alpha (ERα) and FOXA1 and for immune cell composition.

Representative IHC staining patterns of ZEB1 in KTB-normal, NATs, and tumors from AA and EA women are shown in **Fig. 48A** and statistical analyses are presented in **Fig. 48B-D** and **in Table 1**. Highly discriminative 41-ancestry marker profiles of KTB-normal showed >75% African ancestry markers in samples from African American women and >80% European ancestry markers in Caucasian women **(Fig. 49)** [4]. ZEB1 expressing cells are localized outside the ductal structures of the normal breast and in the stromal part of the tumors. KTB-normal of AA women contained significantly higher levels of ZEB1-positive cells compared to KTB-normal breasts of EA women **(Fig. 48B).** NATs of AA women showed a modest increase in ZEB1+ cells compared to KTB-normal **(Fig. 48C and D).** Scenario was completely different in EA women; both NATs and tumors contained significantly higher levels of ZEB1+ cells compared to KTB-normal **(Fig. 48C and D).** NAT to tumor differences were noted only in EA women where an increase in ZEB1+ cells was noted predominantly in ERα+ tumors. Thus, ZEB1+ cells are intrinsically higher in the normal breast of AA women, whereas remarkably elevated ZEB1+ cells in the breasts of EA women are observed only in the context of breast cancer. Increase in ZEB1+ cells in KTB-normal of AA women compared to EA women is less likely related to BMI differences, as when analysis was done by subdividing women above and below BMI of 30, irrespective of genetic ancestry, ZEB1 H-score but not positivity showed a marginal relationship (p=0.04) to BMI above and below 30.

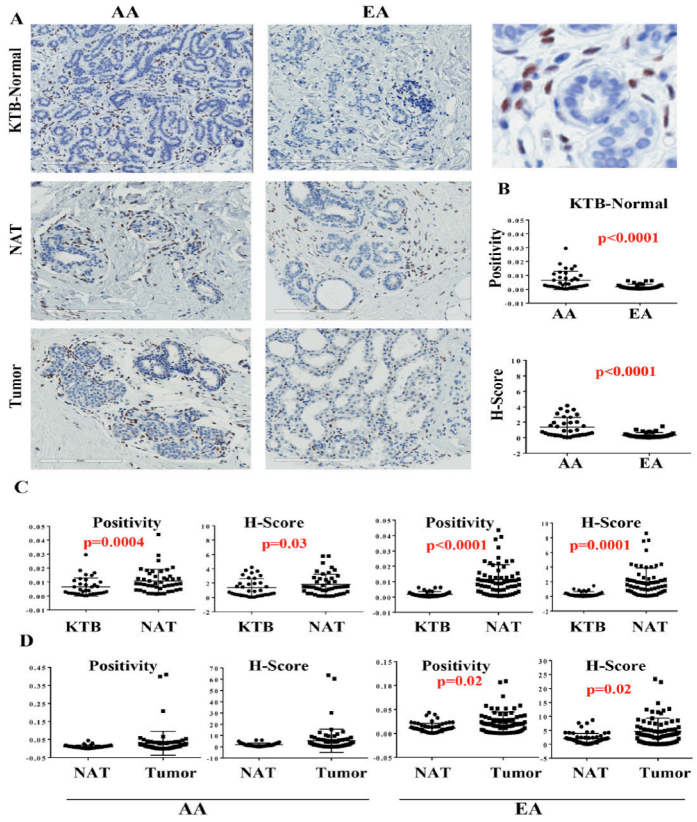


Fig. 48: ZEB1 expression pattern in KTB-normal, normal adjacent to tumor (NATs) and in breast tumors. A) Representative immunohistochemistry of KTB-normal, NATs and tumors of women of African American (AA) and European Ancestry (EA). Enlarged view of a KTB-normal is shown on right (top). B) Differences in ZEB1 expression (positivity and H-score) between KTB-normal of AA and EA women. C) Differences between KTB-normal and NATs in AA and EA women. D) Differences between NATs and tumors in AA and EA women.

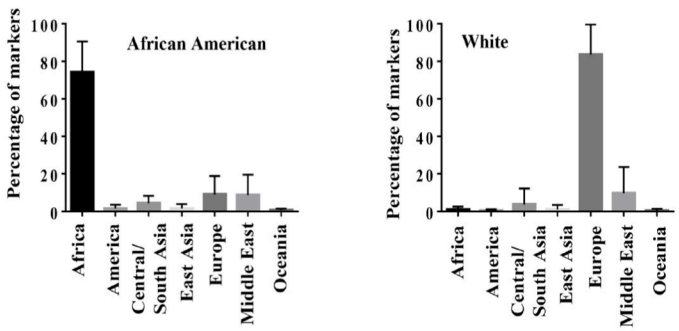


Fig. 49: Ancestry marker distribution patterns in self-reported African American and White women. Normal breast tissues from these women were used as KTB-normal

FOXA1 expression is lower in NATs of only AA women: FOXA1 serves as a pioneer factor that controls chromatin access of various nuclear receptors including ERα and controls the expression of genes enriched in luminal cells compared to basal cells [50-52]. FOXA1 along with another pioneer factor GATA3 and ERα form a lineage restricted hormone-responsive signaling network in the normal breast [7]. While higher expression of FOXA1 in the primary tumor is associated with better outcome, its overexpression in metastatic and/or anti-estrogen resistant tumors is associated with rewiring of ERα signaling and poor outcome [53-55]. In addition, it is suggested that FOXA1 gene is preferentially methylated in tumors of AA women [56]. Because of its relative importance in breast cancer, we assessed our TMA for FOXA1 expression. Representative staining pattern of FOXA1 is shown in **Fig. 50A** and numerical values are presented in **Table 1.** While FOXA1 levels in KTB-normal of AA women was modestly higher than in EA women, NATs of AA women had lower FOXA1 compared to KTB normal **(Fig. 50B).** Thus, tumors through their field effect may decrease FOXA1 in the surrounding breast tissues of AA women.

ERα+ expression is similar in the breasts of AA and EA women: ERα-positive cells in the normal breasts are considered to be highly differentiated non-proliferative cells and control proliferation of ERα-negative cells through paracrine mechanisms [57]. Representative ERα staining pattern is shown in **Fig. 51A** and statistical analyses are presented in **Fig. 51B-D** and **Table 1.** There were no differences between AA and EA women in ERα-positivity in KTB-normal and NATs. The results are not only relevant, but also reassure that our TMA detects only specific changes.

NATs of AA and EA women show differing levels of CD8, PD1 and PDL1+ cells: Results thus far point to the pro-inflammatory state of NATs of EA based on known link between ZEB1 and inflammatory cytokines [58]. To address this further, we stained the above TMAs with CD8 for T cells, CD68 for macrophages, and PD1 for immune cells. We also examined epithelial/tumor cells for PDL1. All staining was done in a CLIA-certified lab with FDA-approved antibodies. In KTB-normal TMAs, there was no staining with CD8 and CD68 in either the AA or EA TMAs. There were less than 1% of the lymphocytes and macrophages stained and these were considered negative. The same negativity was obtained with PD1 and PDL1 immunostains (data not shown). Therefore, we analyzed staining results between NATs of AA and EA women and between NATs and tumors. Representative staining patterns in NATs and tumors are shown in **Fig. 52.** CD8 immunostaining was localized to the inflammatory cells (T lymphocytes) and not in the tumor cells in the breast cancer cores. No background reactivity was observed in all cases. NATs of AA women showed statistically significantly higher CD8 positivity compared to EA women **(Fig. 53).** The tumors in EA had more CD8 immunostaining compared to corresponding NATs but such differences were not seen in the AA women. CD68 staining was localized to macrophages in the breast cancer cores **(Fig. 52).** CD68 had lower positivity compared to CD8 by both visual and the Aperio positive pixel reads. CD68 positivity was higher in tumors compared to their NATs (p=0.02) in EA but no such differences were noted in AA women **(Fig. 53).** PD1 immunostaining was localized to immune cells only and no background staining was observed **(Fig. 52).** There was no staining of tumor cells. NATs of AA women contained significantly higher PD1+ cells, similar to CD8+ cells, compared to NATs of EA women **(Fig. 54**). PD1 staining did not show any differences between NATs and tumors in both groups. PDL1 immunostaining was seen localized in the tumor cell cytoplasm and cell membrane **(Fig. 52).** In a few EA cases, only lymphocytes were stained. PDL1 staining of NATs of AA was significantly lower than EA cases **(Fig. 54).** Although PDL1 staining did not differ between NATs and tumors in case of AA women, its levels were marginally lower in ER+ tumors but not ER- tumors compared to NATs in case of EA women. It is interesting that PD1 and PDL1 staining scores in NATs of AA is reverse of the patterns in EA women. In summation, immune environment in NATs is different from that in KTB-normal with further differences between NATs and tumors, which showed variations based on genetic ancestry.

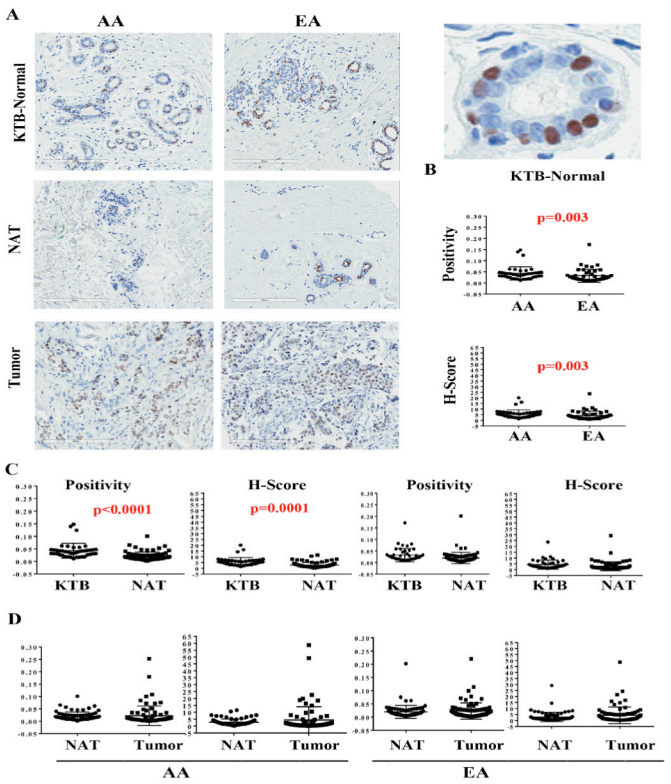


Fig. 50: FOXA1 expression pattern in KTB-normal, normal adjacent to tumor (NATs) and in breast tumors. A) Representative immunohistochemistry of KTB-normal, NATs and tumors of AA and EA women. Enlarged view of a KTB-normal is shown on right (top). B) Differences in FOXA1 expression (positivity and H-score) between KTB-normal of AA and EA women. C) Differences between KTB-normal and NATs in AA and EA women. D) Differences between NATs and tumors in AA and EA women.

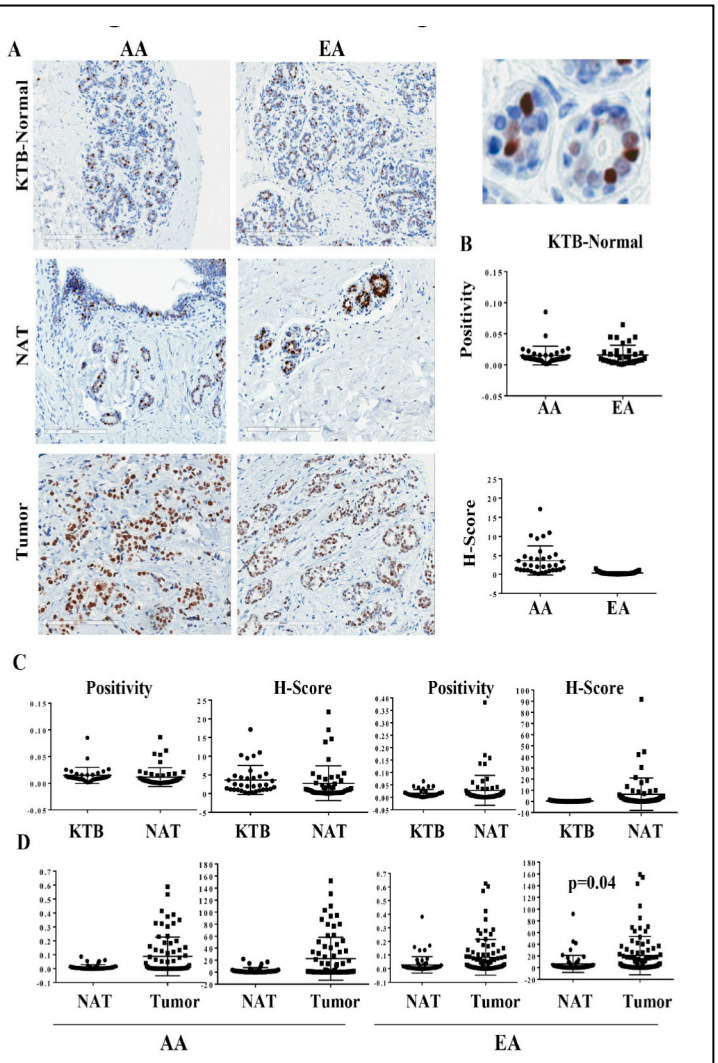


Fig. 51: ERα expression pattern in KTB-normal, normal adjacent to tumor (NATs) and in breast tumors. A) Representative immunohistochemistry of KTB-normal, NATs and tumors of AA and EA women. Enlarged view of a KTB-normal is shown on right (top). B) Differences in ERα expression (positivity and H-score) between KTB-normal of AA and EA women. C) Differences between KTB-normal and NATs in AA and EA women. D) Differences between NATs and tumors in AA and EA women.

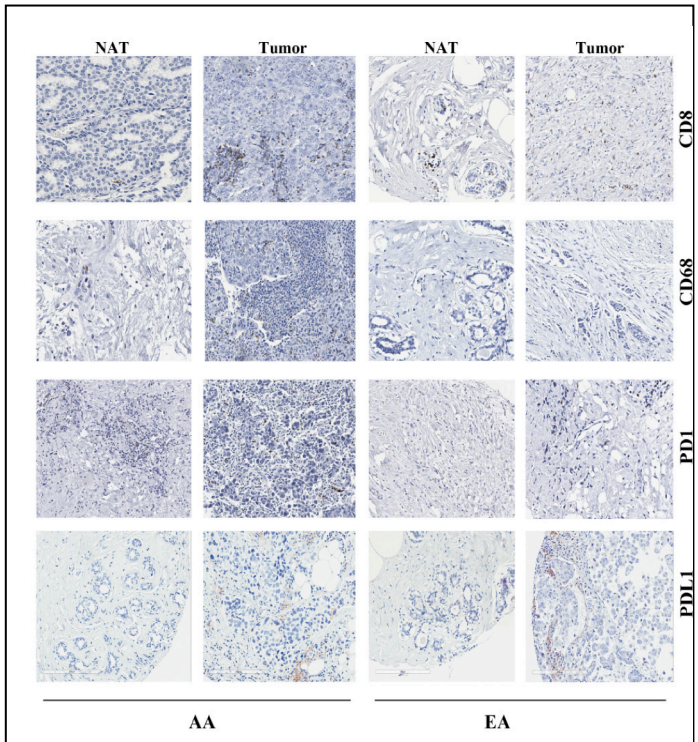


Fig. 52: Representative CD8, CD68, PD1 and PDL1 immunohistochemistry of NATs and Tumors of AA and EA women.

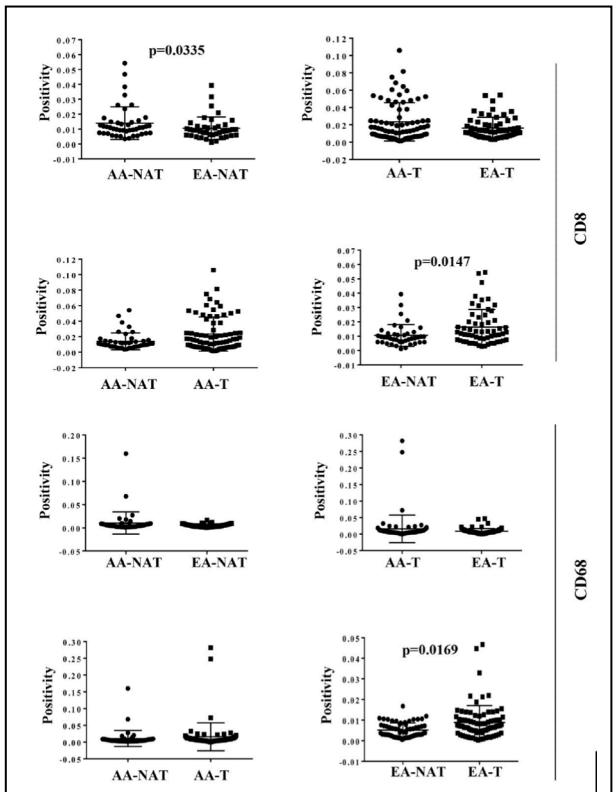


Fig. 53: Statistical analyses of CD8 and CD68 positivity in NATs and tumors (T) of AA and EA women. All statistically significant differences are indicated with *p* values.

4. Impact:

Principal discipline: Impact of our studies are two folds. One is on basic understanding of the normal breast biology and sharing of resources created with the breast cancer research community. The second is a potential early biomarker of breast cancer initiation in Caucasian women.

Normal breast biology: We have shown that ZEB1+ cells are enriched in the normal breasts of African American women compared to Caucasian women. We created immortalized breast epithelial cell lines from women of different genetic ancestry and showed that these cell lines are normal counterparts of intrinsic subtypes of breast cancer. When transformed, these cell lines generate adenocarcinomas. Additional studies of these cell lines further revealed that while tumor type and metastasis patterns are hardwired in the cell type from which tumor originates (cell-of-origin), oncogenic aberrations determine sensitivity to chemotherapy. Thus, developing methods to identify cell-of-origin of a tumor may help to develop predictive markers of breast cancer metastasis in future. Because of unique features of cell lines we have created (enriched for luminal gene expression compared established cell lines in the literature, which have basal gene expression patterns), these cell lines have been distributed to more than 15 labs both within and outside USA.

Early biomarkers of breast cancer initiation in Caucasian women: Our studies showed that while the normal breasts of African American women contain elevated number of ZEB1+ cells, the number of ZEB1+ cells increased only in the breasts with cancer in case of Caucasian women. Thus, an increase in ZEB1+ cells in the breasts of Caucasian women would indicate the presence of cancer. ZEB1-positivity could be used as a biomarker to complement radiologic techniques for early detection of breast cancer in Caucasian women. Based on additional characteristics, we have renamed ZEB1+ cells as PZP cells (PROCR+/ZEB1+/PDGFRalpha+). These cells can transdifferentiate into multiple subtypes including epithelial cells, fibroblasts, and adipocytes. We are seeking funding from several agencies to characterize these cells as we believe that PZP cells have the potential to alter the tumor microenvironment and promote breast cancer growth. These cells can also be targeted through clinically used agents such as nilotinib, which inhibits PDGFRalpha activity. Thus, findings from this study may lead to biomarker-driven clinical trials in future. We also observed elevated number of hormonally responsive GATA3+ cells in the normal breasts of Caucasian women compared to African American women. These results can explain for increased incidence of hormone receptor positive breast cancers in Caucasian women. Our studies also revealed that breast tissues adjacent to tumors are molecularly abnormal and can be used to develop biomarkers of tumor initiation, progression and metastasis.

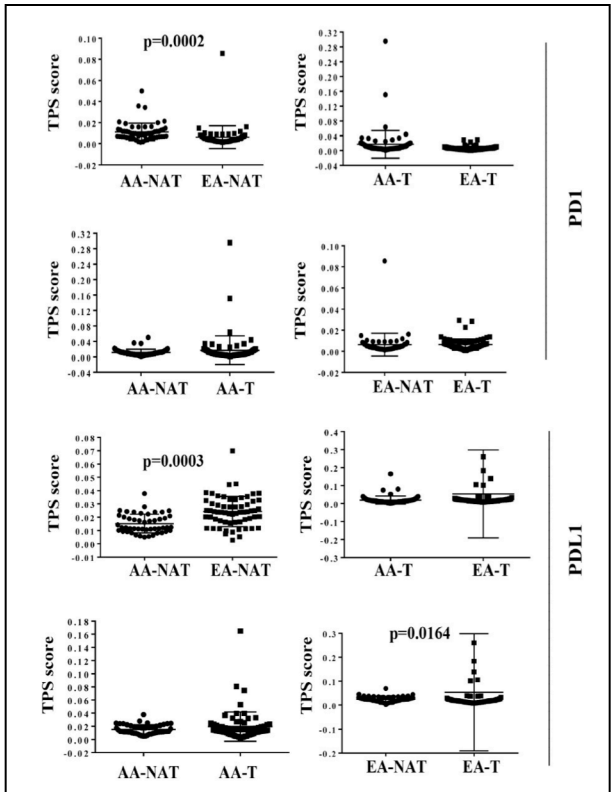


Fig. 54: Statistical analyses of PD1 and PDL1 TPS scores in NATs and tumors (T) of AA and EA women. All statistically significant differences are indicated with *p* values.

Impact on other disciplines: Our studies have revealed the need to consider genetic ancestry in developing biomarker. Although our results are from breast cancer models, the same may apply to other cancers. Second, tissues adjacent to tumors are often used as normal controls. Our results clearly indicate tumor-induced field defects in these tissues. Therefore, normal control tissues need to be derived from healthy individuals.

Impact on technology transfer: Breast epithelial cell lines we have created are unique compared to breast epithelial cell lines described in the literature as these cell lines express luminal-enriched gene expression. Breast cancers typically originate from luminal epithelial cells. Because of these properties, investigators from several institutions have requested and received our cell lines.

Impact on society beyond science and technology: Nothing to report

5. Changes/Problems: Nothing to report

6. Products:

Publications, conference papers, and presentations:

Publications:

1) Kumar B, Prasad MS, Bhat-Nakshatri P, Anjanappa M, Kalra M, Marino N, Storniolo AM, Rao X, Liu S, Wan J, Liu Y and **Nakshatri H (2018).** Normal breast-derived epithelial cells with luminal and intrinsic subtype-enriched gene expression document inter-individual differences in their differentiation cascade. **Cancer Research 78:5107-5123**

2) **Nakshatri H,** Kumar B, Burney H, Cox ML, Jacobsen M, Sandusky G, D'Souza-Schorey C and Storniolo AM **(2019).** Genetic ancestry-dependent differences in breast cancer-induced field defects in the tumor-adjacent normal breast. **Clinical Cancer Research 25:2848-2859**

3) Prasad M, Kumar B, Bhat-Nakshatri P, Anjanappa M, Sandusky G, Miller KD, Storniolo AM and **Nakshatri H. (2019).** TGFβ/BMP pathway inhibition enables expansion and characterization of multiple epithelial cell types of the normal and cancerous breast. **Molecular Cancer Research 17:1556-1570**

Conference papers and presentations:

1. San Antonio Breast Cancer Symposium, San Antonio, December 10-14, 2019
2. Hallmark of Cancer Symposium (Cell Press), Seattle, November 17-19, 2019
3. San Antonio Breast Cancer Symposium, December 4-8, 2018
4. Annual Amelia Project Conference on July 14, 2018 at Indiana University, Kokomo Campus, Indiana.
5. Indiana University Melvin and Bren Simon Cancer Center's Cancer Research Day on May 17, 2018 at IUPUI Campus Center, Indianapolis.
6. AACR Travel Award presentation on May 31, 2018 at Indiana University.
7. AACR Annual Meeting on April 14-18, 2018 at Chicago, IL.
8. Indiana University School of Medicine Postdoc Symposium on October 6, 2017.
9. Biochemistry and Molecular Biology Research Day on October 5, 2017 at IUPUI Campus Center, Indianapolis.
10. Indiana University Melvin and Bren Simon Cancer Center's Cancer Research Day on May 17, 2017 at IUPUI Campus Center, Indianapolis.
11. Annual Amelia Project Conference on March 4, 2017 at Indiana University, Kokomo Campus, Indiana.

Website(s) or other Internet site(s): Nothing to Report
Technologies or techniques: Nothing to Report
Inventions, patent applications, and/or licenses: IU has filled U.S. Provisional Patent Application Nos. 62//773,457, filed November 30, 2018, and 62/795218, filed January 22, 2019. Title: CANCER DETECTION UTILIZING NORMAL TISSUE ADJACENT TO BREAST TUMORS WITH GENETIC ANCESTRY-MAPPING

Other Products: Breast epithelial cell lines derived from the normal breasts. Both immortalized and transformed variants are available.

7. Participants & Other Collaborating Organizations

(1).

Name:	**Prof. Harikrishna Nakshatri**
Project Role:	**PI**
Researcher Identifier (University ID Number):	**20200001165457**
Nearest person month worked:	**48**
Contribution to Project:	**Prepared primary cells, xenograft studies and supervised the entire project and wrote the report with input from Drs. Kumar and Prasad**
Funding Support:	**Partially by W81XWH-15-1-0707**

(2).

Name:	**Brijesh Kumar**
Project Role:	**Post-Doctoral Fellow**
Researcher Identifier (University ID Number):	**20200095058465**
Nearest person month worked:	**40**
Contribution to Project:	**Dr. Kumar has performed all cell lines immortalization, characterization, flow cytometry, mammosphere assays, qRT-PCRs, karyotyping, cell based assay, immunofluorescence, western blotting, transformation of cell lines with oncogenes and xenograft study**
Funding Support:	**W81XWH-15-1-0707**

(3)

Name:	**Mayuri Prasad**
Project Role:	**Post-Doctoral Fellow**
Researcher Identifier (University ID Number):	**20200095026778**
Nearest person month worked:	**30**
Contribution to Project:	**Dr. Prasad standardized hydrogel assay, immunofluorescence, qRT-PCRs and characterized cell lines for luminal and basal features.**
Funding Support:	**W81XWH-15-1-0707**

(3). Collaborating Organizations: Nothing to Report

8. Special Reporting Requirements

Collaborative Awards: Nothing to Report

Quad Charts: Nothing to Report

9. Appendices: Three publications resulting from the award and the patent application.

References:

1. Tata PR, Mou H, Pardo-Saganta A, Zhao R, Prabhu M, Law BM, Vinarsky V, Cho JL, Breton S, Sahay A *et al*: **Dedifferentiation of committed epithelial cells into stem cells in vivo**. *Nature* 2013, **503**:218-223.
2. Nakshatri H, Anjanappa M, Bhat-Nakshatri P: **Ethnicity-Dependent and -Independent Heterogeneity in Healthy Normal Breast Hierarchy Impacts Tumor Characterization**. *Scientific reports* 2015, **5**:13526.
3. Kumar B, Prasad MS, Bhat-Nakshatri P, Anjanappa M, Kalra M, Marino N, Storniolo AMV, Rao X, Liu S, Wan J *et al*: **Normal breast-derived epithelial cells with luminal and intrinsic subtype-enriched gene expression document inter-individual differences in their differentiation cascade**. *Cancer Res* 2018, **78**(17):5107-5123.
4. Nievergelt CM, Maihofer AX, Shekhtman T, Libiger O, Wang X, Kidd KK, Kidd JR: **Inference of human continental origin and admixture proportions using a highly discriminative ancestry informative 41-SNP panel**. *Investig Genet* 2013, **4**(1):13.
5. Prat A, Perou CM: **Mammary development meets cancer genomics**. *Nat Med* 2009, **15**(8):842-844.
6. Han B, Bhowmick N, Qu Y, Chung S, Giuliano AE, Cui X: **FOXC1: an emerging marker and therapeutic target for cancer**. *Oncogene* 2017, **36**(28):3957-3963.
7. Eeckhoute J, Keeton EK, Lupien M, Krum SA, Carroll JS, Brown M: **Positive cross-regulatory loop ties GATA-3 to estrogen receptor alpha expression in breast cancer**. *Cancer Res* 2007, **67**(13):6477-6483.
8. Garbe JC, Vrba L, Sputova K, Fuchs L, Novak P, Brothman AR, Jackson M, Chin K, LaBarge MA, Watts G *et al*: **Immortalization of normal human mammary epithelial cells in two steps by direct targeting of senescence barriers does not require gross genomic alterations**. *Cell Cycle* 2014, **13**(21):3423-3435.
9. Villadsen R, Fridriksdottir AJ, Ronnov-Jessen L, Gudjonsson T, Rank F, LaBarge MA, Bissell MJ, Petersen OW: **Evidence for a stem cell hierarchy in the adult human breast**. *J Cell Biol* 2007, **177**(1):87-101.
10. Visvader JE, Stingl J: **Mammary stem cells and the differentiation hierarchy: current status and perspectives**. *Genes Dev* 2014, **28**(11):1143-1158.
11. Wang D, Cai C, Dong X, Yu QC, Zhang XO, Yang L, Zeng YA: **Identification of multipotent mammary stem cells by protein C receptor expression**. *Nature* 2014, **517**:81-84.
12. Kim J, Villadsen R, Sorlie T, Fogh L, Gronlund SZ, Fridriksdottir AJ, Kuhn I, Rank F, Wielenga VT, Solvang H *et al*: **Tumor initiating but differentiated luminal-like breast cancer cells are highly invasive in the absence of basal-like activity**. *Proc Natl Acad Sci U S A* 2012, **109**(16):6124-6129.
13. Al-Hajj M, Wicha MS, Benito-Hernandez A, Morrison SJ, Clarke MF: **Prospective identification of tumorigenic breast cancer cells**. *Proc Natl Acad Sci U S A* 2003, **100**(7):3983-3988.
14. Elkabets M, Pazarentzos E, Juric D, Sheng Q, Pelossof RA, Brook S, Benzaken AO, Rodon J, Morse N, Yan JJ *et al*: **AXL mediates resistance to PI3Kalpha inhibition by activating the EGFR/PKC/mTOR axis in head and neck and esophageal squamous cell carcinomas**. *Cancer Cell* 2015, **27**(4):533-546.
15. Antony J, Huang RY: **AXL-Driven EMT State as a Targetable Conduit in Cancer**. *Cancer Res* 2017, **77**(14):3725-3732.
16. Ginestier C, Hur MH, Charafe-Jauffret E, Monville F, Dutcher J, Brown M, Jacquemier J, Viens P, Kleer CG, Liu S *et al*: **ALDH1 Is a Marker of Normal and Malignant Human Mammary Stem Cells and a Predictor of Poor Clinical Outcome**. *Cell Stem Cell* 2007, **1**(5):555-567.
17. Korkaya H, Liu S, Wicha MS: **Breast cancer stem cells, cytokine networks, and the tumor microenvironment**. *J Clin Invest* 2011, **121**(10):3804-3809.
18. Liu S, Cong Y, Wang D, Sun Y, Deng L, Liu Y, Martin-Trevino R, Shang L, McDermott SP, Landis MD *et al*: **Breast Cancer Stem Cells Transition between Epithelial and Mesenchymal States Reflective of their Normal Counterparts**. *Stem cell reports* 2014, **2**(1):78-91.
19. Lim E, Wu D, Pal B, Bouras T, Asselin-Labat ML, Vaillant F, Yagita H, Lindeman GJ, Smyth GK, Visvader JE: **Transcriptome analyses of mouse and human mammary cell subpopulations reveal multiple conserved genes and pathways**. *Breast Cancer Res* 2010, **12**(2):R21.
20. Honeth G, Lombardi S, Ginestier C, Hur M, Marlow R, Buchupalli B, Shinomiya I, Gazinska P, Bombelli S, Ramalingam V *et al*: **Aldehyde dehydrogenase and estrogen receptor define a hierarchy of cellular differentiation in the normal human mammary epithelium**. *Breast Cancer Res* 2014, **16**(3):R52.
21. Pan D, Roy S, Gascard P, Zhao J, Chen-Tanyolac C, Tlsty TD: **SOX2, OCT3/4 and NANOG expression and cellular plasticity in rare human somatic cells requires CD73**. *Cell Signal* 2016, **28**(12):1923-1932.
22. Caramel J, Ligier M, Puisieux A: **Pleiotropic Roles for ZEB1 in Cancer**. *Cancer Res* 2017.
23. Nieto MA, Huang RY, Jackson RA, Thiery JP: **Emt: 2016**. *Cell* 2016, **166**(1):21-45.
24. Ye X, Tam WL, Shibue T, Kaygusuz Y, Reinhardt F, Ng Eaton E, Weinberg RA: **Distinct EMT programs control normal mammary stem cells and tumour-initiating cells**. *Nature* 2015, **525**(7568):256-260.
25. Dong J, Hu Y, Fan X, Wu X, Mao Y, Hu B, Guo H, Wen L, Tang F: **Single-cell RNA-seq analysis unveils a prevalent epithelial/mesenchymal hybrid state during mouse organogenesis**. *Genome Biol* 2018, **19**(1):31.
26. Pastushenko I, Brisebarre A, Sifrim A, Fioramonti M, Revenco T, Boumahdi S, Van Keymeulen A, Brown D, Moers V, Lemaire S *et al*: **Identification of the tumour transition states occurring during EMT**. *Nature* 2018, **556**(7702):463-468.

27. Pellacani D, Bilenky, M., Kannan, N., Heravi-Moussavi, A., Knapp, D.J.H.F., Gakkhar, S., Moksa, M., Carles, A., Moore, R., Mungall, A.J., Marra, M.A., Jones, S.J.M., Aparicio, S., Hirst, M., and Eaves, C.J.: **Analysis of normal human mammary epigenomes reveal cell-specfiic active enahcner states and associated transcription factor networks**. *Cell reports* 2016, **17**:2060-2074.
28. Morel AP, Ginestier C, Pommier RM, Cabaud O, Ruiz E, Wicinski J, Devouassoux-Shisheboran M, Combaret V, Finetti P, Chassot C *et al*: **A stemness-related ZEB1-MSRB3 axis governs cellular pliancy and breast cancer genome stability**. *Nat Med* 2017, **23**(5):568-578.
29. Su S, Chen J, Yao H, Liu J, Yu S, Lao L, Wang M, Luo M, Xing Y, Chen F *et al*: **CD10(+)GPR77(+) Cancer-Associated Fibroblasts Promote Cancer Formation and Chemoresistance by Sustaining Cancer Stemness**. *Cell* 2018, **172**(4):841-856 e816.
30. Huang KL, Li S, Mertins P, Cao S, Gunawardena HP, Ruggles KV, Mani DR, Clauser KR, Tanioka M, Usary J *et al*: **Proteogenomic integration reveals therapeutic targets in breast cancer xenografts**. *Nature communications* 2017, **8**:14864.
31. Shu S, Polyak K: **BET Bromodomain Proteins as Cancer Therapeutic Targets**. *Cold Spring Harb Symp Quant Biol* 2016, **81**:123-129.
32. Alsarraj J, Walker RC, Webster JD, Geiger TR, Crawford NP, Simpson RM, Ozato K, Hunter KW: **Deletion of the proline-rich region of the murine metastasis susceptibility gene Brd4 promotes epithelial-to-mesenchymal transition- and stem cell-like conversion**. *Cancer Res* 2011, **71**(8):3121-3131.
33. Heldin CH, Vanlandewijck M, Moustakas A: **Regulation of EMT by TGFbeta in cancer**. *FEBS Lett* 2012, **586**(14):1959-1970.
34. Ewan KB, Oketch-Rabah HA, Ravani SA, Shyamala G, Moses HL, Barcellos-Hoff MH: **Proliferation of estrogen receptor-alpha-positive mammary epithelial cells is restrained by transforming growth factor-beta1 in adult mice**. *Am J Pathol* 2005, **167**(2):409-417.
35. Weeber F, Ooft SN, Dijkstra KK, Voest EE: **Tumor Organoids as a Pre-clinical Cancer Model for Drug Discovery**. *Cell Chem Biol* 2017, **24**(9):1092-1100.
36. Linnemann JR, Miura H, Meixner LK, Irmler M, Kloos UJ, Hirschi B, Bartsch HS, Sass S, Beckers J, Theis FJ *et al*: **Quantification of regenerative potential in primary human mammary epithelial cells**. *Development* 2015, **142**(18):3239-3251.
37. Miller DH, Sokol ES, Gupta PB: **3D Primary Culture Model to Study Human Mammary Development**. *Methods Mol Biol* 2017, **1612**:139-147.
38. Prat A, Karginova O, Parker JS, Fan C, He X, Bixby L, Harrell JC, Roman E, Adamo B, Troester M *et al*: **Characterization of cell lines derived from breast cancers and normal mammary tissues for the study of the intrinsic molecular subtypes**. *Breast Cancer Res Treat* 2013, **142**(2):237-255.
39. Bensen JT, Graff M, Young KL, Sethupathy P, Parker J, Pecot CV, Currin K, Haddad SA, Ruiz-Narvaez EA, Haiman CA *et al*: **A survey of microRNA single nucleotide polymorphisms identifies novel breast cancer susceptibility loci in a case-control, population-based study of African-American women**. *Breast Cancer Res* 2018, **20**(1):45.
40. Wright KL, Adams JR, Liu JC, Loch AJ, Wong RG, Jo CE, Beck LA, Santhanam DR, Weiss L, Mei X *et al*: **Ras Signaling Is a Key Determinant for Metastatic Dissemination and Poor Survival of Luminal Breast Cancer Patients**. *Cancer Res* 2015, **75**(22):4960-4972.
41. Elenbaas B, Spirio L, Koerner F, Fleming MD, Zimonjic DB, Donaher JL, Popescu NC, Hahn WC, Weinberg RA: **Human breast cancer cells generated by oncogenic transformation of primary mammary epithelial cells**. *Genes Dev* 2001, **15**(1):50-65.
42. Witkiewicz AK, Knudsen ES: **Retinoblastoma tumor suppressor pathway in breast cancer: prognosis, precision medicine, and therapeutic interventions**. *Breast Cancer Res* 2014, **16**(3):207.
43. Cancer Genome Atlas N: **Comprehensive molecular portraits of human breast tumours**. *Nature* 2012, **490**(7418):61-70.
44. Ince TA, Richardson AL, Bell GW, Saitoh M, Godar S, Karnoub AE, Iglehart JD, Weinberg RA: **Transformation of different human breast epithelial cell types leads to distinct tumor phenotypes**. *Cancer Cell* 2007, **12**(2):160-170.
45. Lawson DA, Bhakta NR, Kessenbrock K, Prummel KD, Yu Y, Takai K, Zhou A, Eyob H, Balakrishnan S, Wang CY *et al*: **Single-cell analysis reveals a stem-cell program in human metastatic breast cancer cells**. *Nature* 2015, **526**(7571):131-135.
46. Saal LH, Holm K, Maurer M, Memeo L, Su T, Wang X, Yu JS, Malmstrom PO, Mansukhani M, Enoksson J *et al*: **PIK3CA mutations correlate with hormone receptors, node metastasis, and ERBB2, and are mutually exclusive with PTEN loss in human breast carcinoma**. *Cancer Res* 2005, **65**(7):2554-2559.
47. Nakshatri H, Kumar B, Burney HN, Cox ML, Jacobsen M, Sandusky GE, D'Souza-Schorey C, Storniolo AMV: **Genetic ancestry-dependent differences in breast cancer-induced field defects in the tumor-adjacent normal breast**. *Clin Cancer Res* 2019, **25**(9):2848-2859.
48. Prasad MS KB, Bhat-Nakshatri P, Anjanappa M, Sandusky G, Miller KD, Storniolo AM, and Nakshatri H: **Dual TGFbeta/BMP pathway inhibition enables expansion and characterization of multiple epithelial cell types of the normal and cancerous breast**. *Molecular Cancer Research* 2019, **17**(7):1556-1570.
49. Medina P, Navarro S, Bonet E, Martos L, Estelles A, Bertina RM, Vos HL, Espana F: **Functional analysis of two haplotypes of the human endothelial protein C receptor gene**. *Arteriosclerosis, thrombosis, and vascular biology* 2014, **34**(3):684-690.

50. Carroll JS, Liu XS, Brodsky AS, Li W, Meyer CA, Szary AJ, Eeckhoute J, Shao W, Hestermann EV, Geistlinger TR *et al*: **Chromosome-wide mapping of estrogen receptor binding reveals long-range regulation requiring the forkhead protein FoxA1**. *Cell* 2005, **122**(1):33-43.
51. Bernardo GM, Bebek G, Ginther CL, Sizemore ST, Lozada KL, Miedler JD, Anderson LA, Godwin AK, Abdul-Karim FW, Slamon DJ *et al*: **FOXA1 represses the molecular phenotype of basal breast cancer cells**. *Oncogene* 2013, **32**(5):554-563.
52. Bernardo GM, Lozada KL, Miedler JD, Harburg G, Hewitt SC, Mosley JD, Godwin AK, Korach KS, Visvader JE, Kaestner KH *et al*: **FOXA1 is an essential determinant of ERalpha expression and mammary ductal morphogenesis**. *Development* 2010, **137**(12):2045-2054.
53. Badve S, Turbin D, Thorat MA, Morimiya A, Nielsen TO, Perou CM, Dunn S, Huntsman DG, Nakshatri H: **FOXA1 expression in breast cancer--correlation with luminal subtype A and survival**. *Clin Cancer Res* 2007, **13**(15 Pt 1):4415-4421.
54. Thorat MA, Marchio C, Morimiya A, Savage K, Nakshatri H, Reis-Filho JS, Badve S: **Forkhead box A1 expression in breast cancer is associated with luminal subtype and good prognosis**. *J Clin Pathol* 2008, **61**(3):327-332.
55. Hurtado A, Holmes KA, Ross-Innes CS, Schmidt D, Carroll JS: **FOXA1 is a key determinant of estrogen receptor function and endocrine response**. *Nat Genet* 2011, **43**(1):27-33.
56. Espinal AC, Buas MF, Wang D, Cheng DT, Sucheston-Campbell L, Hu Q, Yan L, Payne-Ondracek R, Cortes E, Tang L *et al*: **FOXA1 hypermethylation: link between parity and ER-negative breast cancer in African American women?** *Breast Cancer Res Treat* 2017, **166**(2):559-568.
57. Clarke RB, Howell A, Potten CS, Anderson E: **Dissociation between steroid receptor expression and cell proliferation in the human breast**. *Cancer Res* 1997, **57**(22):4987-4991.
58. Katsura A, Tamura Y, Hokari S, Harada M, Morikawa M, Sakurai T, Takahashi K, Mizutani A, Nishida J, Yokoyama Y *et al*: **ZEB1-regulated inflammatory phenotype in breast cancer cells**. *Molecular oncology* 2017, **11**(9):1241-1262.

Published OnlineFirst July 11, 2018; DOI: 10.1158/0008-5472.CAN-18-0509

Translational Science

Cancer Research

Normal Breast-Derived Epithelial Cells with Luminal and Intrinsic Subtype-Enriched Gene Expression Document Interindividual Differences in Their Differentiation Cascade

Brijesh Kumar[1], Mayuri Prasad[1], Poornima Bhat-Nakshatri[1], Manjushree Anjanappa[1], Maitri Kalra[2], Natascia Marino[2], Anna Maria Storniolo[2], Xi Rao[3], Sheng Liu[3], Jun Wan[3], Yunlong Liu[3], and Harikrishna Nakshatri[1,4,5]

Abstract

Cell-type origin is one of the factors that determine molecular features of tumors, but resources to validate this concept are scarce because of technical difficulties in propagating major cell types of adult organs. Previous attempts to generate such resources to study breast cancer have yielded predominantly basal-type cell lines. We have created a panel of immortalized cell lines from core breast biopsies of ancestry-mapped healthy women that form ductal structures similar to normal breast in 3D cultures and expressed markers of major cell types, including the luminal-differentiated cell-enriched ERα–FOXA1–GATA3 transcription factor network. We have also created cell lines from PROCR $(CD201)^+$/$EpCAM^-$ cells that are likely the "normal" counterpart of the claudin-low subtype of breast cancers. RNA-seq and PAM50-intrinsic subtype clustering identified these cell lines as the "normal" counterparts of luminal A, basal, and normal-like subtypes and validated via immunostaining with basal-enriched KRT14 and luminal-enriched KRT19. We further characterized these cell lines by flow cytometry for distribution patterns of stem/basal, luminal-progenitor, mature/differentiated, multipotent $PROCR^+$ cells, and organogenesis-enriched epithelial/mesenchymal hybrid cells using CD44/CD24, CD49f/EpCAM, CD271/EpCAM, CD201/EpCAM, and ALDEFLUOR assays and E-cadherin/vimentin double staining. These cell lines showed interindividual heterogeneity in stemness/differentiation capabilities and baseline activity of signaling molecules such as NF-κB, AKT2, pERK, and BRD4. These resources can be used to test the emerging concept that genetic variations in regulatory regions contribute to widespread differences in gene expression in "normal" conditions among the general population and can delineate the impact of cell-type origin on tumor progression.

Significance: In addition to providing a valuable resource for the breast cancer research community to investigate cell-type origin of different subtypes of breast cancer, this study highlights interindividual differences in normal breast, emphasizing the need to use "normal" cells from multiple sources as controls to decipher the effects of cancer-specific genomic aberrations. *Cancer Res; 78(17); 5107–23. ©2018 AACR.*

Introduction

Normal breast epithelial cells are hierarchically organized broadly into bipotent mammary stem/basal (MaSC), luminal progenitor, and mature/differentiated luminal cells (1, 2). Luminal progenitor cells have been further classified into bipotent and committed progenitor cells based on cell surface marker profiles and expression patterns of keratins (2). Although basal cells express keratin 14 (KRT14) and luminal cells express keratin 19 (KRT19), cells expressing both keratins show luminal progenitor phenotype (3). Each of these cell types is associated with distinct transcription factor networks: *TP63* and *NFIB* in basal cells, *ELF5* and *EHF* in luminal progenitors, and *ESR1* and *FOXA1* in luminal cells (4). Although 11 different cell types have been described, it is acknowledged that current methods of sorting and classifying cell types based on surface markers and keratin expression may underestimate the level of heterogeneity in the normal breast (5). Furthermore, recent studies have identified interindividual genetic variations in noncoding regions affecting gene expression across tissues, thus supporting the concept of interindividual variability in the normal breast (6–8). Therefore, a clear understanding of the normal breast heterogeneity and signaling pathway differences is needed for better classification of breast tumors and for assessing tumor heterogeneity.

Breast cancers have been subclassified into five intrinsic subtypes based on gene expression patterns in tumors (9). These include estrogen receptor alpha (ERα)-positive luminal A and B subtypes, $HER2^+$ subtype, basal subtype, and normal-like

[1]Departments of Surgery, Indiana University School of Medicine, Indianapolis, Indiana. [2]Department of Medicine, Indiana University School of Medicine, Indianapolis, Indiana. [3]Department of Medical and Molecular Genetics, Indiana University School of Medicine, Indianapolis, Indiana. [4]Deaprtment of Biochemistry and Molecular Biology, Indiana University School of Medicine, Indianapolis, Indiana. [5]VA Roudebush Medical Center, Indianapolis, Indiana.

Note: Supplementary data for this article are available at Cancer Research Online (http://cancerres.aacrjournals.org/).

B. Kumar and M. Prasad contributed equally to this article.

Corresponding Author: Harikrishna Nakshatri, Indiana University School of Medicine, C218C, 980 West Walnut St., Indianapolis, IN 46202. Phone: 317-278-2238; Fax: 317-274-0396; E-mail: hnakshat@iupui.edu

doi: 10.1158/0008-5472.CAN-18-0509

©2018 American Association for Cancer Research.

Downloaded from cancerres.aacrjournals.org on September 4, 2018. © 2018 American Association for Cancer Research.

Published OnlineFirst July 11, 2018; DOI: 10.1158/0008-5472.CAN-18-0509

subtype. Another relatively rare molecular subtype, called the claudin-low, has been added subsequently, which is believed to originate from MaSCs (10). It is suggested that bipotent progenitor or luminal progenitors are the cell-type origin of basal breast cancers (11). $HER2^+$ tumors may arise from late luminal progenitors, whereas luminal A and B breast cancers may originate from differentiated luminal cells (11). Experimental validation of these possibilities is still challenging, because most of the prior culturing methods favored the outgrowth of basal-like breast epithelial cells and primary cells need to be directly used for transformation to obtain tumors with luminal and basal-like characteristics (12). Indeed, the most commonly used human mammary epithelial cells (HMEC) and MCF10A cells have basal-like gene expression pattern, and transformation of these cells gives rise to squamous cell carcinomas instead of adenocarcinomas (13, 14). Only one study has reported a method to generate cells with luminal characteristics and transformed counterpart of these cells, giving rise to tumors resembling human breast adenocarcinomas (13). For unknown reasons, this methodology has not been adapted widely.

The majority of "normal" tissue for breast cancer–related studies is derived from reduction mammoplasty or tissues adjacent to normal. However, a recent study that compared normal breast tissue donated by healthy volunteers with Komen Normal Tissue Bank at the Indiana University, reduction mammoplasty, and tumor-adjacent normal tissues found significant levels of histologic abnormalities in reduction mammoplasty as well as in tumor-adjacent normal specimens (15). Additionally, "normal" tissue adjacent to tumors undergoes extensive DNA methylation changes, specifically targeting transcription factor binding sites specifying chromatin architecture and stem cell differentiation pathways, including Wnt and FGF signaling networks, due to "field effects" attributed to tumors (16). Thus, although several publications in the literature have described generation of breast epithelial cell lines using tissues from reduction mammoplasty (12, 17–26), these cell lines are less likely suitable for mechanistic studies because of inherent genomic abnormalities and limited luminal cell features. Moreover, a maximum of four cell lines were created in each of these studies, which may not be sufficient to study interindividual variations in signaling pathways downstream of specific genomic aberrations. The mechanistic studies in breast cancer field, therefore, would benefit from having multiple normal/immortalized cells with luminal characteristics from healthy women to critically evaluate interplay between cell-type origin, interindividual differences in normal genome, and cancer-specific genomic aberrations driving tumor growth.

We recently reported a modified epithelial reprogramming growth condition that allowed propagation of breast epithelial cells with stem/basal, luminal progenitor, and differentiated/mature cell characteristics from core breast biopsies of healthy women (27). We reported significant interindividual variability in terms of specific cell types and an enrichment of PROCR $(CD201)^+/EpCAM^-$ subpopulation of cells in women of African ancestry. Here we describe creation of a panel of immortalized breast epithelial cell lines with gene expression pattern closely resembling the intrinsic subtypes of breast cancer and with features of cells prevalent during organogenesis (28). This resource will be made available to research community with the goal to address critical issues related to interplay between cell-type origin and specific genomic aberration driving breast cancer and to provide diversity in "normal" cell types reflecting interindividual variation and response to exogenous stimuli.

Materials and Methods

Primary cell culture and immortalization

Primary breast epithelial cells were created from fresh or cryopreserved, deidentified normal breast tissues of healthy Caucasian, Hispanic, and African American women donated to the Komen Tissue Bank (KTB) after informed written consent from subjects. All experiments were carried out in accordance with the approved guidelines of the Indiana University Institutional Review Board. International Ethical Guidelines for Biomedical Research Involving Human Subjects were followed. Tissues were dissociated using a 10% collagenase/hyaluronidase mixture (#07919, Stem Cell Technologies) and 10 μmol/L ROCK inhibitor (ALX-270-333-M005, Enzo Life Sciences) in culture media for 2 hours at 37°C. The dissociated cells were filtered through sterile 70-μm filter, washed in media, and centrifuged for 5 minutes at 1,000 RPM. The cells were cocultured on irradiated murine embryonic fibroblasts (Applied StemCell, Inc.) using media described previously (27). Cells were immortalized by human telomerase gene (hTERT) retrovirus using the vector pLXSN-hTERT. Retrovirus was prepared using the amphophoenix cells and primary cells were infected with viral supernatant along with 8 μg/mL polybrene (H9268, Sigma-Aldrich) for 4 hours and immortalized cells were selected by 100 μg/mL G418 (61-234, Corning). Because all cell lines are derived from normal breast tissues and are not established cell lines, authentication is not possible. However, based on karyotyping (Supplementary Fig. S1A), it is clear that these are not contaminating established cell lines. Most of the cell lines used are <20 passages. Cell lines in the laboratory are typically tested for *Mycoplasma* once in 2 years.

Genetic ancestry mapping and genotype analysis

Leukocyte DNA was obtained from the KTB. Blood was collected using the BD-Vacutainer spray-coated K3EDTA tubes (Becton Dickinson). Tubes were centrifuged for 15 minutes at 2,000 RPM. Once the upper phase (plasma) was removed, the tubes were stored in −80°C. Peripheral blood leukocyte DNA was extracted using AutoGenprep 965 (Autogen, Inc.). Genotyping was performed using the KASP technology (LGC Genomics) and a 41–single-nucleotide polymorphism (SNP) panel (labeled 41-AIM panel) selected from Nievergelt and colleagues (29). Genotype analysis using the 41-SNP panel along with a Bayesian clustering method (Structure Software V2.3.4) was able to discern continental origins including European (Caucasian)/Middle East, East Asia, Central/South Asia, Africa, Americas, and Oceania. A reference set was obtained from the Human Genome Diversity Project.

RNA sequencing

The concentration and quality of total RNA samples were first assessed using Agilent 2100 Bioanalyzer. A RIN (RNA Integrity Number) of five or higher was required to pass the quality control. Five hundred nanograms of RNA per sample were used to prepare dual-indexed strand-specific cDNA library using TruSeq Stranded mRNA Library Prep Kit (Illumina). The resulting libraries were assessed for its quantity and size distribution using Qubit and Agilent 2100 Bioanalyzer. Two hundred picomolar pooled libraries were utilized per flowcell for clustering amplification on cBot using HiSeq 3000/4000 PE Cluster Kit and sequenced with 2 ×

Downloaded from cancerres.aacrjournals.org on September 4, 2018. © 2018 American Association for Cancer Research.

Published OnlineFirst July 11, 2018; DOI: 10.1158/0008-5472.CAN-18-0509

75-bp paired-end configuration on HiSeq4000 (Illumina) using HiSeq 3000/4000 PE SBS Kit. A Phred quality score (Q score) was used to measure the quality of sequencing. More than 90% of the sequencing reads reached Q30 (99.9% base call accuracy).

The sequencing data were first assessed using FastQC (Babraham Bioinformatics) for quality control. All sequenced libraries were mapped to the human genome (UCSC hg19) using STAR RNA-seq aligner with the following parameter: "–outSAMmapqUnique 60." The reads distribution across the genome was assessed using bamutils (from ngsutils; ref. 30). Uniquely mapped sequencing reads were assigned to hg19 refGene genes using featureCounts (from subread) with the following parameters: "-s 2 –p –Q 10." Genes with read count per million (CPM) > 0.5 in more than three of the samples were kept. The data were normalized using TMM (trimmed mean of M values) method. Differential expression analysis was performed using *edgeR* R package (31). False discovery rate (FDR) was computed from *P* values using the Benjamini–Hochberg procedure. The multidimensional scaling (MDS) plot was drawn using *plotMDS* function in *edge*R to visualize the differences between the expression profiles of different samples in two dimensions. Raw sequencing data have been submitted to GEO (accession number GSE108541).

Predictor analysis of microarray classification

Gene expression was averaged from three replicates for each cell line. Based on the nearest predictor analysis of microarray (PAM50) centroid algorithm, intrinsic breast cancer subtypes were assigned using the *Pam50.robust* model from *genefu* R package (32). Luminal cell line MCF7 and basal cell lines HMECs and MCF10A were included as controls. Unsupervised hierarchical clustering of the PAM50 gene was performed and shown by using the function of *heatmap.2* in *gplots* R package.

Heatmap of pairwise comparison of cell lines

Differential expression analysis was performed between every two different cell lines. A total of 11,722 genes with FDR < 0.01 and absolute value of fold change > 1 in any pairwise comparison were selected for unsupervised hierarchical clustering.

Flow cytometry, flow cytometry imaging, and data analysis

Breast epithelial cells were collected by trypsinization, stained using antibodies CD49f-APC (FAB13501A), AXL-APC (FAB154A) (R&D Systems), PROCR (CD201)-PE (130-105-256), EpCAM-PE (130-091-253), EpCAM-APC (130-091-254), EpCAM-VioBlue (130-097-324; Miltenyi Biotech Inc.), CD271-APC (345108) (BioLegend), CD44-APC (559942), CD24-PE (555428; BD Pharmingen), Pan-Cytokeratin-FITC (F3418; Sigma-Aldrich) and ALDEFLUOR (01700; STEMCELL Technologies), and were acquired using a BD LSR II flow cytometer. Data were analyzed using CellQuest or FlowJo software. Forward and side scatter were used to ensure that only live cells were considered in the analysis. Gating was done using appropriate FITC (555573), PE (555749), APC (555576; BD Pharmingen), and VioBlue (130-094-670; Miltenyi Biotech) isotype control antibodies and only a representative isotype control for two fluorescent markers are shown.

For imaging flow cytometry, 500,000 to 1,000,000 cells per sample were stained with the following conjugated antibodies: 1:100 mouse anti-vimentin clone V9-Cy3 (C-9080, Sigma-Aldrich) and 1:20 mouse anti-CD324/E-cadherin clone 67A4-APC (A15717, Molecular Probes). The nucleus was stained with 1:5,000 Hoechst 33342 (H-3570, Molecular Probes). Cells were fixed and permeabilized using BD Cytofix/Cytoperm solution (51-2090KZ, BD Biosciences) for 20 minutes at 4°C. Blocking was done with 5% goat serum in 1× BD Perm/Wash solution (51-2091KZ, BD Biosciences) for 15 minutes at 4°C. Single-color control samples and cocktail sample with all the antibodies were incubated for 30 minutes at 4°C in dark. Cells were washed twice for 5 minutes between each step with 1× BD Perm/Wash solution. The last step of washing was done twice with 1× PBS to eliminate the detergent from the cell suspension. Cells were transferred to the UltraClear boil proof microcentrifuge tubes (Inc-607-GMT, Dot Scientific). The stained cells were stored in 1% formalin in dark at 4°C. A total of 5,000 images were analyzed using ImageStream Data Exploration and Analysis Software. In-focus cells were evaluated after gating on live single cells based on an aspect ratio near 1 and a low area of the bright field.

Antibodies and Western blotting analysis

Primary antibody against p16 (ab108349) was purchased from Abcam. An antibody against ERK (sc-94) was purchased from Santa Cruz Biotechnologies. Antibodies against pERK (9101), AKT1 (2967), AKT2 (5239), pAKT1 (9018), pAKT2 (8599), AKT (9272), and BRD4 (13440) were purchased from Cell Signaling Technologies. Antibody against E-cadherin (610181) was purchased from BD Biosciences. Antibody against vimentin (C9080) and β-actin (A5441) was purchased from Sigma-Aldrich. Horseradish peroxidase (HRP)-linked secondary antibodies against mouse, rabbit, and goat were purchased from Santa Cruz Biotechnologies. Cell lysates prepared in radioimmunoassay buffer were analyzed by Western blotting as described previously (33).

Mammosphere formation assay and cell proliferation assay

Single-cell suspensions were cultured to form mammospheres in ultralow attachment 6-well plate at a density of 5,000 cells/well in MammoCult basal medium (human) supplemented with MammoCult proliferation supplement (human), Heparin, 1% penicillin/streptomycin (STEMCELL Technologies Inc.), and hydrocortisone. Phase-contrast images of mammospheres were taken at day 6. For secondary and tertiary cultures, mammospheres were collected by centrifugation, washed, trypsinized, filtered through 40-μm filter and an equal number of cells were replated in mammosphere media. Mammospheres were counted using ImageJ. For cell proliferation assay, cells were seeded in 96-well plates followed by treatment with TGFβ1 (1 ng/mL; 7754-BH, R&D Systems) for 3 to 4 days. At the end of the treatment, bromodeoxyuridine (BrdUrd) incorporation-ELISA was done using BrdUrd Proliferation Assay Kit (Calbiochem/Millipore) according to the manufacturer's instructions.

Karyotyping of immortalized cell lines

Metaphase chromosome spreads from KTB cell lines at passage between 10 and 20 were prepared by culturing proliferative cells with 0.05 μg/mL colcemid (10295892001, Sigma-Aldrich) for 2 to 12 hours depending on the cell proliferation rate, followed by resuspension in hypotonic solution (75 mmol/L KCl) for 15 minutes at 37°C. Cells were fixed by resuspension in Carnoy's Fixative (3:1 methanol:acetic acid) onto a clean dry slide and allowed to air dry. Fixed cells were stained with Hoechst 33342 (Molecular Probes H-3570) and analyzed under microscope.

Downloaded from cancerres.aacrjournals.org on September 4, 2018. © 2018 American Association for Cancer Research.

Published OnlineFirst July 11, 2018; DOI: 10.1158/0008-5472.CAN-18-0509

More than five metaphase chromosome spreads were analyzed for each cell line.

3D *in vitro* Matrigel, collagen, and hydrogel assays

Eight-chamber cover glass system (155409, Lab-Tek II) was coated with 50 μL of Matrigel (354234, Corning) per well and incubated for 30 minutes at 37°C for the gel to solidify. A total of 6,000 cells in 400 μL of overlay media comprising primary cell media with 2% Matrigel were added to the Matrigel-coated wells. The media were changed every 3 to 4 days, and cells were cultured for 10 to 12 days.

Floating Collagen scaffolds were prepared based on a protocol from Linnemann and colleagues (34) with some modifications. A total of 5,000 cells per well in a 48-well plate were used. The final concentration of the rat-tail collagen I (A10483-01, Corning) used for the scaffold was 1.3 mg/mL. HEPES buffer measuring 550 mmol/L, amounting to 1/10th volume of the collagen, was used as a neutralizing solution. The neutralizing solution was first added to the cell suspension, followed by a quick addition of rat-tail collagen. Two hundred microliters of the collagen–cell mixture was immediately plated into 48-well plate kept on ice. The gels were incubated for 1 hour at 37°C to promote polymerization. Five hundred microliters of mammary epithelial basal medium (CC-3151, Lonza) with 3 μmol/L Y-27632 and 10 μmol/L forskolin were carefully added to wells. The gels were then detached from the bottom of the well by encircling pipet tip around the edge of the gel. The media were changed every 3 days, and cells were cultured for up to 10 days. After day 3, Y-27632 was withdrawn from the medium.

Floating hydrogel scaffold was prepared based on protocol from Miller and colleagues (35). Briefly, 10,000 cells per well were used in a 4-chamber slide. Extracellular matrix (25×) was made by adding 0.5 mg/mL laminin (23017015, Thermo Fisher), 0.25 mg/mL hyaluronic acid (385908, Millipore), and 0.5 mg/mL fibronectin (33016015, Thermo Fisher) in PBS. The final concentration of rat-tail collagen I was 1.3 mg/mL. To make 1 mL of the hydrogel, collagen I was first neutralized with 12.5% of 0.1N NaOH followed by the addition of ice cold MEGM to make the volume 910 μL. Extracellular matrix (40 μL) was quickly added. Lastly, 50 μL of cell suspension was added to the mixture and pipetted up and down 3 to 4 times to uniformly distribute the cells in the mixture. The hydrogel mixture (200 μL) was added to the 4-well chamber slide kept on ice. The gel was incubated for 1 hour at 37°C. MEGM (1 mL) was added after the polymerization was complete. The gels were then detached from the bottom of the well by encircling a P200 pipet tip around the edge of the gel. All the steps were done on ice. Media were changed twice a week, and the cells were cultured for 10 to 14 days.

RNA extraction and quantitative real-time PCR

Total RNA was isolated using the RNeasy Kit (74106, Qiagen), and 1 μg of RNA was used to synthesize cDNA with the Bio-Rad iScript cDNA Synthesis Kit (170–8891). Quantitative real-time PCR (qRT-PCR) was performed using TaqMan universal PCR mix and predesigned gene expression assays with best coverage from Applied Biosystems. The following assays were used in our study: *ACTB* (Hs01060665_g1), *ESR1* (Hs01046816_m1), *GATA3* (Hs00231122_m1), *FOXA1* (Hs04187555_m1), *FOXC1* (Hs00559473_s1), *ZEB1* (Hs01566408_m1), *SNAI1* (Hs00195591_m1), *SNAI2* (Hs00161904_m1), *TWIST1* (Hs00361186_m1), *TWIST2* (Hs00382379_m1), *EHF* (Hs00171917_m1), and *TP63* (Hs00978340_m1).

Electrophoretic mobility shift assay

Electrophoretic mobility shift assay (EMSA) with whole-cell extracts from immortalized cell lines was performed as described previously (33). Oligonucleotides with consensus DNA binding sites for NF-κB (E3292), AP-1 (E3201), and OCT-1 (E3242) were purchased from Promega Corporation.

Immunofluorescence of monolayer culture and 3D collagen structures

Cells were seeded in monolayer glass bottomed microwell dish (P35G-0-14-C, MatTek Corporation) overnight and fixed with 4% paraformaldehyde for 10 minutes at room temperature (RT). Cells were blocked with a blocking buffer comprising 5% goat serum in 1% BSA and 0.1% Triton-X-100 in PBS for 1 hour at RT or overnight at 4°C. KRT14 (ab7800, Abcam) and KRT19 (Abcam, ab52625) primary antibodies were diluted to 1:200 in the antibody diluent (S0809, Dako) and added to cells for 1.5 hours at RT. Secondary goat anti-rabbit Alexa Fluor 555 (A21428, Life Technologies) and secondary rabbit anti-mouse Alexa Fluor 488 (A11059, Life Technologies) antibodies were diluted to 1:400 and added to cells for 1 hour in dark at RT. Nuclear staining was done with 1:5,000 Hoechst 33342 (H-3570, Molecular Probes) for 5 minutes. Cells were washed 3 times for 2 minute each, between every step. Cells were stored in PBS in dark and images were taken within 48 hours of staining with Olympus FV1000 MPE inverted confocal microscope.

Cells in the collagen and hydrogel scaffolds were rinsed twice for 10 minutes with 0.05% Tween 20 in PBS (PBST) and fixed with 4% paraformaldehyde for 20 minutes, followed by quenching with 0.15 M glycine for 10 minutes at RT. Cells were permeabilized with 0.15% Triton-X-100 in PBST at RT overnight and were blocked with 5% goat serum and 5% fetal calf serum in PBST for at least 2 hours. KRT14 and KRT19 primary antibodies were diluted to 1:200 in PBST and cells were incubated overnight at 4°C. Secondary goat anti-rabbit Alexa Fluor 555 and rabbit anti-mouse Alexa Fluor 488 antibodies were diluted to 1:300 and added to cells for 2 hour in dark at RT. The nuclei were stained with 1:2500 Hoechst 33342 for 15 minutes. All steps were carried out on a rocker and cells were washed twice for 10 minutes with PBST between every step. The stained scaffolds were stored in PBS and imaged with Olympus FV1000 MPE inverted confocal microscope by placing the gel on the glass surface of the microwell dish with PBS to help the gel stick to the bottom without floating. The XLUMPLFLN-W 20× long working distance objective was used to enable imaging of cells embedded in the gel and 3D images were processed using Imaris software.

Statistical analyses

All experiments were conducted in three biological replicates. Statistical analyses were performed using Prism software program (version 6.0). qPCR, FACS, and cell proliferation assay were analyzed using one-way ANOVA. P values of <0.05 were considered statistically significant.

Results

Isolation and immortalization of primary breast epithelial cells

We generated immortalized cell lines from core biopsies of normal breast by overexpressing human telomerase gene (hTERT)

Downloaded from cancerres.aacrjournals.org on September 4, 2018. © 2018 American Association for Cancer Research.

Published OnlineFirst July 11, 2018; DOI: 10.1158/0008-5472.CAN-18-0509

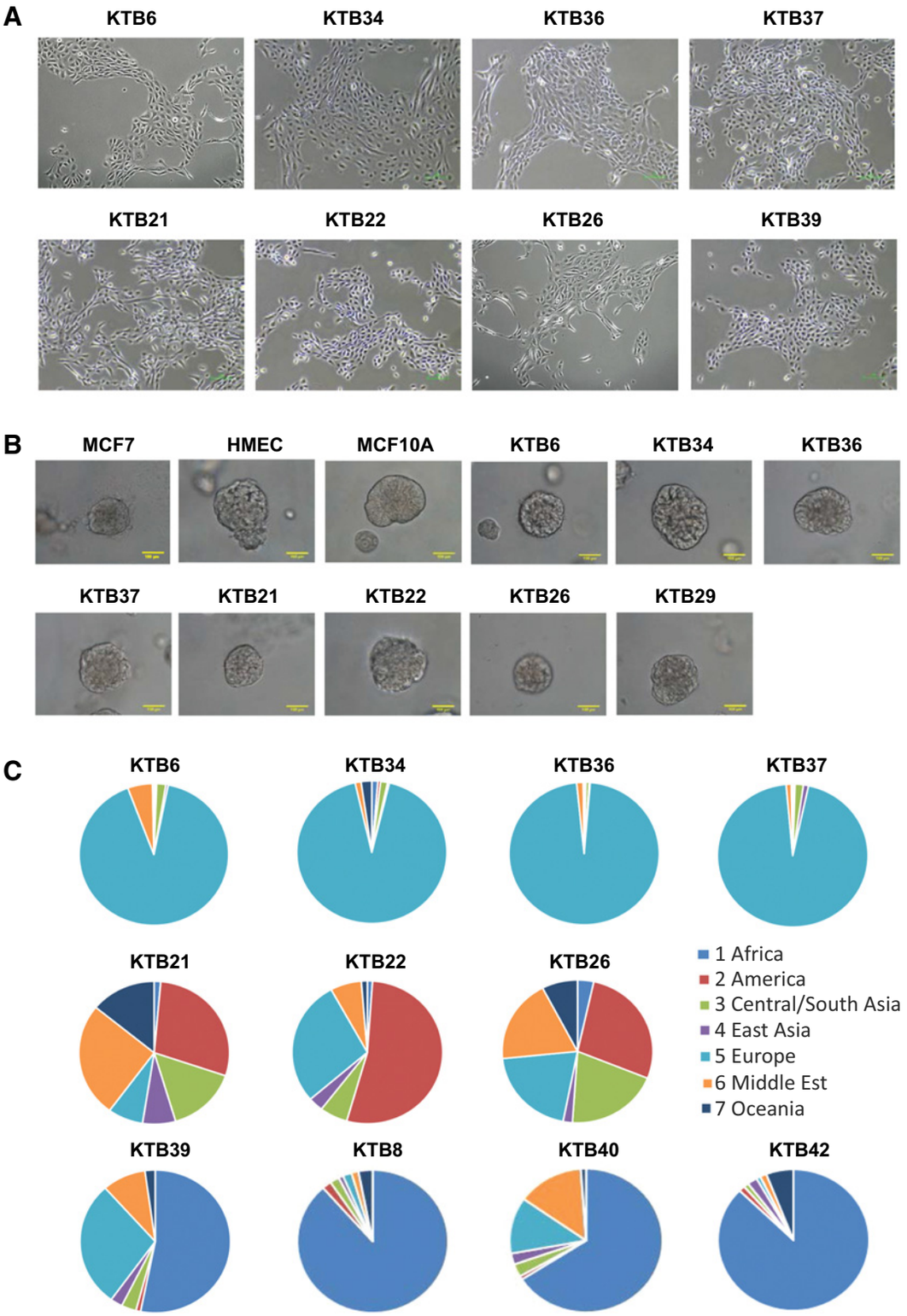


Figure 1.

Generation of cell lines from core biopsies of ancestry-mapped healthy women (KTB cell lines). **A,** Phase contrast images of KTB cell lines. **B,** KTB cell lines form acini on Matrigel. **C,** Highly discriminative ancestry informative 41-SNP genomic analyses of KTB cell lines show expected heterogeneity in Hispanic women (KTB21, KTB22, and KTB26) compared with Caucasian (KTB6, KTB34, KTB36, and KTB37) and African American women (KTB8, KTB39, KTB40, and KTB42).

Downloaded from cancerres.aacrjournals.org on September 4, 2018. © 2018 American Association for Cancer Research.

Published OnlineFirst July 11, 2018; DOI: 10.1158/0008-5472.CAN-18-0509

in primary cells isolated and propagated under modified epithelial reprogramming assay (Fig. 1A). To evaluate the ability of immortalized cell lines (called KTB cell lines hereafter) to form acini resembling normal breast, we grew cells on Matrigel. KTB immortalized cell lines formed acini of varying sizes, similar to the acini formed by the most commonly used "normal" cell line MCF10A (Fig. 1B).

To ensure that KTB cell lines are representative of women of different genetic ancestry, all samples were subjected to highly discriminative ancestry informative 41-SNP genomic analyses (29). As shown in Fig. 1C, while self-reported white women were enriched for European ancestry markers (KTB6, KTB34, KTB36, and KTB37), expectedly, the self-reported Hispanic women (KTB21, KTB22, and KTB26) displayed highly heterogeneous ancestry marker distribution. Four self-reported African American women (KTB8, KTB39, KTB40, and KTB42) had inherited >50% of African ancestry markers. To ensure diploid nature of immortalized cell lines, karyotype analyses were performed after arresting cells at metaphase of the cell cycle by colcemid treatment, which confirmed 46 chromosomes in all samples (Supplementary Fig. S1A). Several cell lines have undergone >20 passages and remained diploid.

Molecular subtype classification of KTB cell lines

To determine whether gene expression patterns in immortalized KTB cell lines differ from the commonly used normal cell lines MCF10A and HMECs, we performed RNA-seq analyses. A cell line derived from a BRCA2-mutant carrier 1505-10B was also included in the analyses. Principal component analyses (PCA) revealed distinct differences in the gene expression pattern between KTB cell lines compared with HMEC, MCF10A, and the luminal breast tumor cell line MCF7 (Fig. 2A). Independent PCA analyses of KTB cell lines showed significant differences between cell lines, further confirming interindividual variations in characteristics of "normal" breast cell lines (Fig. 2B).

Unsupervised clustering analyses separated KTB cell lines from HMEC and MCF10A, with the exception of KTB40 and KTB42, which clustered with HMEC and MCF10A (Fig. 2C). Previous studies have shown enrichment of a basal cell gene expression pattern in HMEC and MCF10A but the KTB cell lines displayed a distinct gene expression pattern compared with HMEC and MCF10A cell lines (14). We performed ingenuity pathway analyses of 733 genes expressed at higher level in KTB luminal cell lines compared with HMEC, MCF10A, KTB40, and KTB42 (Supplementary Table S1; subcluster genes). Canonical pathways typically enriched in luminal cells, including ErbB and PAK signaling, were elevated in KTB cell lines (Supplementary Table S1, IPA_Pathway; refs. 11, 36).

Encouraged by these results, we next determined whether KTB cell lines represent "normal" counterparts of intrinsic subtypes of breast cancer. Intrinsic subtypes have been predicted to originate from normal counterparts but it has never been demonstrated experimentally (11). Using the PAM50 model in genefu, we calculated the subtype probability score for each cell line. KTB34 represented luminal A, KTB21, KTB22, and KTB36 represented basal, and KTB26, KTB37, KTB39, and 1505-10B (BRCA2 mutant) represented normal-like subtypes (Fig. 2D and E). These results document the presence of immortalized cells with a gene expression pattern similar to intrinsic subtypes of breast cancer.

To further characterize KTB cell lines for luminal and basal-enriched gene expression, we measured the expression levels of *ESR1*, *FOXA1*, *GATA3*, and *FOXC1*. *ESR1*, *FOXA1*, and *GATA3* are expressed predominantly in luminal cells and form a hormone-responsive lineage-specific transcription factor network, whereas *FOXC1* expression identifies cells with basal cell–enriched gene expression (37, 38). Indeed, all four of these transcription factors were expressed at variable levels in KTB cell lines and, to our knowledge, this is the first report to document immortalized breast epithelial cell lines expressing *ESR1*, *GATA3*, and *FOXA1* (Fig. 2F). We also noted another difference in KTB cell lines and general observation about immortalization of breast epithelial cells. Immortalization is often associated with the loss of cell-cycle inhibitor and senescence promoter p16(INK4) (39). However, RNA-seq results showed significant expression of p16(INK4) transcripts in KTB cell lines. Western blotting also confirmed expression of p16(INK4) in several of KTB cell lines (Fig. 2G). Thus, immortalized KTB cell lines are unique in their characteristics compared with available cell lines, including the expression of luminal-enriched transcription factors and p16(INK4).

To rule out the possibility that intrinsic subtype phenotype is acquired during immortalization, we performed RNA-seq analyses of primary cells prior to immortalization. PAM50 classification identified primary cells with luminal A, luminal B, basal, and normal-like features (Fig. 2H). It does appear that immortalization selects for specific subtypes as intrinsic subtype classification prior and after immortalization did not overlap. In particular, KTB40 and KTB42 prior to immortalization clustered as "normal-like" subtype. However, we do note a caveat in our analyses, as RNA-seq of primary and immortalized cells was done at different times due to volume of samples and these comparisons have limitations. Nonetheless, our results clearly show that cells derived from the normal breast prior and after immortalization have gene expression patterns overlapping the gene expression patterns in intrinsic subtypes of breast cancer.

Keratin expression pattern identifies distinct subpopulation of cells within immortalized KTB cell lines

Keratin 14 (KRT14) is typically expressed at higher levels in basal cells, whereas keratin 19 (KRT19) is expressed in luminal cells. Cells that express both KRT14 and KRT19 ($KRT14^+$/$KRT19^+$) are considered to have luminal progenitor properties (3). We used immunofluorescence to determine expression pattern of these two keratins. Consistent with the results of PAM50 classification, the luminal A cell line KTB34 was strongly positive for KRT19, similar to MCF7, although low-level KRT14 expression was seen in this cell line, whereas the basal cell line KTB22 was strongly positive for KRT14 (Fig. 3A). The normal-like cell line KTB39 was $KRT14^+$/$KRT19^+$.

Phenotypic heterogeneity in immortalized KTB cell lines

In order to confirm that the immortalized cell lines show the interindividual variability, as previously shown for the primary breast epithelial cells (27), we subjected KTB cell lines prior to and after immortalization to phenotypic analyses using additional markers that define stem/basal, luminal progenitor, luminal differentiated, luminal, and basal cells (1). $CD49f^+$/$EpCAM^-$, $CD49f^+$/$EpCAM^+$, and $CD49f^-$/$EpCAM^+$ cells are described as breast stem, luminal progenitor, and mature/differentiated cells, respectively (1). Primary cells prior to immortalization showed interindividual variation in the levels of all three populations of cells (Supplementary Fig. S1B). The CD44/CD24 staining pattern also confirmed interindividual heterogeneity. After

Downloaded from cancerres.aacrjournals.org on September 4, 2018. © 2018 American Association for Cancer Research.

Published OnlineFirst July 11, 2018; DOI: 10.1158/0008-5472.CAN-18-0509

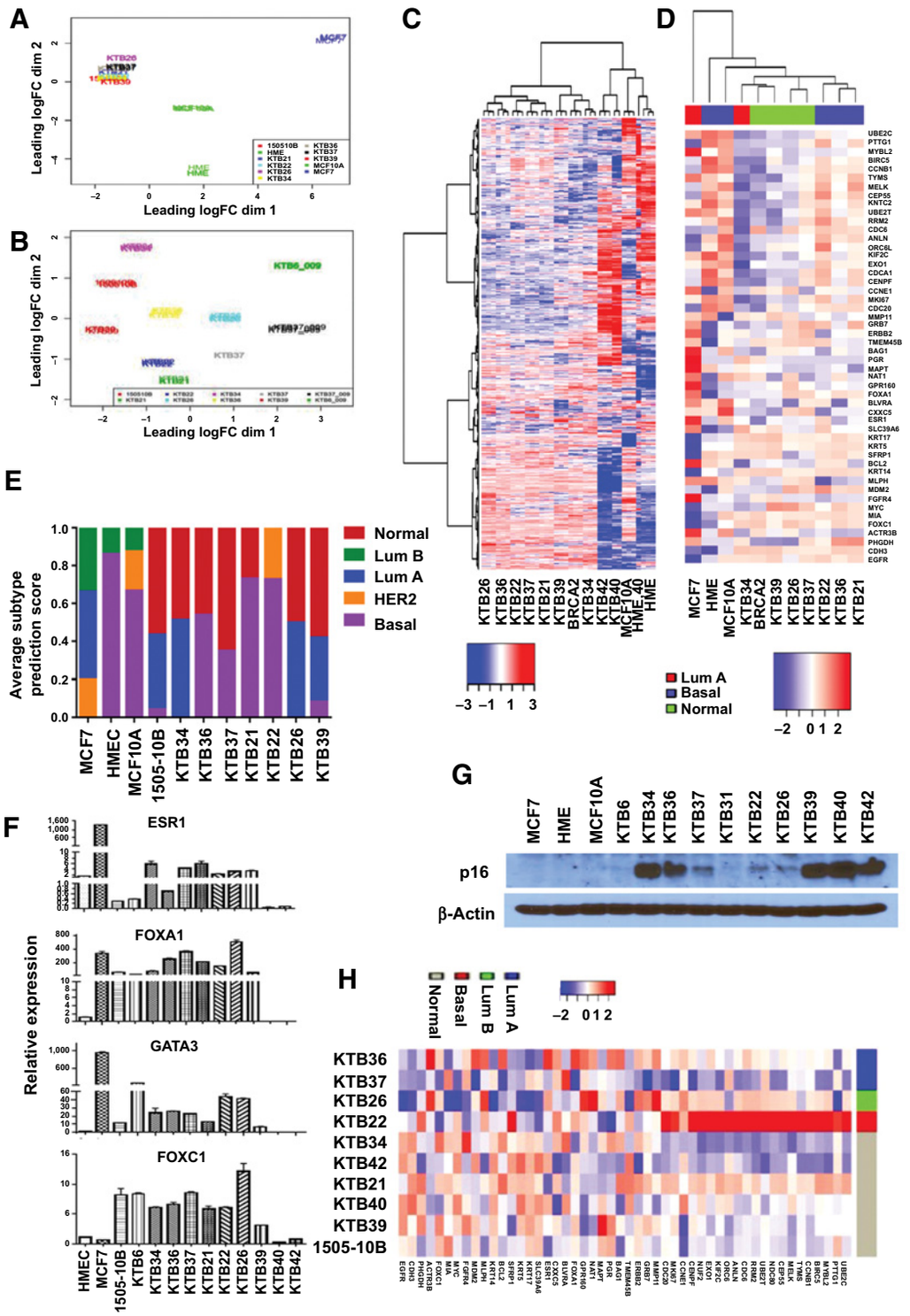


Figure 2.

Classification of KTB cell lines based on PAM50 subtype predictor. **A,** PCA analyses of RNA-seq data of all immortalized cell lines including HME, MCF10A, and MCF7. 1505-10B is a BRCA2-mutant cell line. **B,** PCA analyses of RNA-seq data of KTB cell lines show heterogeneity in gene expression. **C,** Unsupervised clustering of RNA-seq data shows a distinct gene expression pattern in KTB cell lines except KTB40 and KTB42 compared with MCF10A and HME. KTB40, and KTB42 clustered with HME and MCF10A. **D,** Classification of KTB cell lines based on PAM50 subtype classifier. **E,** Subtype prediction scores of individual cell lines. **F,** Expression levels of luminal-enriched *ESR1, FOXA1, GATA3*, and basal-enriched *FOXC1* in different KTB cell lines. **G,** p16(INK4) expression in different cell lines. **H,** Classification of primary cells prior to immortalization based on the PAM50 subtype classifier.

Downloaded from cancerres.aacrjournals.org on September 4, 2018. © 2018 American Association for Cancer Research.

Published OnlineFirst July 11, 2018; DOI: 10.1158/0008-5472.CAN-18-0509

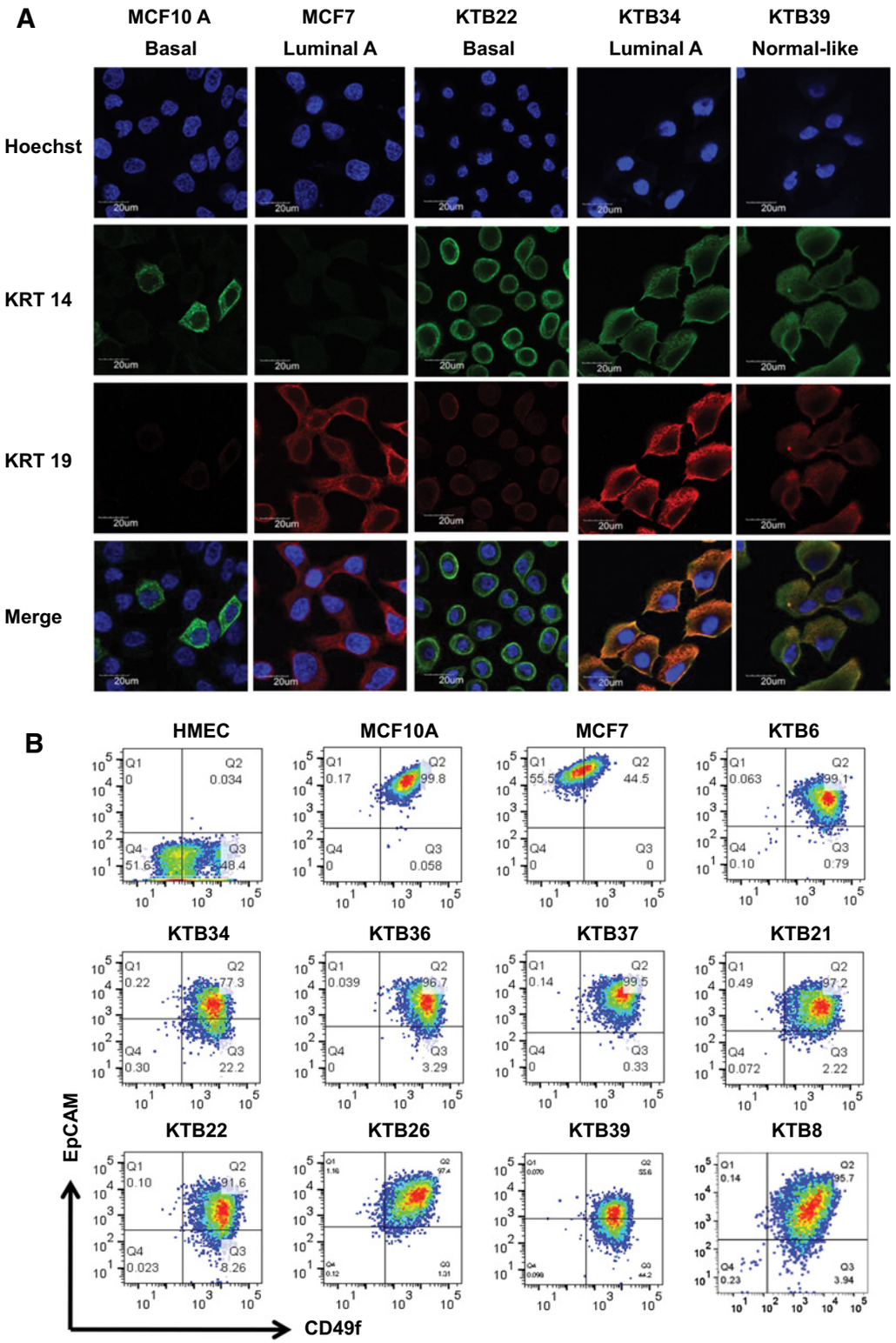


Figure 3.

Luminal and basal cell–enriched features of KTB cell lines. Representative data are presented. **A,** KRT14 and KRT19 staining patterns of different cell lines. MCF10A and MCF7 were used as controls. **B,** CD49f and EpCAM staining patterns of cell lines. CD49f$^+$/EpCAM$^-$, CD49f$^+$/EpCAM$^+$, and CD49f$^-$/EpCAM$^+$ cells correspond to stem/basal, luminal progenitor, and mature luminal cells, respectively.

Downloaded from cancerres.aacrjournals.org on September 4, 2018. © 2018 American Association for Cancer Research.

Published OnlineFirst July 11, 2018; DOI: 10.1158/0008-5472.CAN-18-0509

immortalization, this variation decreased considerably with the majority of cell lines displaying luminal progenitor phenotype ($CD49f^+/EpCAM^+$; Fig. 3B). However, there was some variation in the intensity of staining as well as the presence of $CD49f^+/EpCAM^-$ stem/basal subpopulation (Fig. 3B). CD49f/EpCAM staining patterns in immortalized KTB cell lines were different from MCF10A, HMEC, and MCF7. CD201 (PROCR), which identifies cells with multipotent stem cell activity in the mouse mammary gland (40), was expressed at variable levels in immortalized KTB cell lines with few cell lines containing a mix of $CD201^+/EpCAM^+$ and $CD201^-/EpCAM^+$ cells, while the others with only $CD201^-/EpCAM^+$ subpopulation (Supplementary Fig. S2). Cell lines showed similar differences in the staining pattern for CD271, which is expressed predominantly in basal cells (Supplementary Fig. S2; ref. 41).

CD44/CD24 and ALDEFLUOR staining identify interindividual differences in normal counterparts of breast cancer stem cells that transition between epithelial and mesenchymal states

Recent studies have shown that breast cancer stem cells transition between epithelial and mesenchymal states with proliferative epithelial-like stem cells being positive for aldehyde dehydrogenase (ALDH) and invasive/quiescent mesenchymal stem-like cells being $CD44^+/CD24^-$ (42). The gene expression pattern in epithelial and mesenchymal cancer stem cells is remarkably similar to luminal and basal cells of the normal breast. We characterized KTB cell lines for the presence of these two subpopulations of cells. Although most of the cells in the KTB cell lines were $CD44^+/CD24^+$, the $CD44^+/CD24^-$ population varied between cell lines (Fig. 4A). As with the CD49f/EpCAM staining pattern, immortalized cells compared with cells prior to immortalization showed more homogeneity in CD44/CD24 staining. Significant variations in the levels of $ALDEFLUOR^+$ cells were noted between cell lines (Fig. 4B and Supplementary Fig. S3A). AXL, which has recently been shown to drive epithelial-to-mesenchymal transition (EMT) and confers resistance to PI3K inhibitors (43), was expressed at variable levels in cell lines (Supplementary Fig. S3B). Collectively, these multiple assays document enormous interindividual heterogeneity in the normal breast, which is at least partially reflected in immortalized cell lines.

Stemness properties of immortalized cell lines show interindividual variability

To determine whether KTB cell lines show interindividual variability in stemness, we performed mammosphere assay in all cell lines. KTB cell lines showed variable size of spheres, indicating differences in the stemness property of each cell line (Fig. 4C). Mammospheres were subjected to flow cytometry for CD49f/EpCAM, CD201/EpCAM, and CD44/CD24 expression to distinguish stem, luminal progenitor, basal, and mature/differentiated cells. This assay suggested interindividual differences in differentiation capacity (Fig. 4D). For example, CD44 and CD24 staining of mammospheres showed interindividual variability in the ratio between $CD44^+/CD24^-$, $CD44^-/CD24^+$, and $CD44^+/CD24^+$ cells. Notably, more $CD44^-/CD24^+$ cells, which were never observed under 2D culture of primary cells before or after immortalization (Fig. 4A; ref. 27), were found in mammosphere cultures of most of the cell lines (Fig. 4D). Similar increase in $CD49f^-/EpCAM^+$ cells, representing mature/differentiated cells, was noted in mammosphere cultures of few cell lines (Supplementary Fig. S4). We further confirmed interindividual differences in differentiation under mammosphere condition by comparing the expression levels of *ESR1*, *FOXA1*, *GATA3*, and *FOXC1* under 2D and mammosphere conditions. The expression levels of *ESR1*, *FOXA1*, and *GATA3* increased substantially in cells from mammosphere cultures compared with 2D cultures (Fig. 4E). Thus, stemness and differentiation capabilities show interindividual variability, and cell lines/assay systems described here are useful for assessment at individual level.

PROCR $(CD201)^+/EpCAM^-$ cells enriched in African American women correspond to naturally occurring ZEB1–positive cells with EMT phenotype

KTB40 and KTB42 cell lines distinguished themselves from the rest of the cell lines with respect to gene expression patterns and cell surface marker profiles. These cell lines were derived from African American women. During our efforts to immortalize breast epithelial cells from African American women, we consistently observed rapid immortalization of cells with EMT features, which overtook colonies with epithelial phenotype. We placed additional efforts to characterize these cell lines. These cells at low density displayed epithelial morphology but acquired fibroblastic phenotype with increasing density (Fig. 5A). These cell lines contained heterogeneous population of cells with $CD201^+/EpCAM^-$, $CD201^-/EpCAM^-$, $CD201^+/CD90^-$, $CD201^+/CD90^+$, $CD201^-/CD90^+$, and $CD44^+/CD24^-$ cell surface marker profiles (Fig. 5B). The majority of cells stained positive for pan-keratin antibody (Fig. 5B). We included CD90 in this analysis because CD90 is expressed in MaSC, fibroblasts, adipocytes, and mesenchymal stem cells (44). Unlike other KTB cell lines described above, these cells expressed EMT and stemness-associated transcription factors *ZEB1*, *TWIST1*, *TWIST2*, *SNAI1*, and *SNAI2* (Fig. 5C and D; refs. 45, 46). KTB40 and KTB42 did not demonstrate E-cadherin positivity in Western blotting while the remaining cell lines expressed significant levels of E-cadherin (Fig. 5E). Vimentin expression was significantly higher in KTB40 and KTB42 compared with other KTB cell lines.

Lack of E-cadherin but elevated vimentin provided further evidence for EMT features of KTB40 and KTB42 cell lines. Because EMT features partially overlap with stemness (47), serial dilution mammosphere assay was performed to determine whether KTB40 and KTB42 differ in stemness relative to other KTB cell lines. KTB40 and KTB42 displayed higher mammosphere-forming ability than other KTB cell lines and only these two cell lines were able to form tertiary mammospheres (Fig. 5F and G).

It is puzzling to detect low levels of vimentin expression in our E-cadherin–positive KTB cell lines. This could be due to the presence of distinct subpopulation of cells expressing either E-cadherin or vimentin. Alternatively, there could be a population of cells that express both proteins. Recent studies support the concept of epithelial/mesenchymal or luminal/basal hybrid cells playing a role in organogenesis and tumor progression (28, 48). To distinguish between these possibilities, we measured the expression levels of luminal progenitor cell-enriched *EHF* and basal cell-enriched *TP63* transcription factors (4). Although KTB40 and KTB42 did not express these two genes, the remaining KTB lines expressed variable levels of both *EHF* and *TP63* (Fig. 5H). The most commonly used HMEC cells expressed only *TP63*. To confirm the presence of hybrid cells, we double stained cells for E-cadherin and vimentin, and single cells were analyzed by imaging cytometry. A significant number of cells in the KTB34

Downloaded from cancerres.aacrjournals.org on September 4, 2018. © 2018 American Association for Cancer Research.

Published OnlineFirst July 11, 2018; DOI: 10.1158/0008-5472.CAN-18-0509

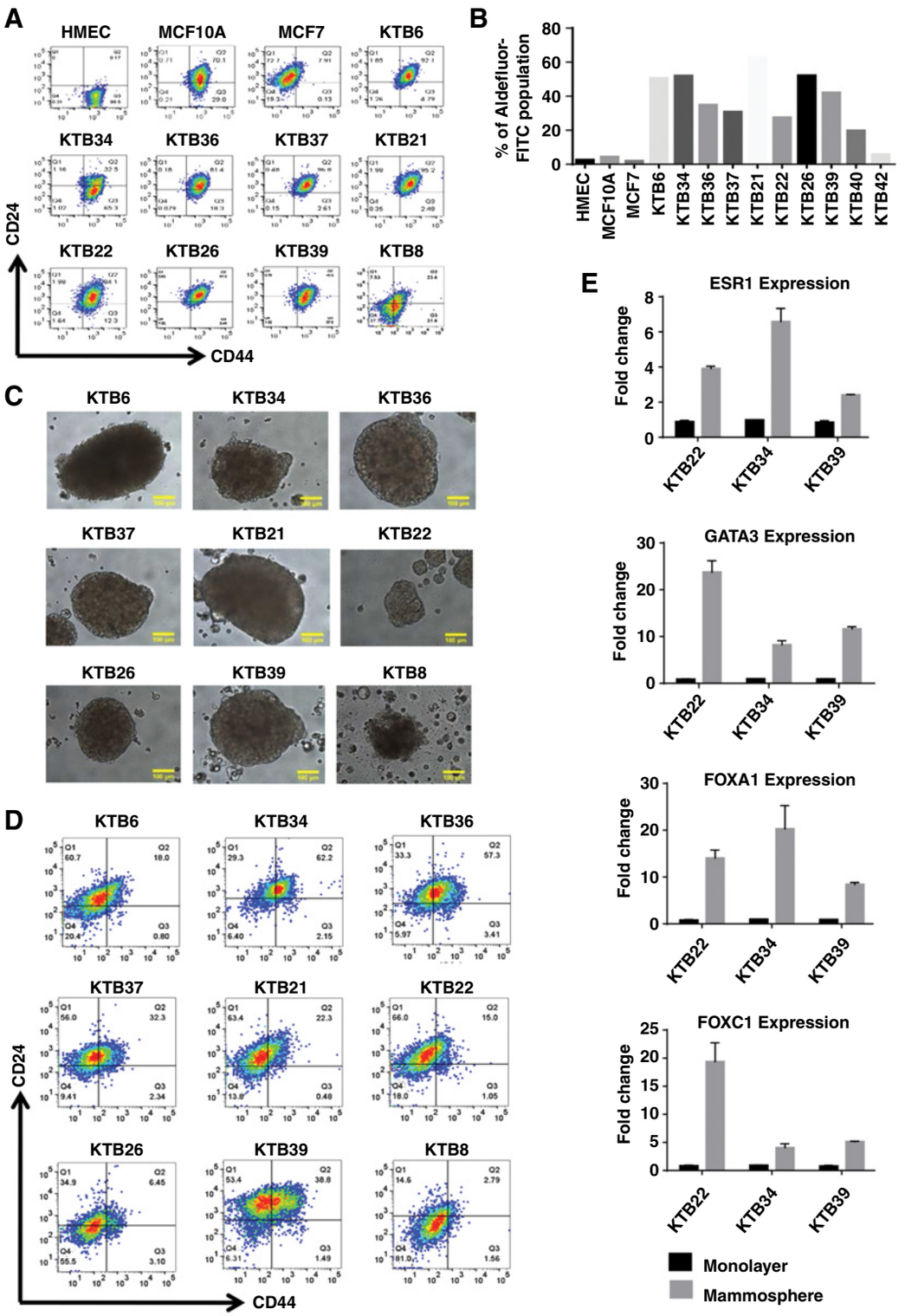


Figure 4.

KTB cell lines show different levels of $CD44^{+}/CD24^{-}$ and $ALDEFLUOR^{+}$ stem cells based on culture condition. **A,** CD44 and CD24 staining pattern of KTB cell lines. **B,** $ALDEFLUOR^{+}$ stem cells in different cell lines. **C,** KTB cell lines form mammospheres of variable sizes. **D,** Cells in mammospheres show different levels of differentiation based on CD44/CD24 staining pattern. **E,** Cells in mammospheres show elevated levels of stem/basal (*FOXC1*) and luminal differentiation markers (*ESR1, FOXA1,* and *GATA3*).

Downloaded from cancerres.aacrjournals.org on September 4, 2018. © 2018 American Association for Cancer Research.

Published OnlineFirst July 11, 2018; DOI: 10.1158/0008-5472.CAN-18-0509

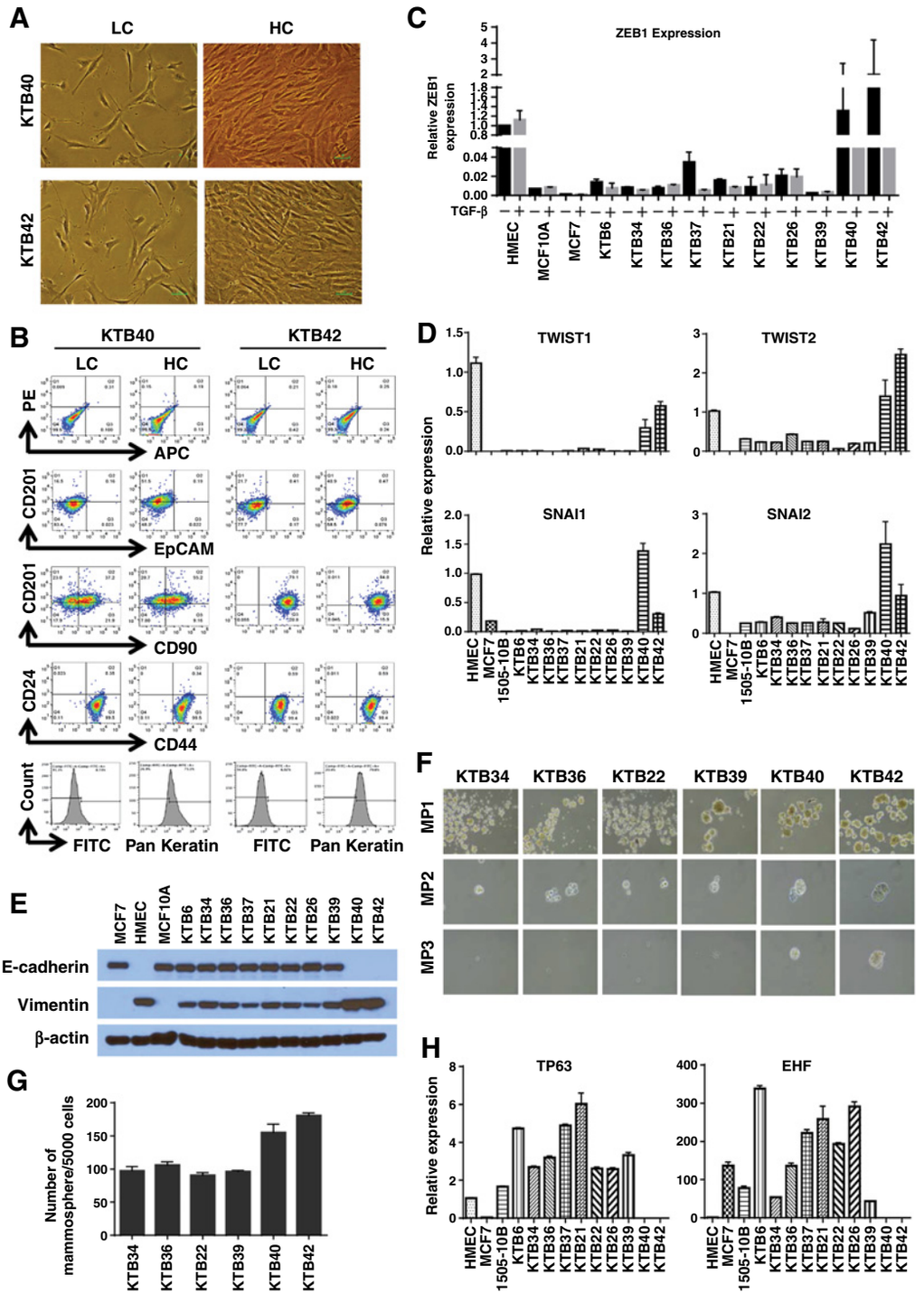


Figure 5.

CD201$^+$/EpCAM$^-$ cells express pan-keratin but display features of EMT. **A,** Confluence-dependent changes in the morphology of CD201$^+$/EpCAM$^-$ cells. **B,** CD24, CD44, CD90, CD201, EpCAM, and pan-keratin staining pattern of cell lines. **C,** *ZEB1* expression levels in various KTB cell lines with and without TGFβ1 treatment. *ZEB1* expression in KTB40 and KTB42 was significantly higher than in other KTB cell lines ($P < 0.0001$). **D,** *TWIST1, TWIST2, SNAI1,* and *SNAI2* expression levels in various KTB cell lines. A *TWIST1, TWIST2, SNAI1,* and *SNAI2* expression in KTB40 and KTB42 was significantly higher than in other KTB cell lines. **E,** E-Cadherin and vimentin expression levels in various KTB cell lines. **F,** Self-renewal capacity of various cell lines as measured by mammosphere assay. Primary mammospheres of KTB40 and KTB42 but not other KTB cell lines formed tertiary mammospheres. MP1, mammosphere passage 1; MP2 mammosphere passage 2; MP3, mammosphere passage 3. **G,** Mammosphere-forming efficiency of various KTB cell lines. KTB40 and KTB42 cell lines were more efficient than other cell lines in generating mammospheres. **H,** *TP63* and *EHF* expression levels in various KTB cell lines.

Downloaded from cancerres.aacrjournals.org on September 4, 2018. © 2018 American Association for Cancer Research.

Published OnlineFirst July 11, 2018; DOI: 10.1158/0008-5472.CAN-18-0509

cell line were double positive. Interestingly, although Western blotting did not detect E-cadherin expression, a small fraction of KTB42 expressed E-cadherin (Supplementary Fig. S5A).

Because our cell lines were derived from core biopsies of healthy women, there is a possibility that a fraction of cells in our cell lines display myoepithelial features. $CD10^+/EpCAM^-$ cells are considered myoepithelial cells, although a recent study showed CD10 positivity of cancer associated fibroblasts (49, 50). Although KTB40 and KTB42 were $CD10^+/EpCAM^-$, the remaining cell lines showed variable levels of CD10 positivity without any distinct subpopulation with $CD10^+/EpCAM^-$ phenotype (Supplementary Fig. S5B).

Interindividual differences in basal activity of select signaling networks

Because both transcriptome and phenotypic analyses documented interindividual heterogeneity in cell lines, we next examined whether basal signaling events were also different in these cell lines. We measured phospho-ERK, phospho-AKT1, phospho-AKT2, OCT-1, AP-1, and NF-κB DNA binding activity because intrinsic subtypes are known to differ in activity of signaling molecules (51). Phospho-ERK levels varied between cell lines with KTB40 and KTB42 containing highest levels of both phospho-ERK1 and the remaining cell lines showing variable levels of phospho-ERK2 but very low levels of phospho-ERK1 (Supplementary Fig. S6A). Phospho-AKT2 levels varied between cell lines, while phospho-AKT1 levels were nearly similar between cell lines (Supplementary Fig. S6A). NF-κB DNA binding activity also showed cell line variability with KTB40 and KTB42 showing predominant p65:p50 DNA binding activity (Supplementary Fig. S6B). OCT-1 and AP-1 DNA binding activity was least variable among KTB cell lines with the exception of very little OCT-1 DNA binding activity in KTB40 and KTB42 cell lines (Supplementary Figs. S6C and S6D). These results suggest that few of the reported activation of signaling pathways in cancer could be due to higher baseline activity in cells from which cancer originated instead of acquired changes during transformation.

BET bromodomain (BRD) proteins have recently been identified as major regulators of oncogenic transcription factors, and BRD4 among them has been targeted therapeutically (52). Two isoforms of BRD4 with opposing functions in cancer progression have been described; a long isoform with tumor suppressor activity and a short isoform with prometastatic functions (53). KTB cell lines expressed different levels of both long and short forms of BRD4 (Supplementary Fig. S6E), which suggest that tumors originating from cells expressing short BRD4 isoform are predisposed for metastatic progression.

The effect of TGFβ on cell proliferation of KTB cell lines

TGFβ regulates various biological processes, including cell proliferation, extracellular matrix (ECM) synthesis, angiogenesis, immune response, apoptosis, and differentiation (54). TGFβ also inhibits proliferation of $ER\alpha^+$ normal breast epithelial cells, and it is believed that loss of this TGFβ-mediated growth inhibition is essential for the initiation of $ER\alpha^+$ breast cancer (55). To determine the effect of TGFβ on cell proliferation, BrdUrd-incorporation-ELISA cell proliferation assay was performed. TGFβ1 inhibited proliferation but response varied between cell lines with highest inhibition of KTB34, KTB36, KTB37, KTB21, KTB26, and moderate inhibition of KTB6, KTB22, and KTB39 (Fig. 6A). Note that KTB40 and 42, cell lines derived from $CD201^+/EpCAM^-$ cells, were least sensitive to TGFβ1-mediated growth inhibition. However, there was no correlation between intrinsic subtype classification of these cell lines and the level of response to TGFβ1. We further examined the effect of TGFβ1 on characteristics of the cells with moderate change in proliferation by analyzing the levels of CD49f, CD201, CD271, and EpCAM markers with and without TGFβ1 treatment. TGFβ1-treated KTB6, KTB21, KTB36, KTB37, and KTB39 cells contained higher proportion of cells with basal/stem cell characteristics compared with untreated cells as evident from a higher number of $CD49f^+/EpCAM^-$ and $CD271^+/EpCAM^-$ subpopulations (Fig. 6B and Supplementary Fig. S7). Also, TGFβ1 had cell line–specific effects on CD49f, CD201, CD271, and EpCAM expression (Supplementary Fig. S7). Overall, these results document interindividual variability in response to TGFβ1, which includes growth inhibition, dedifferentiation, or lack of response.

KTB cell lines generate ductal-like structure in 3D *in vitro* breast development assays

There is growing interest in developing an organoid culture system to evaluate growth characteristics and individualized drug screening (56). We used hydrogel and collagen gel 3D culture models to determine whether KTB cell lines could create structures similar to those found in normal breast (34, 35). In both hydrogel and collagen scaffolds assays, KTB cell lines formed a complex branched tubular structure when cultured for 10 to 12 days (Fig. 7A and B). These ductal structures had some resemblance to the interlobular ducts that form through the process of ductal initiation and branching (Fig. 7A and B). Although the outer layer of these structures contained either $KRT14^+$ or $KRT14^+/KRT19^+$ cells, the inner layer contained few cells that stained only for $KRT19^+$ cells (Fig. 7C and D). We further confirmed that cells in these structures are epithelial by staining for KRT17, which is expressed in all epithelial cell types, including mature luminal cells or $KRT14^+/KRT19^+$ luminal progenitor cells (Fig. 7E; ref. 57). Collectively, resources reported in this study and assays utilized to characterize individual cell lines highlight the need to use a panel of "normal" cells from multiple sources as controls in signaling studies.

Discussion

Breast cancer research over many decades benefited from the availability of various breast cancer cell lines that maintain genetic and transcriptional features of breast tumors (57). However, modeling features of normal breast epithelial cells *in vitro* has been exceedingly difficult because culture conditions favored outgrowth of cells with basal cell gene expression patterns with few showing luminal progenitor properties and almost all previously generated breast epithelial cell lines (listed in Supplementary Table S1, labeled table comparing similar studies), including those with gene expression patterns overlapping with intrinsic subtypes and luminal progenitor cells, were derived from reduction mammoplasty samples or normal tissues adjacent to tumors with aberrant genomes (13, 15, 16, 57). Moreover, these cell lines have not been characterized for interindividual heterogeneity because of limited numbers (2–4 immortalized cell lines per study) and have limited luminal characteristics. Thus, limitation in culturing methods and tissue resource severely restricted our ability to

Downloaded from cancerres.aacrjournals.org on September 4, 2018. © 2018 American Association for Cancer Research.

Published OnlineFirst July 11, 2018; DOI: 10.1158/0008-5472.CAN-18-0509

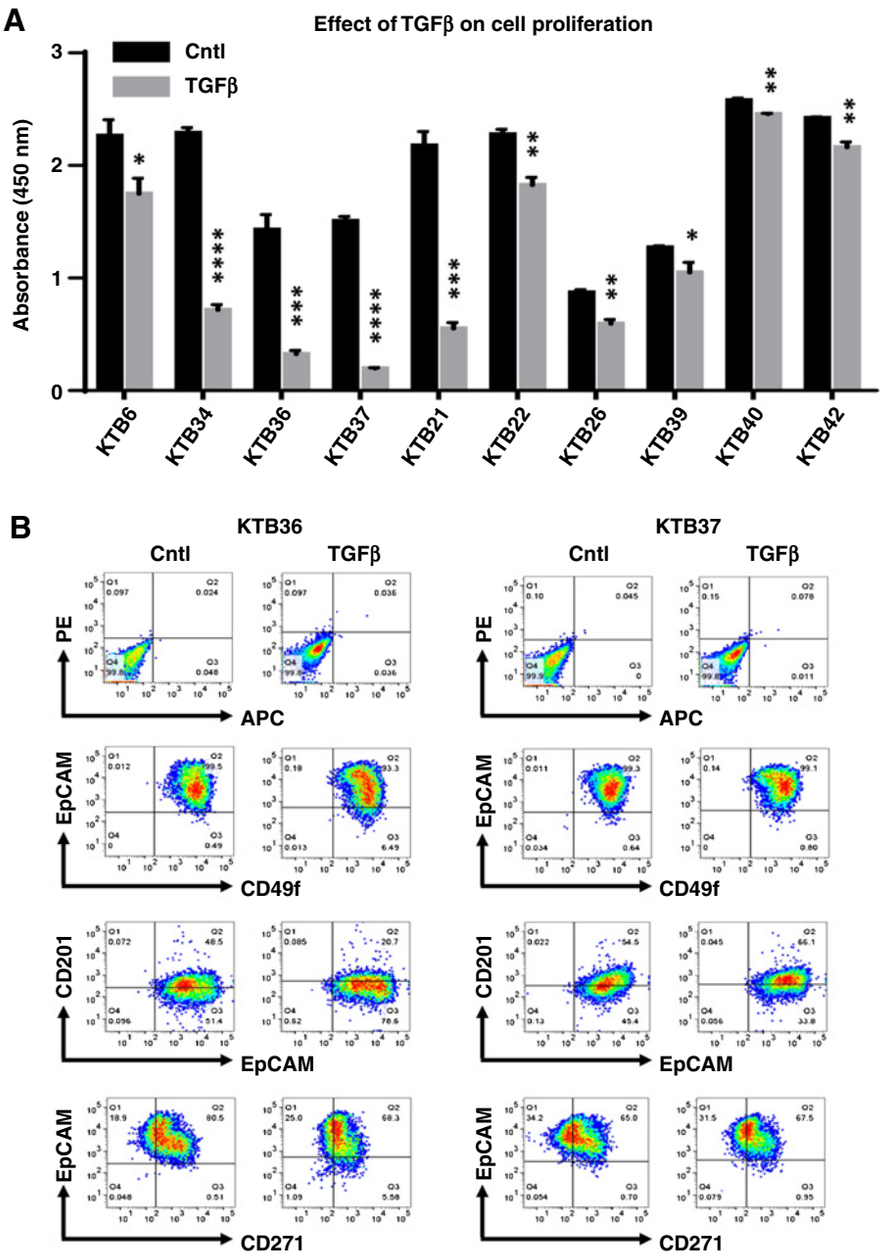


Figure 6.
The effect of TGFβ1 on the proliferation and differentiation of KTB cell lines. **A,** KTB cell lines show variable sensitivity to TGFβ1 in cell proliferation assay. The effects indicated by stars are statistically significant, but the extent of effects varied between cell lines (*, $P < 0.05$; **, $P < 0.004$; ***, $P < 0.0007$; ****, $P < 0.0001$). **B,** The effect of TGFβ1 on cell surface markers that define stem/progenitor/mature cell status.

mechanistically examine the role of specific genomic aberrations noted in luminal breast cancers, particularly ERα-positive breast cancers (58). For example, PIK3CA mutations are more frequently observed in ERα-positive breast cancers, but the mechanistic studies to identify downstream signaling events in nontransformed cells had to use basal-like MCF10A cells that do not express ERα (59). In addition, due to the heterogeneity noted in normal breast as well as the observation that most of the SNPs are located in gene regulatory regions affecting gene expression in *cis* or in *trans* (6), there is a need to use multiple isogenic cell lines with and without cancer-specific genome manipulations to identify cancer-specific signaling pathways.

KTB cell lines described in this article fulfill many of the needs of the research community. First, these cell lines are derived from biopsies of healthy women and several of these cell lines express p16(INK), which allows examination of oncogenic process by defined oncogenes with and without p16(INK)-mediated cell-cycle and senescence effects. Although most of the *in vitro* studies have suggested that loss of p16(INK) is necessary for immortalization and consequently for transformation (39), breast tumors in some patients express p16 (INK), indicating that its loss of expression is not always needed for transformation. Second, under mammosphere growth conditions, several of KTB cell lines differentiated into

Downloaded from cancerres.aacrjournals.org on September 4, 2018. © 2018 American Association for Cancer Research.

Published OnlineFirst July 11, 2018; DOI: 10.1158/0008-5472.CAN-18-0509

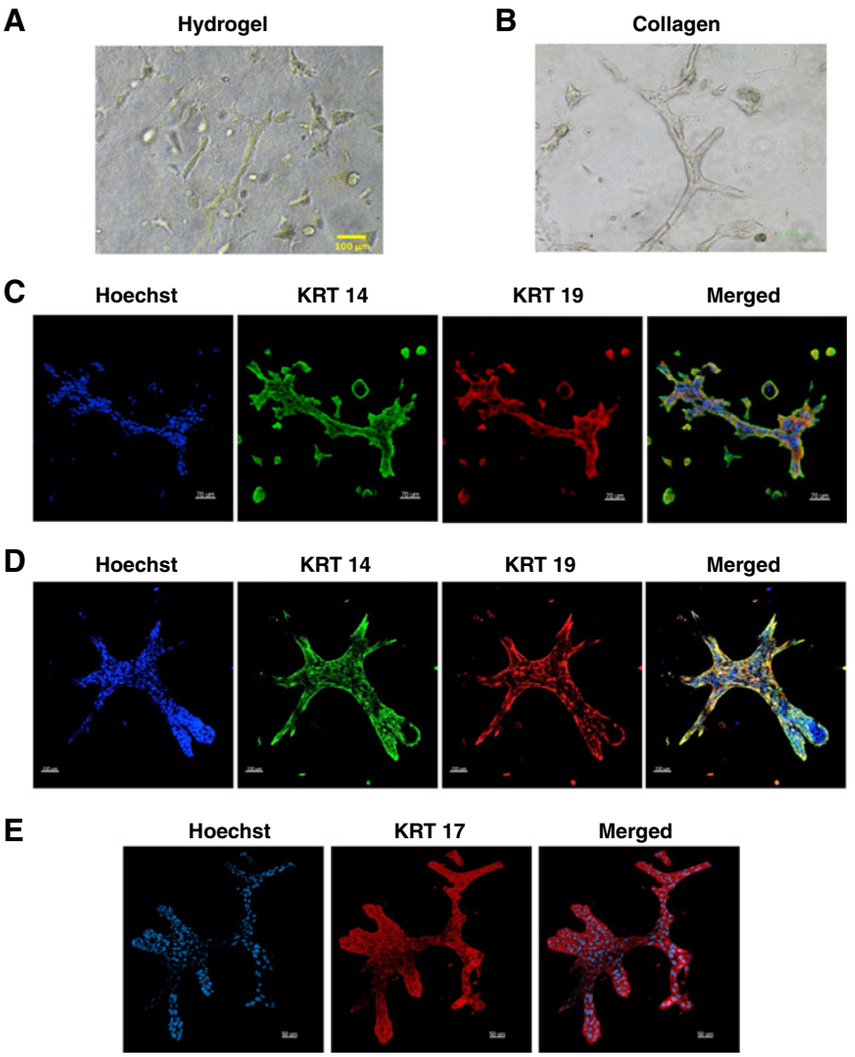


Figure 7.
KTB cell lines form ductal structures under 3D culture conditions. **A,** Phase contrast images of hydrogel cultures. **B,** Phase contrast images of collagen cultures. **C,** KRT14 and KRT19 staining patterns of cells in hydrogel cultures. **D,** KRT14 and KRT19 staining patterns of collagen cultures. **E,** KRT17 staining patterns of collagen cultures.

cells that express higher levels of luminal restricted transcription factors such as *ESR1*, *FOXA1*, and *GATA3* as well as $CD49f^{-}/EpCAM^{+}$ and $CD44^{-}/CD24^{+}$ phenotypes. Recent genomic analyses of several nontransformed and breast cancer cell lines have revealed nonexistence of cell lines corresponding to luminal A breast cancer (57). To our knowledge, immortalized breast epithelial cell lines expressing luminal markers have not been described, but we have for the first time demonstrated that the gene expression pattern in one of the KTB cell lines (KTB34) is comparable to luminal A breast cancer. Third, we have identified cell lines representing basal-like breast cancer. Thus, cell lines with luminal- and basal-enriched gene expression patterns can now be used to generate isogenic cell lines that overexpress amplicons, with deletion of chromosomal regions or mutations unique to specific subtypes of breast cancer. For example, $ER^{+}/HER2^{-}$ tumors are enriched for amplification of *CCND1* and *FGFR1* and mutations in *PIK3CA* and *ESR1* (60). $HER2/ERBB2^{+}$ breast cancers have been classified into at least two distinct subtypes: one with luminal-enriched gene expression and other with basal-enriched gene expression including strong coregulation of hypoxia/EGFR/TNFα/TGFβ/STAT3 pathways (61). KTB cell lines showed differences in response to TGFβ1, baseline DNA activity of transcription factors such as NF-κB and expression levels of pERK, pAKT2, and BRD4. ERBB2 overexpression studies in multiple "normal" cell lines would help to establish the role of cell-type origin in determining pathways that are coregulated along with this oncogene. Fourth, in a limited number of cases, we have cells prior to immortalization that have been cryopreserved and gene expression was analyzed, which can be used for sequential characterization of genomic/epigenomic changes during immortalization and transformation.

Claudin-low breast cancers and cell lines contain highly motile cells with fibroblastic features that have lost the expression of epithelial markers and at the same time gained a gene expression pattern similar to bone marrow–derived mesenchymal stem cells (57). Cell lines derived from $CD201^{+}/EpCAM^{-}$ cells, which are enriched in African American women (27), displayed fibroblastic and stromal features together with high expression level of *ZEB1*, *TWIST1*, *TWIST2*, *SNAI1*, *SNAI2*,

Downloaded from cancerres.aacrjournals.org on September 4, 2018. © 2018 American Association for Cancer Research.

Published OnlineFirst July 11, 2018; DOI: 10.1158/0008-5472.CAN-18-0509

vimentin, and elevated NF-κB and ERK activity. These cell lines also showed enhanced stemness phenotype compared with other KTB cell lines and have partial epithelial characteristics, as E-cadherin expression was detectable in a fraction of cells. ZEB1 has previously been shown to induce claudin-low mammary tumors and we had demonstrated the ability of NF-κB to induce ZEB1 (62, 63). Collectively, these results suggest that $CD201^+/EpCAM^-$ cells are the cell-type origin of claudin-low breast cancers with inherent features of EMT and elevated basal NF-κB activity. Triple-negative breast cancers (TNBC) have recently been classified into six subgroups with mesenchymal stem–like subgroup showing resemblance to claudin-low breast cancers (64). This TNBC subtype also contains elevated NF-κB activity and show EMT features. African American women develop basal-like and mesenchymal stem-like TNBC with higher levels of intratumor heterogeneity compared with white women (65) and elevated presence of $CD201^+/EpCAM^-$ cells in their normal breast may contribute to an increase in mesenchymal stem–like breast cancers.

Overall, the panel of KTB cell lines described in this study would allow us to ask several fundamental questions: (i) to what extent interindividual differences in baseline signaling activity affect cancer-specific signaling? (ii) can we predict the behavior of tumor based on cell-type origin, which can be linked back to a specific intrinsic subtype? and (iii) how do we define normal gene expression pattern and expression in how many "normal" cell lines need to be measured as controls for cancer studies? In this respect, Susan G Komen normal tissue bank at Indiana University has cryopreserved >1,000 breast biopsies of healthy women with richly annotated demographics and basic health data and we have a >90% success rate in generating primary cells from cryopreserved samples. Having demonstrated the feasibility of using the cryopreserved tissues to generate cell lines, these resources can be used in the future to generate additional cell lines and expand the number of "control" cell lines needed for breast cancer research.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Authors' Contributions

Conception and design: H. Nakshatri
Development of methodology: B. Kumar, M. Prasad, P. Bhat-Nakshatri, J. Wan, Y. Liu, H. Nakshatri
Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): B. Kumar, M. Prasad, P. Bhat-Nakshatri, M. Anjanappa, M. Kalra, A.M. Storniolo, Y. Liu, H. Nakshatri
Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): B. Kumar, M. Prasad, P. Bhat-Nakshatri, M. Kalra, X. Rao, S. Liu, J. Wan, Y. Liu, H. Nakshatri
Writing, review, and/or revision of the manuscript: B. Kumar, M. Prasad, A.M. Storniolo, J. Wan, Y. Liu, H. Nakshatri
Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): P. Bhat-Nakshatri, M. Anjanappa, N. Marino, H. Nakshatri
Study supervision: H. Nakshatri

Acknowledgments

We thank Dr. Vimla Band (University of Nebraska) for hTERT vector, flow cytometry, and tissue procurement cores at the IU Simon Cancer Center and Susan G Komen Tissue Bank for various tissues and reagents. We also thank the countless number of women for donating their breast tissue for research purpose as well as volunteers who facilitated tissue collection. This work is supported by DOD-W81XWH-15-1-0707 and R03CA195250-01A1 to H. Nakshatri. Susan G. Komen for the Cure, Breast Cancer Research Foundation and Vera Bradley Foundation for Breast Cancer Research provided funding support to Komen Normal Tissue Bank. The Walther Cancer Foundation provided funding support for the Collaborative Core for Cancer Bioinformatics (C3B), which performed RNA-seq data analyses and PAM50 classification.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received February 14, 2018; revised May 30, 2018; accepted July 6, 2018; published first July 11, 2018.

References

1. Visvader JE, Stingl J. Mammary stem cells and the differentiation hierarchy: current status and perspectives. Genes Dev 2014;28:1143–58.
2. Raouf A, Zhao Y, To K, Stingl J, Delaney A, Barbara M, et al. Transcriptome analysis of the normal human mammary cell commitment and differentiation process. Cell Stem Cell 2008;3:109–18.
3. Villadsen R, Fridriksdottir AJ, Ronnov-Jessen L, Gudjonsson T, Rank F, LaBarge MA, et al. Evidence for a stem cell hierarchy in the adult human breast. J Cell Biol 2007;177:87–101.
4. Pellacani D, Bilenky M, Kannan N, Heravi-Moussavi A, Knapp DJHF, Gakkhar S, et al. Analysis of normal human mammary epigenomes reveal cell-specfiic active enahcner states and associated transcription factor networks. Cell reports 2016;17:2060–74.
5. Santagata S, Thakkar A, Ergonul A, Wang B, Woo T, Hu R, et al. Taxonomy of breast cancer based on normal cell phenotype predicts outcome. J Clin Invest 2014;124:859–70.
6. Consortium GT, Laboratory DA, Coordinating Center -Analysis Working G, Statistical Methods groups-Analysis Working G, Enhancing Gg, Fund NIHC, , et al. et al. Genetic effects on gene expression across human tissues. Nature 2017;550:204–13.
7. Lappalainen T, Sammeth M, Friedlander MR, t Hoen PA, Monlong J, Rivas MA, et al. Transcriptome and genome sequencing uncovers functional variation in humans. Nature 2013;501:506–11.
8. Wagner JR, Busche S, Ge B, Kwan T, Pastinen T, Blanchette M. The relationship between DNA methylation, genetic and expression interindividual variation in untransformed human fibroblasts. Genome Biol 2014;15:R37.
9. Sorlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. Proc Natl Acad Sci U S A 2001;98:10869–74.
10. Prat A, Parker JS, Karginova O, Fan C, Livasy C, Herschkowitz JI, et al. Phenotypic and molecular characterization of the claudin-low intrinsic subtype of breast cancer. Breast Cancer Res 2010;12:R68.
11. Prat A, Perou CM. Mammary development meets cancer genomics. Nat Med 2009;15:842–4.
12. Keller PJ, Arendt LM, Skibinski A, Logvinenko T, Klebba I, Dong S, et al. Defining the cellular precursors to human breast cancer. Proc Natl Acad Sci U S A 2012;109:2772–7.
13. Ince TA, Richardson AL, Bell GW, Saitoh M, Godar S, Karnoub AE, et al. Transformation of different human breast epithelial cell types leads to distinct tumor phenotypes. Cancer Cell 2007;12:160–70.
14. Kao J, Salari K, Bocanegra M, Choi YL, Girard L, Gandhi J, et al. Molecular profiling of breast cancer cell lines defines relevant tumor models and provides a resource for cancer gene discovery. PloS One 2009;4:e6146.
15. Degnim AC, Visscher DW, Hoskin TL, Frost MH, Vierkant RA, Vachon CM, et al. Histologic findings in normal breast tissues: comparison to reduction mammaplasty and benign breast disease tissues. Breast Cancer Res Treat 2012;133:169–77.

Downloaded from cancerres.aacrjournals.org on September 4, 2018. © 2018 American Association for Cancer Research.

Published OnlineFirst July 11, 2018; DOI: 10.1158/0008-5472.CAN-18-0509

16. Teschendorff AE, Gao Y, Jones A, Ruebner M, Beckmann MW, Wachter DL, et al. DNA methylation outliers in normal breast tissue identify field defects that are enriched in cancer. Nat Commun 2016;7:10478.
17. Fridriksdottir AJ, Villadsen R, Morsing M, Klitgaard MC, Kim J, Petersen OW, et al. Proof of region-specific multipotent progenitors in human breast epithelia. Proc Natl Acad Sci U S A 2017;114:E10102–E11.
18. Hopkinson BM, Klitgaard MC, Petersen OW, Villadsen R, Ronnov-Jessen L, Kim J. Establishment of a normal-derived estrogen receptor-positive cell line comparable to the prevailing human breast cancer subtype. Oncotarget 2017;8:10580–93.
19. Fridriksdottir AJ, Kim J, Villadsen R, Klitgaard MC, Hopkinson BM, Petersen OW, et al. Propagation of oestrogen receptor-positive and oestrogen-responsive normal human breast cells in culture. Nat Commun 2015;6:8786.
20. Zhao X, Malhotra GK, Lele SM, Lele MS, West WW, Eudy JD, et al. Telomerase-immortalized human mammary stem/progenitor cells with ability to self-renew and differentiate. Proc Natl Acad Sci U S A 2010; 107:14146–51.
21. Arendt LM, Keller PJ, Skibinski A, Goncalves K, Naber SP, Buchsbaum RJ, et al. Anatomical localization of progenitor cells in human breast tissue reveals enrichment of uncommitted cells within immature lobules. Breast Cancer Res 2014;16:453.
22. Zhao X, Malhotra GK, Band H, Band V. Derivation of myoepithelial progenitor cells from bipotent mammary stem/progenitor cells. PloS One 2012;7:e35338.
23. Carter EP, Gopsill JA, Gomm JJ, Jones JL, Grose RP. A 3D in vitro model of the human breast duct: a method to unravel myoepithelial-luminal interactions in the progression of breast cancer. Breast Cancer Res 2017;19:50.
24. Sun W, Kang KS, Morita I, Trosko JE, Chang CC. High susceptibility of a human breast epithelial cell type with stem cell characteristics to telomerase activation and immortalization. Cancer Res 1999;59:6118–23.
25. Gudjonsson T, Villadsen R, Nielsen HL, Ronnov-Jessen L, Bissell MJ, Petersen OW. Isolation, immortalization, and characterization of a human breast epithelial cell line with stem cell properties. Genes Dev 2002; 16:693–706.
26. Chanson L, Brownfield D, Garbe JC, Kuhn I, Stampfer MR, Bissell MJ, et al. Self-organization is a dynamic and lineage-intrinsic property of mammary epithelial cells. Proc Natl Acad Sci U S A 2011;108:3264–9.
27. Nakshatri H, Anjanappa M, Bhat-Nakshatri P. Ethnicity-Dependent and -Independent Heterogeneity in Healthy Normal Breast Hierarchy Impacts Tumor Characterization. Sci Rep 2015;5:13526.
28. Dong J, Hu Y, Fan X, Wu X, Mao Y, Hu B, et al. Single-cell RNA-seq analysis unveils a prevalent epithelial/mesenchymal hybrid state during mouse organogenesis. Genome Biol 2018;19:31.
29. Nievergelt CM, Maihofer AX, Shekhtman T, Libiger O, Wang X, Kidd KK, et al. Inference of human continental origin and admixture proportions using a highly discriminative ancestry informative 41-SNP panel. Investig Genet 2013;4:13.
30. Breese MR, Liu Y. NGSUtils: a software suite for analyzing and manipulating next-generation sequencing datasets. Bioinformatics 2013;29:494–6.
31. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 2010;26:139–40.
32. Gendoo DM, Ratanasirigulchai N, Schroder MS, Pare L, Parker JS, Prat A, et al. Genefu: an R/Bioconductor package for computation of gene expression-based signatures in breast cancer. Bioinformatics 2016;32:1097–9.
33. Bhat-Nakshatri P, Sweeney CJ, Nakshatri H. Identification of signal transduction pathways involved in constitutive NF-kappaB activation in breast cancer cells. Oncogene 2002;21:2066–78.
34. Linnemann JR, Miura H, Meixner LK, Irmler M, Kloos UJ, Hirschi B, et al. Quantification of regenerative potential in primary human mammary epithelial cells. Development 2015;142:3239–51.
35. Miller DH, Sokol ES, Gupta PB. 3D primary culture model to study human mammary development. Methods Mol Biol 2017;1612:139–47.
36. Eswaran J, Li DQ, Shah A, Kumar R. Molecular pathways: targeting p21-activated kinase 1 signaling in cancer–opportunities, challenges, and limitations. Clin Cancer Res 2012;18:3743–9.
37. Eeckhoute J, Keeton EK, Lupien M, Krum SA, Carroll JS, Brown M. Positive cross-regulatory loop ties GATA-3 to estrogen receptor alpha expression in breast cancer. Cancer Res 2007;67:6477–83.
38. Han B, Bhowmick N, Qu Y, Chung S, Giuliano AE, Cui X. FOXC1: an emerging marker and therapeutic target for cancer. Oncogene 2017; 36:3957–63.
39. Garbe JC, Vrba L, Sputova K, Fuchs L, Novak P, Brothman AR, et al. Immortalization of normal human mammary epithelial cells in two steps by direct targeting of senescence barriers does not require gross genomic alterations. Cell Cycle 2014;13:3423–35.
40. 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 2014;517:81–84.
41. Kim J, Villadsen R, Sorlie T, Fogh L, Gronlund SZ, Fridriksdottir AJ, et al. Tumor initiating but differentiated luminal-like breast cancer cells are highly invasive in the absence of basal-like activity. Proc Natl Acad Sci U S A 2012;109:6124–9.
42. Liu S, Cong Y, Wang D, Sun Y, Deng L, Liu Y, et al. Breast Cancer Stem Cells Transition between Epithelial and Mesenchymal States Reflective of their Normal Counterparts. Stem Cell Rep 2014;2:78–91.
43. Antony J, Huang RY. AXL-Driven EMT State as a Targetable Conduit in Cancer. Cancer Res 2017;77:3725–32.
44. Pan D, Roy S, Gascard P, Zhao J, Chen-Tanyolac C, Tlsty TD. SOX2, OCT3/4 and NANOG expression and cellular plasticity in rare human somatic cells requires CD73. Cell Signal 2016;28:1923–32.
45. Caramel J, Ligier M, Puisieux A. Pleiotropic roles for ZEB1 in Cancer. Cancer Res 2017.
46. Nieto MA, Huang RY, Jackson RA, Thiery JP. Emt: 2016. Cell 2016;166: 21–45.
47. Ye X, Tam WL, Shibue T, Kaygusuz Y, Reinhardt F, Ng Eaton E, et al. Distinct EMT programs control normal mammary stem cells and tumour-initiating cells. Nature 2015;525:256–60.
48. Pastushenko I, Brisebarre A, Sifrim A, Fioramonti M, Revenco T, Boumahdi S, et al. Identification of the tumour transition states occurring during EMT. Nature 2018;556:463–68.
49. Morel AP, Ginestier C, Pommier RM, Cabaud O, Ruiz E, Wicinski J, et al. A stemness-related ZEB1-MSRB3 axis governs cellular pliancy and breast cancer genome stability. Nat Med 2017;23:568–78.
50. Su S, Chen J, Yao H, Liu J, Yu S, Lao L, et al. CD10(+)GPR77(+) Cancer-Associated Fibroblasts Promote Cancer Formation and Chemoresistance by Sustaining Cancer Stemness. Cell 2018;172:841–56 e16.
51. Huang KL, Li S, Mertins P, Cao S, Gunawardena HP, Ruggles KV, et al. Proteogenomic integration reveals therapeutic targets in breast cancer xenografts. Nat Commun 2017;8:14864.
52. Shu S, Polyak K. BET Bromodomain Proteins as Cancer Therapeutic Targets. Cold Spring Harb Symp Quant Biol 2016;81:123–29.
53. Alsarraj J, Walker RC, Webster JD, Geiger TR, Crawford NP, Simpson RM, et al. Deletion of the proline-rich region of the murine metastasis susceptibility gene Brd4 promotes epithelial-to-mesenchymal transition- and stem cell-like conversion. Cancer Res 2011;71:3121–31.
54. Heldin CH, Vanlandewijck M, Moustakas A. Regulation of EMT by TGFbeta in cancer. FEBS Lett 2012;586:1959–70.
55. Ewan KB, Oketch-Rabah HA, Ravani SA, Shyamala G, Moses HL, Barcellos-Hoff MH. Proliferation of estrogen receptor-alpha-positive mammary epithelial cells is restrained by transforming growth factor-beta1 in adult mice. Am J Pathol 2005;167:409–17.
56. Weeber F, Ooft SN, Dijkstra KK, Voest EE. Tumor organoids as a pre-clinical cancer model for drug discovery. Cell Chem Biol 2017;24:1092–100.
57. Prat A, Karginova O, Parker JS, Fan C, He X, Bixby L, et al. Characterization of cell lines derived from breast cancers and normal mammary tissues for the study of the intrinsic molecular subtypes. Breast Cancer Res Treat 2013;142:237–55.
58. Cancer Genome Atlas Research N. Comprehensive molecular characterization of gastric adenocarcinoma. Nature 2014;513:202–9.
59. Gustin JP, Karakas B, Weiss MB, Abukhdeir AM, Lauring J, Garay JP, et al. Knockin of mutant PIK3CA activates multiple oncogenic pathways. Proc Natl Acad Sci U S A 2009;106:2835–40.
60. Turner NC, Neven P, Loibl S, Andre F. Advances in the treatment of advanced oestrogen-receptor-positive breast cancer. Lancet 2017; 389:2403–14.
61. Gatza ML, Kung HN, Blackwell KL, Dewhirst MW, Marks JR, Chi JT. Analysis of tumor environmental response and oncogenic pathway activation identifies distinct basal and luminal features in HER2-related breast tumor subtypes. Breast Cancer Res 2011;13:R62.

Downloaded from cancerres.aacrjournals.org on September 4, 2018. © 2018 American Association for Cancer Research.

Published OnlineFirst July 11, 2018; DOI: 10.1158/0008-5472.CAN-18-0509

62. Morel AP, Hinkal GW, Thomas C, Fauvet F, Courtois-Cox S, Wierinckx A, et al. EMT inducers catalyze malignant transformation of mammary epithelial cells and drive tumorigenesis towards claudin-low tumors in transgenic mice. PLoS Genet 2012;8:e1002723.
63. Chua HL, Bhat-Nakshatri P, Clare SE, Morimiya A, Badve S, Nakshatri H. NF-kappaB represses E-cadherin expression and enhances epithelial to mesenchymal transition of mammary epithelial cells: potential involvement of ZEB-1 and ZEB-2. Oncogene 2007;26:711–24.
64. Lehmann BD, Bauer JA, Chen X, Sanders ME, Chakravarthy AB, Shyr Y, et al. Identification of human triple-negative breast cancer subtypes and preclinical models for selection of targeted therapies. J Clin Invest 2011; 121:2750–67.
65. Keenan T, Moy B, Mroz EA, Ross K, Niemierko A, Rocco JW, et al. Comparison of the genomic landscape between primary breast cancer in African American versus white women and the association of racial differences with tumor recurrence. J Clin Oncol 2015;33:3621–7.

Downloaded from cancerres.aacrjournals.org on September 4, 2018. © 2018 American Association for Cancer Research.

Published OnlineFirst February 4, 2019; DOI: 10.1158/1078-0432.CCR-18-3427

Translational Cancer Mechanisms and Therapy

Clinical Cancer Research

Genetic Ancestry–dependent Differences in Breast Cancer–induced Field Defects in the Tumor-adjacent Normal Breast

Harikrishna Nakshatri[1,2,5], Brijesh Kumar[1], Heather N. Burney[4], Mary L. Cox[3,5], Max Jacobsen[3], George E. Sandusky[3,5], Crislyn D'Souza-Schorey[6], and Anna Maria V. Storniolo[5,7]

Abstract

Purpose: Genetic ancestry influences evolutionary pathways of cancers. However, whether ancestry influences cancer-induced field defects is unknown. The goal of this study was to utilize ancestry-mapped true normal breast tissues as controls to identify cancer-induced field defects in normal tissue adjacent to breast tumors (NATs) in women of African American (AA) and European (EA) ancestry.

Experimental Design: A tissue microarray comprising breast tissues of ancestry-mapped 100 age-matched healthy women from the Komen Tissue Bank (KTB) at Indiana University (Indianapolis, IN) and tumor-NAT pairs from 100 women (300 samples total) was analyzed for the levels of ZEB1, an oncogenic transcription factor that is central to cell fate, mature luminal cell–enriched estrogen receptor alpha (ERα), GATA3, FOXA1, and for immune cell composition.

Results: ZEB1$^+$ cells, which were localized surrounding the ductal structures of the normal breast, were enriched in the KTB-normal of AA compared with KTB-normal of EA women. In contrast, in EA women, both NATs and tumors compared with KTB-normal contained higher levels of ZEB1$^+$ cells. FOXA1 levels were lower in NATs compared with KTB-normal in AA but not in EA women. We also noted variations in the levels of GATA3, CD8$^+$ T cells, PD1$^+$ immune cells, and PDL1$^+$ cell but not CD68$^+$ macrophages in NATs of AA and EA women. ERα levels did not change in any of our analyses, pointing to the specificity of ancestry-dependent variations.

Conclusions: Genetic ancestry–mapped tissues from healthy individuals are required for proper assessment and development of cancer-induced field defects as early cancer detection markers. This finding is significant in light of recent discoveries of influence of genetic ancestry on both normal biology and tumor evolution.

Introduction

Recent data demonstrating a correlation between lymph node positivity at the time of detection, and the probability of disease recurrence even decades postdetection, only solidifies the principle that detection of breast cancer prior to lymph node metastasis can appreciably improve clinical outcomes (1). Although the last decade witnessed significant improvements in imaging technologies including 3D-mammography, false negatives remain a significant concern (2). One way to overcome these false negatives is to complement radiologic techniques with molecular assays that measure "transcriptomic and epigenetic field effect" of tumors on adjacent "normal" (NAT) tissues. Teschendorff and colleagues demonstrated tumor-induced epigenetic field defects in NATs specifically targeting transcription factor–binding sites specifying chromatin architecture and stem cell differentiation pathways (3). These include Wnt and FGF signaling networks. Unfortunately, the Tumor Genome Atlas (TCGA) of breast cancer utilized reduction mammoplasty or NATs as their controls in transcriptome analyses (4). These are often substituted for "normal" controls in comparative analyses with breast cancers. This limitation was highlighted in another study, which compared TCGA "normal" breast transcriptome with the transcriptome of epithelial cells from the breast of healthy women. Significant differences were noted between these 2 sources of normal tissues (5). Reduction mammoplasty samples are also histologically abnormal compared with breast tissues from healthy women (6).

While molecular markers of cancers, particularly gene expression signatures, are traditionally developed by comparing gene expression between available "normal" and cancer tissues, the possibility of genetic ancestry of samples having an impact on gene expression under normal and abnormal conditions is rarely taken into consideration. The effects of genetic ancestry on tumor evolution and gene expression are just beginning to be

[1]Department of Surgery, Indiana University School of Medicine, Indianapolis, Indiana. [2]Department of Biochemistry and Molecular Biology, Indiana University School of Medicine, Indianapolis, Indiana. [3]Department of Pathology and Laboratory Medicine, Indiana University School of Medicine, Indianapolis, Indiana. [4]Department of Biostatistics, Indiana University School of Medicine, Indianapolis, Indiana. [5]IU Simon Cancer Center, Indiana University School of Medicine, Indianapolis, Indiana. [6]Department of Biological Sciences, University of Notre Dame, South Bend, Indiana. [7]Department of Medicine, Indiana University School of Medicine, Indianapolis, Indiana.

Note: Supplementary data for this article are available at Clinical Cancer Research Online (http://clincancerres.aacrjournals.org/).

Corresponding Author: Harikrishna Nakshatri, Indiana University School of Medicine, C218C, 980 West Walnut St., Indianapolis, IN 46202. Phone: 317-278-2238; Fax: 317-274-0396; E-mail: hnakshat@iupui.edu

doi: 10.1158/1078-0432.CCR-18-3427

©2019 American Association for Cancer Research.

Downloaded from clincancerres.aacrjournals.org on May 5, 2019. © 2019 American Association for Cancer Research.

Published OnlineFirst February 4, 2019; DOI: 10.1158/1078-0432.CCR-18-3427

Translational Relevance

Breast cancer diagnosis prior to lymph node metastasis can appreciably improve clinical outcomes. While radiologic techniques have improved early diagnosis, molecular markers that can complement radiologic techniques are needed to improve specificity. This study aimed to investigate how both genetic ancestry and appropriate control tissues influence detection of cancer-induced changes in the breast. We show that alterations in ZEB1$^+$ cells in tissues surrounding tumors are observed predominantly in women of European (EA) ancestry, whereas FOXA1$^+$ cells were altered in normal tissues adjacent to tumors of women of African American (AA) ancestry. Immune cell activation in tumors, as well as surrounding tissue showed genetic ancestry–dependent variations as evident from differences in PD1$^+$ and PDL1$^+$ cells in the normal tissue adjacent to tumors of women of AA and EA ancestry. Thus, biomarker discovery needs to consider not only sample size and statistical methods but also genetic ancestry and true normal control tissues.

recognized (7). This observation is highly relevant in the context of known differences in cancer incidence and/or outcome based on genetic ancestry. For example, women of African American ancestry (AA) suffer higher mortality from the aggressive breast cancer subtype, triple-negative breast cancer (TNBC), than women of European ancestry (EA; ref. 8). In contrast, breast cancer in Hispanic and Native American women is less prevalent and these women have better outcomes (9, 10). Whether the worse outcome in AA women is due to an increased incidence of TNBC or unique biological factors that promote aggressive biology is an important but unresolved challenge in cancer disparities research. Dietze and colleagues (8) recently highlighted that key molecular pathways, including Aurora A-PLK, EZH2, and Wnt-stem cell signaling networks, are significantly upregulated in TNBCs of AA women compared with TNBCs of EA women. The review (8) further emphasized that it remains unknown whether genomic aberrations unique to TNBCs in AA women result in activation of these signaling pathways in tumors or whether the basal activity of these pathways in normal AA women's breasts is inherently different compared with EA women's breasts. It remains possible that normal breast biology varies based on genetic ancestry. Evidence for this possibility comes from a recent discovery of breast cancer protective alleles in Latinas (11). SNPs in the protective allele are located on gene regulatory regions affecting the expression of genes linked to differentiation. Our own studies have discovered enrichment of a unique population of cells in the normal breast of AA women (12). Furthermore, a breast cancer susceptibility locus in AA women, potentially altering the expression levels of miRNA miR-3065, has recently been described (13).

Here, we took advantage of genetic ancestry–mapped true normal breast tissues to identify differences between true normal and NATs. These differences can potentially be developed into the earliest markers of breast cancer initiation. A tissue microarray (TMA) comprising breast tissues from clinically normal breasts, NATs, and tumors were analyzed for markers that are expressed in cells with stem or mature luminal cell properties. We also examined the TMA for CD8$^+$ T cells, CD68$^+$ macrophages, PD1$^+$ immune cells, and PDL1$^+$ epithelial cells to determine whether immune cell composition of tumors and NATs in AA women differ from those of EA women.

Materials and Methods

Generation of TMA

Breast core biopsies from healthy women donated to the Komen Tissue Bank (KTB) at Indiana University (Indianapolis, IN) and surgical material left over after pathologic assessment as part of a treatment protocol were obtained after informed written consent from the subjects. All experiments were carried out in accordance with the approved guidelines of the Indiana University Institutional Review Board (Indianapolis, IN). International Ethical Guidelines for Biomedical Research Involving Human Subjects were followed. We created a TMA comprising healthy breast tissue from the KTB (KTB-normal), matched normal adjacent to tumor (NAT), and tumor tissue of approximately 50 each of AA and Caucasian women (total ~300 samples). KTB-normal tissues were age- and race-matched to NATs/tumors. Body mass index (BMI) of AA women who donated tissues to KTB was 32.3 ± 9, whereas it was 28.3 ± 8.5 in case of Caucasian women. Each sample was spotted in duplicate in cases of NATs and tumors.

IHC and statistical analyses

TMA was analyzed for ZEB1, MSRB3, estrogen receptor alpha (ERα), FOXA1, and GATA3 expression. All IHC was done in a Clinical Laboratory Improvement Amendments (CLIA)-certified histopathology laboratory and evaluated by 3 pathologists in a blinded manner. Quantitative measurements were done using the automated Aperio Imaging system and analysis was done using an FDA-approved algorithm. Positivity and H-scores were scored and statistically analyzed as described previously (14, 15). With respect to PD-L1, a tumor proportion score (TPS) was conducted alongside a pathology hand score. The TPS describes the ratio of positive viable tumor cells against all viable tumor cells. The PD-L1 TPS followed the prescribed FDA reading: a negative score having <1% positive staining on the cell membrane, a positive score having 1%–49% tumor cells partially or completely expressing PD-L1 on the cell membrane, and a strong positive having ≥50% tumor cells partially or completely expressing PD-L1 on the cell membrane at a stronger intensity (16, 17). Data were analyzed in 3 different ways: (i) expression differences between AA and EA KTB-normal; (ii) expression differences between KTB-normal and NATs; and (iii) expression differences between NATs and tumors. The statistical software SAS version 9.4 was used to complete the statistical analyses with $P < 0.05$ considered significant. Nonparametric Wilcoxon rank-sum tests were used for unpaired analyses, as positivity and H-scores were not normally distributed, whereas nonparametric Wilcoxon signed-rank tests were used for paired analyses. The following antibodies were used: CD8 (Dako IR623), CD68 KP1 (Dako IR609), ER clone:EP1 (Dako IR 084), FOXA1 (Santa Cruz Biotechnology sc-6553), GATA3 (Santa Cruz Biotechnology sc-268), MSRB3 (HPA014432, rabbit polyclonal, Sigma), PD1 (Cell Marque Corporation 315M-98), PDL1 (Keytruda; clone 22c3, Dako IHC 22c3), and ZEB1 (3G6, catalog No. 14-9741-82, eBioscience).

Downloaded from clincancerres.aacrjournals.org on May 5, 2019. © 2019 American Association for Cancer Research.

Published OnlineFirst February 4, 2019; DOI: 10.1158/1078-0432.CCR-18-3427

Results

ZEB1$^+$ cells are enriched in the normal breasts of AA women compared with EA women

In the mouse mammary gland, PROCR$^+$/EpCAM$^-$ cells are purported to function as multipotent stem cells (18). In our previous study focused on evaluating ethnicity-dependent differences in the normal breast, we observed specific enrichment of PROCR$^+$/EpCAM$^-$ cells in cultured normal breast epithelial cells from biopsies of healthy AA women compared with EA women (12). These cells are enriched for the expression of stemness-related transcription factor ZEB1 and have enhanced Wnt pathway activity compared with PROCR$^{+/-}$/EpCAM$^+$ cells (12). ZEB1 has recently been demonstrated to limit oncosuppressive p53-driven DNA damage response in stem cells and thus increase the stem cells' intrinsic susceptibility to malignant transformation (19). ZEB1$^+$ cells coexpress the methionine sulfoxidase reductase (MSRB3), which protects against DNA damage (19). These observations raised the possibility that PROCR$^+$/ZEB1$^+$ cells are naturally present at higher levels in the normal breasts of AA women and that failure to consider natural variation in gene expression pattern, influenced at least partially by genetic ancestry, could have an impact on identifying cancer-induced field effect on the adjacent normal breast. Measuring PROCR itself in the breast tissue is complicated because there are 4 haplotypes of PROCR due to SNPs and only 1 among them is a cell surface protein (20). Because ZEB1 expression is enriched in PROCR$^+$/EpCAM$^-$ cells, we used ZEB1 as a surrogate marker for PROCR$^+$/EpCAM$^-$ cells in unmanipulated breast tissues.

Representative IHC staining patterns of ZEB1 in KTB-normal, NATs, and tumors from AA and EA women are shown in Fig. 1A and statistical analyses are presented in Fig. 1B–D and in Table 1. Descriptive statistics of ethnicity, age, menstrual status, pregnancy and breastfeeding history, hormone replacement therapy, and family history of breast cancer for the KTB-normal cohort is shown in Supplementary Table S1. Highly discriminative 41-ancestry marker profiles of KTB-normal showed >75% African ancestry markers in samples from AA women and >80% EA ancestry markers in Caucasian women (Supplementary Fig. S1A; ref. 21). Characteristics of breast cancer in the tumor cohort are shown in Supplementary Table S2. ZEB1-expressing cells were localized outside the ductal structures of the normal breast and in the stromal part of the tumors (enlarged version on right side of Fig. 1A). KTB-normal breast tissue of AA women contained significantly higher levels of ZEB1$^+$ cells compared with KTB-normal breast of EA women (Fig. 1B). NATs of AA women showed a modest increase in ZEB1$^+$ cells compared with those of KTB-normal (Fig. 1C and D). The scenario was completely different in EA women; both NATs and tumors contained significantly higher levels of ZEB1$^+$ cells compared with KTB-normal tissue (Fig. 1C and D). NAT to tumor differences were noted only in EA women where an increase in ZEB1$^+$ cells was noted predominantly in ERα$^+$ tumors (Supplementary Table S3). Thus, ZEB1$^+$ cells are intrinsically higher in the normal breasts of AA women, whereas remarkably elevated ZEB1$^+$ cells in the breasts of EA women were observed only in the context of breast cancer. Increases in ZEB1$^+$ cells in KTB-normal tissue of AA women compared with EA women is less likely related to BMI differences. This was demonstrated by subdividing women above and below BMI of 30, irrespective of genetic ancestry; ZEB1 H-score but not positivity showed a marginal relationship ($P = 0.04$) to BMI above and below 30 (Supplementary Table S4).

MSRB3 has recently been shown to be one of the downstream transcriptional targets of ZEB1 and it cooperates with ZEB1 during transformation of stem-like cells (19). To correlate ZEB1 expression with its activity, we measured the levels of MSRB3 using the same antibody used in the above study. We could measure positivity but not H-score because of low-level expression. The expression pattern was similar to that of ZEB1, as cells surrounding the ducts showed expression (Supplementary Fig. S1B). However, KTB-normal tissues of AA and EA women expressed similar levels of MSRB3 (Table 1; Supplementary Fig. S1C), which could be due to regulation by other transcription factors or to the low expression levels, making data interpretation difficult. Furthermore, except for a modest change in expression in NATs compared with KTB-normal tissues, no other differences were noted (Supplementary Fig. S1C and S1D).

FOXA1 expression is lower in NATs of only AA women

FOXA1 serves as a pioneer factor that controls chromatin access of various nuclear receptors including ERα and controls the expression of genes enriched in luminal cells compared with basal cells (22–24). FOXA1 along with another pioneer factor GATA3 and ERα form a lineage-restricted hormone-responsive signaling network in the normal breast (25). While higher expression of FOXA1 in the primary tumor is associated with better outcome, its overexpression in metastatic and/or anti-estrogen–resistant tumors is associated with rewiring of ERα signaling and poor outcome (26–29). In addition, it is suggested that *FOXA1* gene is preferentially methylated in tumors of AA women (30). Because of its relative importance in breast cancer, we assessed our TMA for FOXA1 expression. Representative staining pattern of FOXA1 is shown in Fig. 2A and numerical values are presented in Table 1. While FOXA1 levels in KTB-normal tissues of AA women were modestly higher than in EA women, NATs of AA women had lower FOXA1 compared with KTB-normal tissues (Fig. 2B). Thus, tumors through their field effect may decrease FOXA1 in the surrounding breast tissues of AA women.

GATA3 levels are higher in KTB-normal of EA compared with AA women

We also examined expression levels of GATA3 to determine whether hormonal signaling networks show genetic ancestry–dependent variation. Consistent with this possibility, GATA3 H-score and positivity were higher in KTB-normal tissues of EA women compared with those of AA women (Table 1; Supplementary Fig. S2B). Furthermore, GATA3 is a likely candidate for cancer-induced field defects in EA women as its levels were significantly lower in NATs of EA but not AA women compared with their KTB-normal counterparts (Supplementary Fig. S2C).

ERα$^+$ cells remain stable

ERα$^+$ cells in the normal breast are considered to be highly differentiated nonproliferative cells and control proliferation of ERα$^-$ cells through paracrine mechanisms (31). Representative ERα staining pattern is shown in Fig. 3A and statistical analyses are presented in Fig. 3B–D and Table 1. Neither KTB-normal tissues nor NATs showed genetic ancestry–dependent differences in ERα levels. The results are not only relevant but also reassure that our TMA detects only specific changes.

Downloaded from clincancerres.aacrjournals.org on May 5, 2019. © 2019 American Association for Cancer Research.

Published OnlineFirst February 4, 2019; DOI: 10.1158/1078-0432.CCR-18-3427

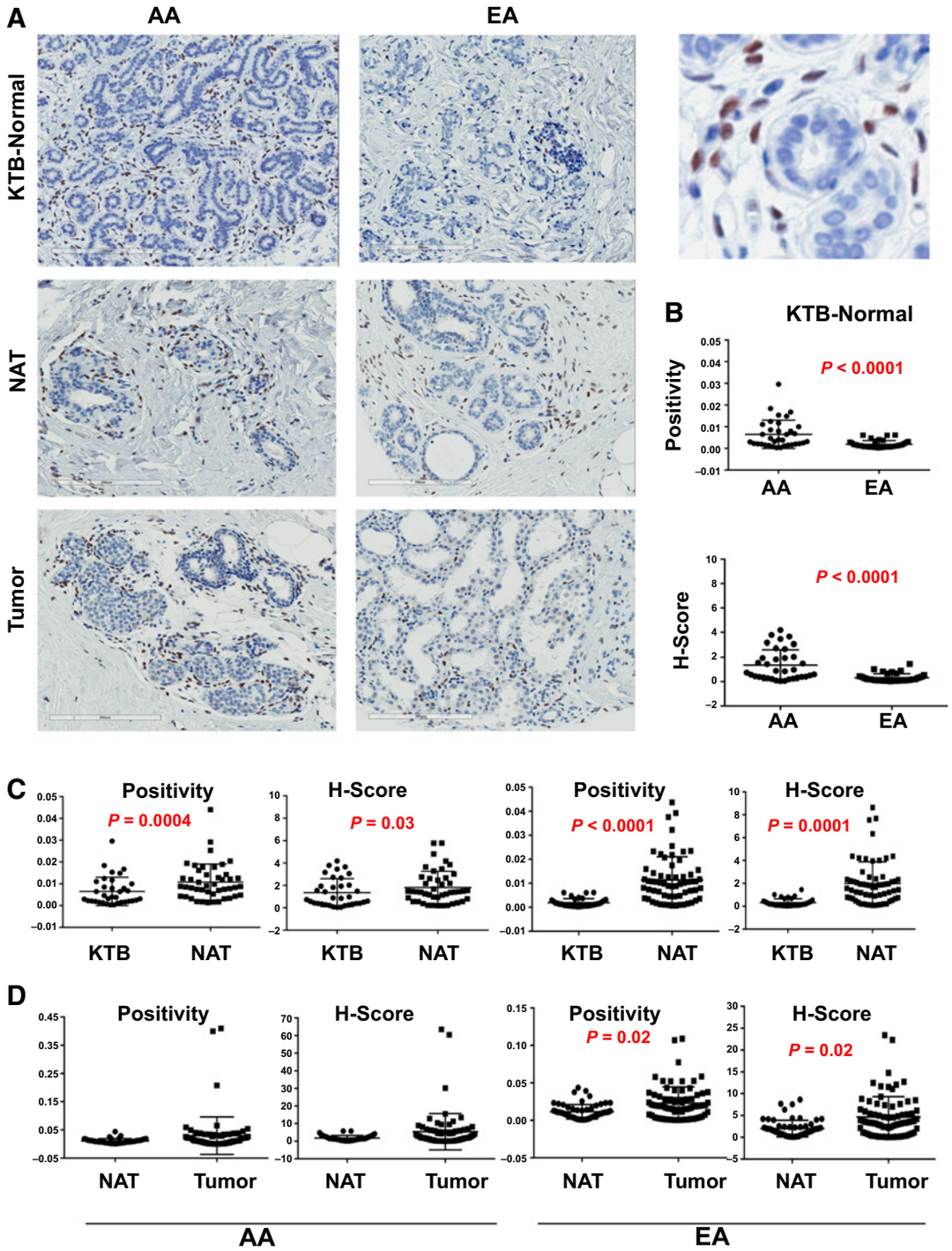


Figure 1.

ZEB1 expression pattern in KTB-normal, NATs, and in breast tumors. **A,** Representative IHC of KTB-normal, NATs, and tumors of women of AA and EA ancestry. Enlarged view of a KTB-normal is shown on right (top). **B,** Differences in ZEB1 expression (positivity and H-score) between KTB-normal of AA and EA women. **C,** Differences between KTB-normal and NATs in AA and EA women. **D,** Differences between NATs and tumors in AA and EA women.

ERα status in tumors influences differences between NATs and tumors

Although we observed differences in ZEB1, GATA3, and ERα expression between NATs and tumors (Figs. 1D, 2D, 3D; Supplementary Fig. S2D), interpretation of these data is difficult because of differences in characteristics of breast cancer subtypes, particularly ERα$^+$ and ERα$^-$ (32). To determine whether ERα$^+$ and ERα$^-$ tumors have distinct effects on NATs, FOXA1, GATA3, and

Downloaded from clincancerres.aacrjournals.org on May 5, 2019. © 2019 American Association for Cancer Research.

Published OnlineFirst February 4, 2019; DOI: 10.1158/1078-0432.CCR-18-3427

Table 1. Differences in expression levels of ERα, FOXA1, GATA3, MSRB3, and ZEB1 in KTB-normal between women of AA and EA ancestry

	African American				European ancestry				Two-sided
Variable name	***N***	**Median**	**Minimum**	**Maximum**	***N***	**Median**	**Minimum**	**Maximum**	**Wilcoxon test *P***
ER Positivity	38	0.009837	0.001769	0.085101	39	0.010387	0.000000	0.064659	0.8345
ER H-Score	38	2.165487	0.315269	21.777591	38	2.265046	0.103586	17.165665	0.7514
ZEB1 Positivity	38	0.004324	0.000316	0.025044	41	0.001224	0.000221	0.028532	<0.0001
ZEB H-Score	38	0.931903	0.045633	5.926867	41	0.157922	0.026349	3.927299	<0.0001
FOXA1 Positivity	42	0.037941	0.010844	0.147725	47	0.021856	0.007987	0.171964	0.0033
FOXA1 H-Score	42	5.108708	1.414022	20.066698	47	3.126083	1.033697	23.666012	0.0031
GATA3 Positivity	27	0.009031	0.001339	0.048353	32	0.018617	0.003970	0.067257	0.0009
GATA3 H-Score	27	1.656681	0.170409	10.399523	32	4.020432	0.589316	17.773060	0.0003
MSRB3 Positivity	29	0.006854	0.002061	0.035347	26	0.006474	0.002085	0.034037	0.4040

ZEB1 expression data in NATs and tumors were subdivided on the basis of ERα status of the tumor and reanalyzed. ZEB1-positivity and H-scores were higher in ERα$^+$ but not ERα$^-$ tumors compared with NATs (Supplementary Table S3). Despite small sample size, these differences were noted only in EA women with ERα$^+$ breast cancers (Supplementary Table S3). With respect to FOXA1, H-score but not positivity was marginally higher in ERα$^+$ tumors compared with NATs of EA women (Supplementary Table S3). ERα$^-$ tumors of EA but not AA women showed a significant decline in both positivity and H-score of FOXA1 compared with NATs (Supplementary Table S3). ERα$^+$ tumors but not ERα$^-$ tumors showed further increase in GATA3 positivity and H-scores in EA women, which further confirm the role of GATA3 in hormonal regulation of breast cancer (Supplementary Fig. S2D; Supplementary Table S3). When the analyses was done with paired NAT and tumors, the above noted differences between NATs and tumors in ZEB1, GATA3, and FOXA1 levels remained significant, although sample size was too small to subdivide samples based on genetic ancestry (Supplementary Table S5).

NATs of AA and EA women show differing levels of CD8, PD1, and PDL1$^+$ cells

Results thus far point to proinflammatory state of NATs of EA women based on the known link between ZEB1 and inflammatory cytokines (33). To address this further, we stained the above TMAs with CD8 for T cells, CD68 for macrophages, and PD1 for immune cells. We also examined epithelial/tumor cells for PDL1. All staining was done in a CLIA-certified laboratory with FDA-approved antibodies. In KTB-normal TMAs, there was no staining with CD8 and CD68 in either the AA or EA TMAs. Less than 1% of the lymphocytes and macrophages stained and these were considered negative. The same negativity was observed with PD1 and PDL1 immunostains (data not shown). Therefore, we analyzed staining results between NATs of AA and EA women and between NATs and tumors. Representative staining patterns in NATs and tumors are shown in Fig. 4.

CD8 immunostaining was localized to inflammatory cells (T lymphocytes) and not to tumor cells in the breast cancer cores. No background reactivity was observed in any case. NATs of AA women showed statistically significantly higher CD8 positivity compared with EA women (Fig. 5; Supplementary Table S6). The tumors in EA women had more CD8 immunostaining compared with corresponding NATs but such differences were not seen in the AA women.

CD68 staining was localized to macrophages in the breast cancer cores (Fig. 4). CD68 had lower positivity compared with CD8 by both visual and the Aperio-positive pixel reads. CD68 positivity was higher in tumors compared with their NATs ($P = 0.02$) in EA women but no such differences were noted in AA women (Fig. 5).

PD1 immunostaining was localized to immune cells only and no background staining was observed (Fig. 4). There was no staining of tumor cells. NATs of AA women contained significantly higher PD1$^+$ cells, similar to CD8$^+$ cells, compared with NATs of EA women (Fig. 6; Supplementary Table S6). PD1 staining did not show any differences between NATs and tumors in both groups (Supplementary Table S3).

PDL1 immunostaining was seen localized in the tumor cell cytoplasm and cell membrane (Fig. 4). In a few EA cases, only lymphocytes were stained. PDL1 staining of NATs of AA women was significantly lower than EA cases (Fig. 6; Supplementary Table S6). Although PDL1 staining did not differ between NATs and tumors of AA women, its levels were marginally lower in ER$^+$ tumors but not ER$^-$ tumors compared with NATs in case of EA women (Fig. 6; Supplementary Table S3). It is interesting that PD1 and PDL1 staining scores in NATs of AA is the reverse of the patterns seen in EA women. In summation, the immune environment in NATs is different from that in KTB-normal tissue with further differences between NATs and tumors, showing variations based on genetic ancestry.

Discussion

Recent studies have shown cancer-induced field defects influencing gene expression patterns in histologically normal tissues surrounding cancer (3, 34, 35). These observations raise a concern, as well as provide an opportunity for further investigation. The concern is the use of tumor adjacent normal as a "normal" control, whereas the opportunity pertains to the development of cancer-induced field defects in the adjacent normal as early markers of cancer. However, recent discovery of inter-individual differences in gene expression patterns due to SNPs in gene regulatory regions and genetic ancestry–dependent enrichment of SNPs with breast cancer protective or elevated risk characteristics necessitate the use of ancestry-matched control samples from healthy individuals to develop molecular features of tumor adjacent normal as cancer-initiation or progression markers (11, 13, 36, 37). Ethnicity contributing to inter-individual differences in normal biology is just beginning to be explored, as evident from a recent study that demonstrated distinct gut microbiota in different ethnic groups with shared geography (38). Furthermore, genetic ancestry has been shown to influence mutation patterns in cancer (7). Resources available at the Komen Tissue Bank at Indiana University (Indianapolis, IN), namely ancestry-mapped breast tissues from >5,000 healthy women, should enable us to take these factors into consideration as we

Downloaded from clincancerres.aacrjournals.org on May 5, 2019. © 2019 American Association for Cancer Research.

Published OnlineFirst February 4, 2019; DOI: 10.1158/1078-0432.CCR-18-3427

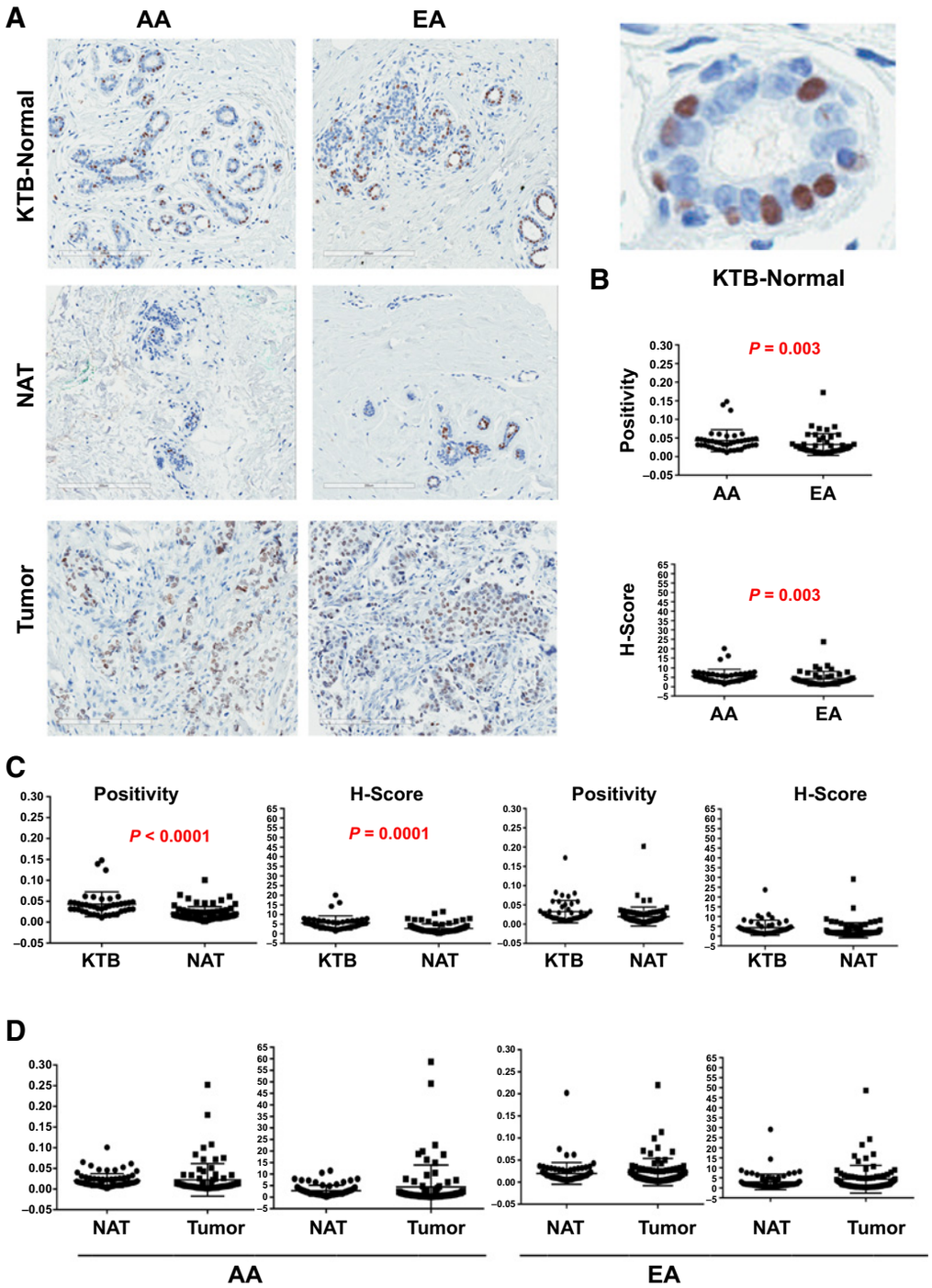


Figure 2.

FOXA1 expression pattern in KTB-normal, NATs, and in breast tumors. **A,** Representative IHC of KTB-normal, NATs, and tumors of AA and EA women. Enlarged view of a KTB-normal is shown on right (top). **B,** Differences in FOXA1 expression (positivity and H-score) between KTB-normal of AA and EA women. **C,** Differences between KTB-normal and NATs in AA and EA women. **D,** Differences between NATs and tumors in AA and EA women.

develop molecular features of NATs as cancer detection markers. Utilizing a small fraction of those tissues, we provide evidence for ancestry-dependent differences in the number of $ZEB1^+$ and $GATA3^+$ cells in the normal breast, as well as cancer-induced field effects on ZEB1, GATA3, and $FOXA^+$ cells in the tumor-adjacent normal tissue.

Recently discovered functions of ZEB1 have raised considerable interest in this molecule within the oncology field. The

Downloaded from clincancerres.aacrjournals.org on May 5, 2019. © 2019 American Association for Cancer Research.

Published OnlineFirst February 4, 2019; DOI: 10.1158/1078-0432.CCR-18-3427

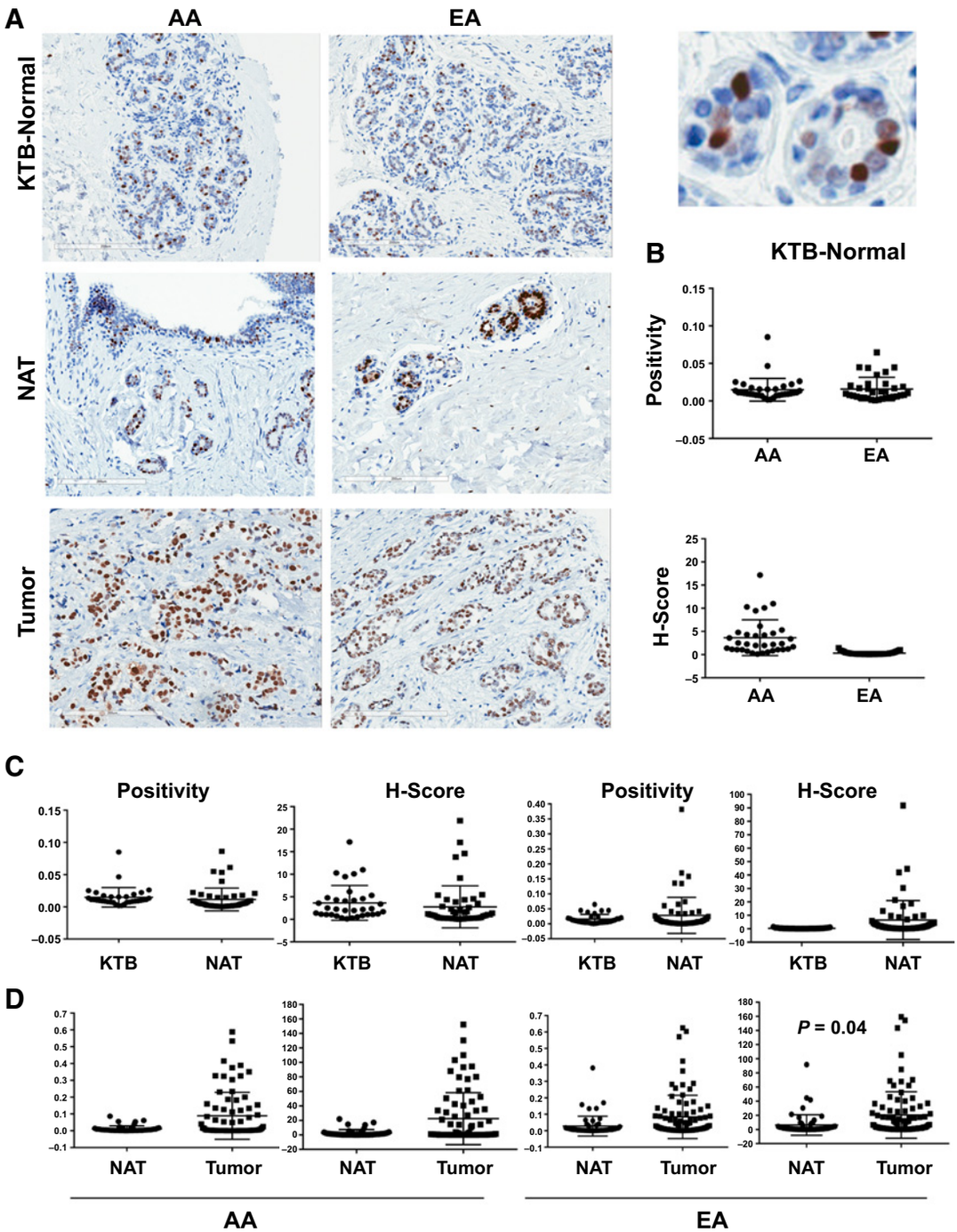


Figure 3.

ERα expression pattern in KTB-normal, NATs, and in breast tumors. **A,** Representative IHC of KTB-normal, NATs, and tumors of AA and EA women. Enlarged view of a KTB-normal is shown on right (top). **B,** Differences in ERα expression (positivity and H-score) between KTB-normal of AA and EA women. **C,** Differences between KTB-normal and NATs in AA and EA women. **D,** Differences between NATs and tumors in AA and EA women.

regulatory regions of this gene remain in a bivalent state, enabling the regulatory regions to respond readily to the tumor microenvironment and increase breast cancer plasticity and tumorigenicity (39). Another study showed elevated ZEB1 expression in normal breast stem cells, and it functionally protects stem cells from p53-mediated cell death in response to oncogene activation–induced DNA damage and promotes tumorigenicity with limited genomic instability (19). It was

Downloaded from clincancerres.aacrjournals.org on May 5, 2019. © 2019 American Association for Cancer Research.

Published OnlineFirst February 4, 2019; DOI: 10.1158/1078-0432.CCR-18-3427

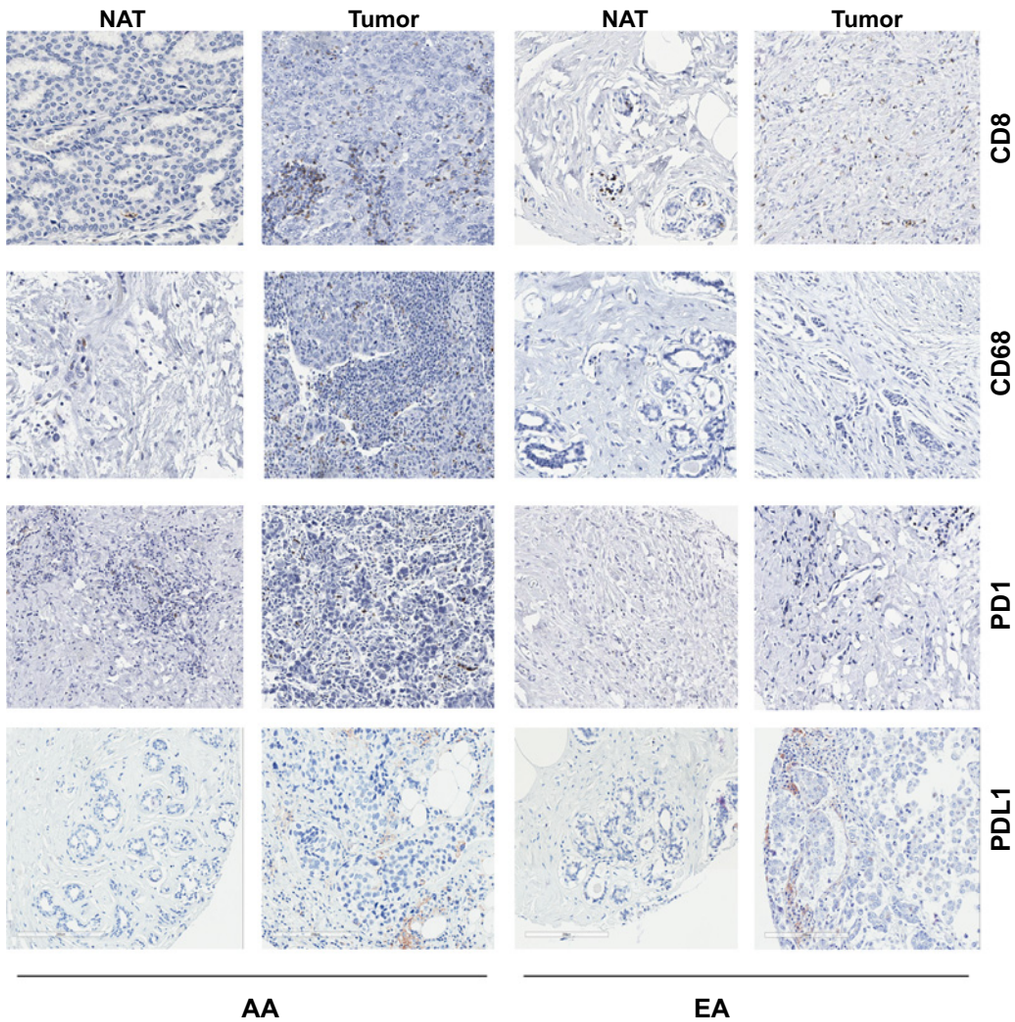


Figure 4.
Representative CD8, CD68, PD1, and PDL1 IHC of NATs and tumors of AA and EA women.

also reported that ZEB1 is expressed in both tumor and stromal cells of the breast (40). ZEB1 directly increases the expression of proinflammatory cytokines such as IL6 and IL8, and it promotes vascular mimicry of breast cancer cells by remodeling extracellular matrix (33, 41). We had previously demonstrated that cytokines such as TNF induce the expression of ZEB1 (42). These observations along with our unique observations of genetic ancestry–dependent differences in $ZEB1^+$ cells in the normal breast, elevated number of $ZEB1^+$ cells in NATs compared with healthy breast tissues of women of EA ancestry, and its localization outside the ductal structures raise several questions about the function of $ZEB1^+$ cells in the normal and tumor adjacent normal breast. We have shown previously that cytokeratin-positive, $PROCR^+/EpCAM^-$ cells of the normal breast, which are enriched in the normal breast of AA women compared with EA women, express 50-fold higher ZEB1 compared with cytokeratin-positive, $PROCR^-/EpCAM^+$ cells of the breast (12, 43). Thus, we suspect that $ZEB1^+$ cells in the normal breast correspond to $PROCR^+/EpCAM^-$ cells and that cancer-induced field effect leads to expansion/proliferation of such cells in the breast of EA women. Signaling pathways leading to proliferation of $ZEB1^+$ cells in NATs of EA women are unknown, but the Wnt pathway is the prime suspect as it is activated in cells surrounding cancer due to altered DNA methylation (3). In this respect, Wnt and ZEB1 constitute a reciprocal feed-forward signaling loop where ZEB1 enhances TCF4/β-Catenin–mediated transcription and Wnt signaling converts ZEB1 from a transcription repressor to an activator (44).

The reason for an intrinsically higher number of $ZEB1^+$ cells in AA women is unknown. TNBCs in AA compared with EA women display elevated Wnt pathway activation and it could be that Wnt pathway activity is intrinsically higher in AA women leading to elevated ZEB1 expression (8). It has also been demonstrated that vitamin D through Vitamin D Receptor (VDR) represses ZEB1 expression and serum vitamin D levels are significantly lower in AA than EA individuals (45, 46). Therefore, lower VDR activity and resulting increase in the activity of proinflammatory cytokines could be responsible for higher number of $ZEB1^+$ cells in the normal breast of AA women, which needs further investigation.

Downloaded from clincancerres.aacrjournals.org on May 5, 2019. © 2019 American Association for Cancer Research.

Published OnlineFirst February 4, 2019; DOI: 10.1158/1078-0432.CCR-18-3427

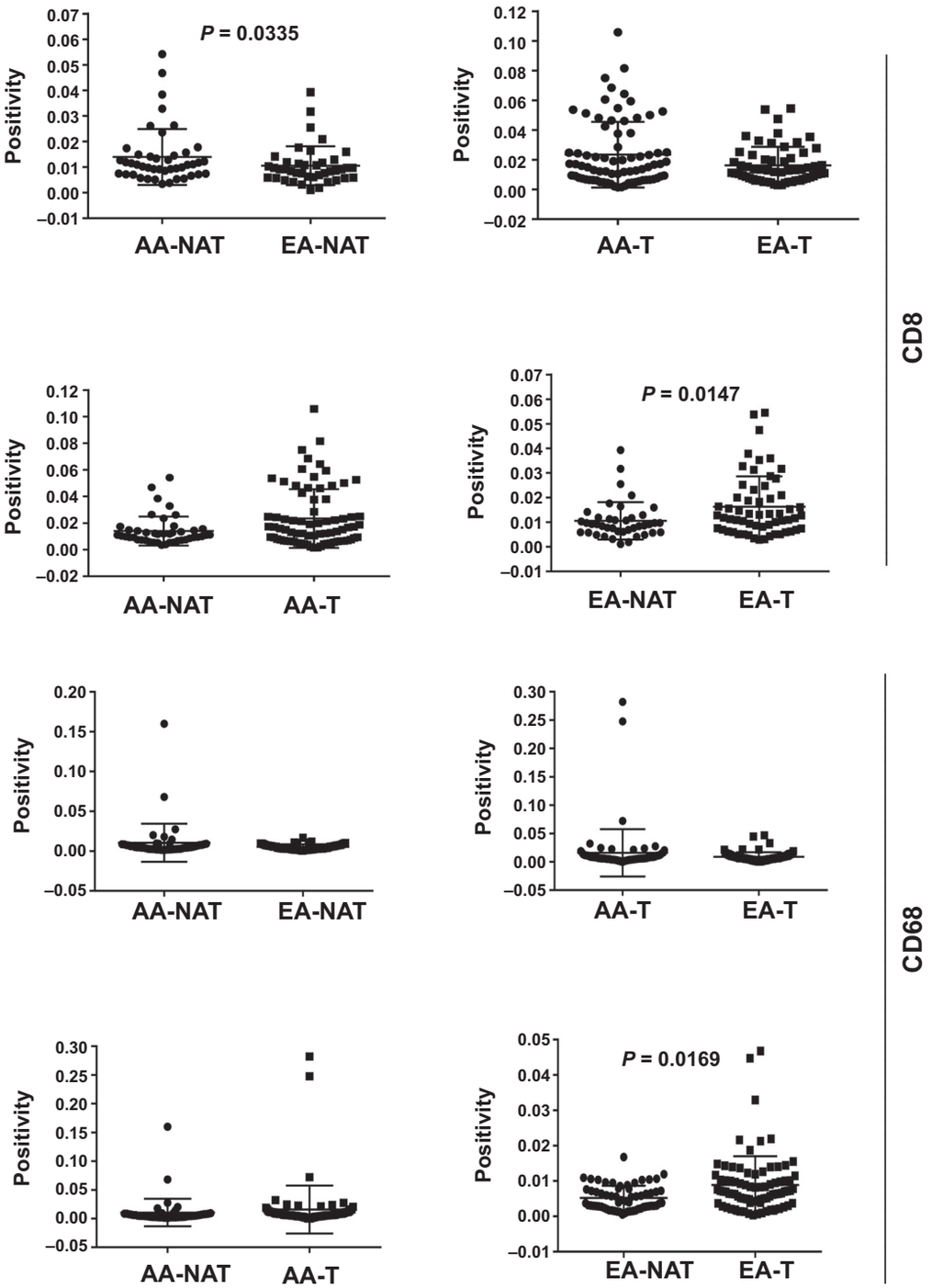


Figure 5.
Statistical analyses of CD8 and CD68 positivity in NATs and tumors (T) of AA and EA women. All statistically significant differences are indicated with *P*.

In contrast to stemness-associated ZEB1, FOXA1, and GATA3, which are expressed predominantly in differentiated luminal cells, showed opposite pattern in AA women. While the normal breasts of AA women had higher number of FOXA1$^+$ cells compared with EA women, a decline in FOXA1$^+$ cells in NATs as a consequence of cancer field effect was observed only in AA women. How tumors cause downregulation of FOXA1 in NATs is unknown but could involve inflammatory cytokines, as cytokine-inducible transcription repressors such as TWIST1 repress FOXA1 expression (47, 48). In this regard, we observed genetic

Downloaded from clincancerres.aacrjournals.org on May 5, 2019. © 2019 American Association for Cancer Research.

Published OnlineFirst February 4, 2019; DOI: 10.1158/1078-0432.CCR-18-3427

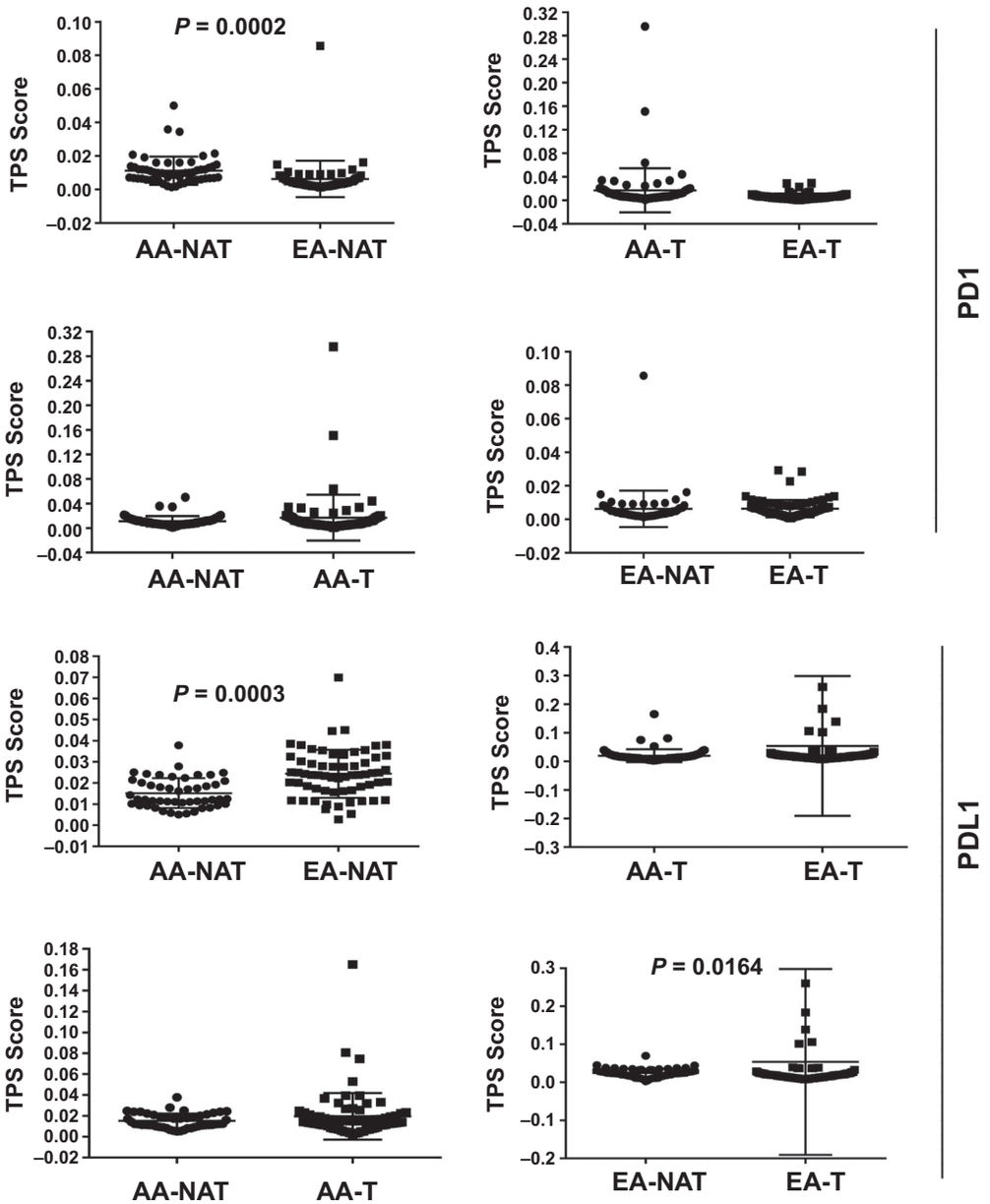


Figure 6.
Statistical analyses of PD1 and PDL1 TPS scores in NATs and tumors (T) of AA and EA women. All statistically significant differences are indicated with *P*.

ancestry–dependent differences in the levels of immune cells in NATs; NATs of AA women contained an elevated number of $CD8^+$ T cells and $PD1^+$ immune cells compared with NATs of EA women. In addition, FOXA1 regulatory regions are highly susceptible for DNA methylation and transcriptional repression, particularly in the context of BRCA1 deficiency (49). Furthermore, ER^- tumors in AA women show elevated FOXA1 DNA methylation compared with ER^- tumors of EA women (30). Recent studies have also demonstrated racial differences in plasma levels of cytokines with CCL2, CCL11, IL4, and IL10 being higher in EA women, and IL1RA and IFNα2 being higher in AA women (50).

Differential expression of GATA3 in the normal breasts of AA and EA women is intriguing, as GATA3 is one of the major signaling molecules required for hormonal response and differentiation of normal breast epithelial cells (25). Our results suggest that hormonal- and differentiation-signaling networks show genetic ancestry–dependent differences and it is likely that ERα:GATA3–dependent transcriptional program is more active in the normal breast of EA compared with AA women. Whether such difference between EA and AA persist in $ER\alpha^+$ tumors is unknown and potentially worth investigating as it is relevant for response to antiestrogen therapy.

Downloaded from clincancerres.aacrjournals.org on May 5, 2019. © 2019 American Association for Cancer Research.

Published OnlineFirst February 4, 2019; DOI: 10.1158/1078-0432.CCR-18-3427

Collectively, data presented in this study suggest the need to consider the following aspects for cancer biomarker discovery: (i) NATs are molecularly abnormal and thus are not suitable as controls; (ii) these abnormalities can be detected only when true normal breast tissues are used as controls and differences in normal gene expression attributable to genetic ancestry are taken into consideration; (iii) ZEB1 and GATA3 show unique expression pattern in the normal breast, which is influenced by the genetic ancestry and could potentially be developed as biomarkers of breast cancer initiation of women of EA ancestry; and (iv) genetic ancestry has an influence on the immune environment of tumors as well as NATs.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Authors' Contributions

Conception and design: H. Nakshatri, C. D'Souza-Schorey
Development of methodology: H. Nakshatri, M.L. Cox, G.E. Sandusky
Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): H. Nakshatri, M.L. Cox, M. Jacobsen, A.M.V. Storniolo
Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): H. Nakshatri, B. Kumar, H.N. Burney, M. Jacobsen, G.E. Sandusky, C. D'Souza-Schorey
Writing, review, and/or revision of the manuscript: H. Nakshatri, B. Kumar, C. D'Souza-Schorey, A.M.V. Storniolo
Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): H. Nakshatri, M.L. Cox, M. Jacobsen
Study supervision: H. Nakshatri
Other (pathology of the tissue samples and analysis): G.E. Sandusky

Acknowledgments

We thank the members of IU Simon Cancer Center tissue procurement facility, IU School of Medicine immunohistochemistry core, Komen Tissue Bank for their service, and countless number of women who donated their breast tissue for research purpose, as well as volunteers who facilitated tissue collection. Department of Defense DOD-W81XWH-15-1-0707, Susan G Komen for the Cure SAC110025, 100 Voices of Hope (to H. Nakshatri), and Catherine Peachy Foundation (to H. Nakshatri and C. D'Souza-Schorey) funded this study. Susan G. Komen for the Cure, Breast cancer Research Foundation and Vera Bradley Foundation for Breast Cancer Research provide funding support to the Komen Normal Tissue Bank.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received October 18, 2018; revised December 23, 2018; accepted January 25, 2019; published first February 4, 2019.

References

1. Pan H, Gray R, Braybrooke J, Davies C, Taylor C, McGale P, et al. 20-Year risks of breast-cancer recurrence after stopping endocrine therapy at 5 years. N Engl J Med 2017;377:1836–46.
2. Alzaghal AA, DiPiro PJ. Applications of advanced breast imaging modalities. Curr Oncol Rep 2018;20:57.
3. Teschendorff AE, Gao Y, Jones A, Ruebner M, Beckmann MW, Wachter DL, et al. DNA methylation outliers in normal breast tissue identify field defects that are enriched in cancer. Nat Commun 2016;7:10478.
4. Cancer Genome Atlas Network. Comprehensive molecular portraits of human breast tumours. Nature 2012;490:61–70.
5. Radovich M, Clare SE, Atale R, Pardo I, Hancock BA, Solzak JP, et al. Characterizing the heterogeneity of triple-negative breast cancers using microdissected normal ductal epithelium and RNA-sequencing. Breast Cancer Res Treat 2014;143:57–68.
6. Degnim AC, Visscher DW, Hoskin TL, Frost MH, Vierkant RA, Vachon CM, et al. Histologic findings in normal breast tissues: comparison to reduction mammaplasty and benign breast disease tissues. Breast Cancer Res Treat 2012;133:169–77.
7. Yuan J, Hu Z, Mahal BA, Zhao SD, Kensler KH, Pi J, et al. Integrated analysis of genetic ancestry and genomic alterations across cancers. Cancer Cell 2018;34:549–60.
8. Dietze EC, Sistrunk C, Miranda-Carboni G, O'Regan R, Seewaldt VL. Triple-negative breast cancer in African-American women: disparities versus biology. Nat Rev Cancer 2015;15:248–54.
9. Sineshaw HM, Gaudet M, Ward EM, Flanders WD, Desantis C, Lin CC, et al. Association of race/ethnicity, socioeconomic status, and breast cancer subtypes in the National Cancer Data Base (2010–2011). Breast Cancer Res Treat 2014;145:753–63.
10. DeSantis CE, Fedewa SA, Goding Sauer A, Kramer JL, Smith RA, Jemal A. Breast cancer statistics, 2015: convergence of incidence rates between black and white women. CA Cancer J Clin 2016;66:31–42.
11. Fejerman L, Ahmadiyeh N, Hu D, Huntsman S, Beckman KB, Caswell JL, et al. Genome-wide association study of breast cancer in Latinas identifies novel protective variants on 6q25. Nat Commun 2014;5:5260.
12. Nakshatri H, Anjanappa M, Bhat-Nakshatri P. Ethnicity-dependent and -independent heterogeneity in healthy normal breast hierarchy impacts tumor characterization. Sci Rep 2015;5:13526.
13. Bensen JT, Graff M, Young KL, Sethupathy P, Parker J, Pecot CV, et al. A survey of microRNA single nucleotide polymorphisms identifies novel breast cancer susceptibility loci in a case-control, population-based study of African-American women. Breast Cancer Res 2018;20:45.
14. Perkins SM, Bales C, Vladislav T, Althouse S, Miller KD, Sandusky G, et al. TFAP2C expression in breast cancer: correlation with overall survival beyond 10 years of initial diagnosis. Breast Cancer Res Treat 2015;152:519–31.
15. Padua MB, Bhat-Nakshatri P, Anjanappa M, Prasad MS, Hao Y, Rao X, et al. Dependence receptor UNC5A restricts luminal to basal breast cancer plasticity and metastasis. Breast Cancer Res 2018;20:35.
16. Scheel AH, Dietel M, Heukamp LC, Jöhrens K, Kirchner T, Reu S, et al. Harmonized PD-L1 immunohistochemistry for pulmonary squamous-cell and adenocarcinomas. Mod Pathol 2016;29:1165–72.
17. NSCLC-Keytruda. PsGtP-LTi. Dako Denmark A/S PD-L1 IHC 22C3 pharmDX Interpretation Manual 10. https://www.agilent.com/cs/library/usermanuals/public/29158_pd-l1-ihc-22C3-pharmdx-nsclc-interpretation-manual.pdf.
18. 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 2014;517:81–4.
19. Morel AP, Ginestier C, Pommier RM, Cabaud O, Ruiz E, Wicinski J, et al. A stemness-related ZEB1-MSRB3 axis governs cellular pliancy and breast cancer genome stability. Nat Med 2017;23:568–78.
20. Medina P, Navarro S, Bonet E, Martos L, Estelles A, Bertina RM, et al. Functional analysis of two haplotypes of the human endothelial protein C receptor gene. Arterioscler Thromb Vasc Biol 2014;34:684–90.
21. Nievergelt CM, Maihofer AX, Shekhtman T, Libiger O, Wang X, Kidd KK, et al. Inference of human continental origin and admixture proportions using a highly discriminative ancestry informative 41-SNP panel. Investig Genet 2013;4:13.
22. Carroll JS, Liu XS, Brodsky AS, Li W, Meyer CA, Szary AJ, et al. Chromosome-wide mapping of estrogen receptor binding reveals long-range regulation requiring the forkhead protein FoxA1. Cell 2005;122:33–43.
23. Bernardo GM, Lozada KL, Miedler JD, Harburg G, Hewitt SC, Mosley JD, et al. FOXA1 is an essential determinant of ERalpha expression and mammary ductal morphogenesis. Development 2010;137:2045–54.
24. Bernardo GM, Bebek G, Ginther CL, Sizemore ST, Lozada KL, Miedler JD, et al. FOXA1 represses the molecular phenotype of basal breast cancer cells. Oncogene 2013;32:554–63.
25. Eeckhoute J, Keeton EK, Lupien M, Krum SA, Carroll JS, Brown M. Positive cross-regulatory loop ties GATA-3 to estrogen receptor alpha expression in breast cancer. Cancer Res 2007;67:6477–83.
26. Badve S, Turbin D, Thorat MA, Morimiya A, Nielsen TO, Perou CM, et al. FOXA1 expression in breast cancer–correlation with luminal subtype A and survival. Clin Cancer Res 2007;13:4415–21.

Downloaded from clincancerres.aacrjournals.org on May 5, 2019. © 2019 American Association for Cancer Research.

Published OnlineFirst February 4, 2019; DOI: 10.1158/1078-0432.CCR-18-3427

27. Thorat MA, Marchio C, Morimiya A, Savage K, Nakshatri H, Reis-Filho JS, et al. Forkhead box A1 expression in breast cancer is associated with luminal subtype and good prognosis. J Clin Pathol 2008;61: 327–32.
28. Hurtado A, Holmes KA, Ross-Innes CS, Schmidt D, Carroll JS. FOXA1 is a key determinant of estrogen receptor function and endocrine response. Nat Genet 2011;43:27–33.
29. Fu X, Jeselsohn R, Pereira R, Hollingsworth EF, Creighton CJ, Li F, et al. FOXA1 overexpression mediates endocrine resistance by altering the ER transcriptome and IL-8 expression in ER-positive breast cancer. Proc Natl Acad Sci U S A 2016;113:E6600–E09.
30. Espinal AC, Buas MF, Wang D, Cheng DT, Sucheston-Campbell L, Hu Q, et al. FOXA1 hypermethylation: link between parity and ER-negative breast cancer in African American women? Breast Cancer Res Treat 2017;166: 559–68.
31. Clarke RB, Howell A, Potten CS, Anderson E. Dissociation between steroid receptor expression and cell proliferation in the human breast. Cancer Res 1997;57:4987–91.
32. Sorlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. Proc Natl Acad Sci U S A 2001;98:10869–74.
33. Katsura A, Tamura Y, Hokari S, Harada M, Morikawa M, Sakurai T, et al. ZEB1-regulated inflammatory phenotype in breast cancer cells. Mol Oncol 2017;11:1241–62.
34. Chan JS, Tan MJ, Sng MK, Teo Z, Phua T, Choo CC, et al. Cancer-associated fibroblasts enact field cancerization by promoting extratumoral oxidative stress. Cell Death Dis 2017;8:e2562.
35. Moller M, Strand SH, Mundbjerg K, Liang G, Gill I, Haldrup C, et al. Heterogeneous patterns of DNA methylation-based field effects in histologically normal prostate tissue from cancer patients. Sci Rep 2017;7: 40636.
36. GTEx Consortium, Laboratory, Data Analysis &, Coordinating Center (LDACC)–Analysis Working Group, Statistical Methods groups-Analysis Working Group, Enhancing GTEx (eGTEx) Groups, NIH Common Fund, NIH/NCI, et al. Genetic effects on gene expression across human tissues. Nature 2017;550:204–13.
37. Lappalainen T, Sammeth M, Friedlander MR, t Hoen PA, Monlong J, Rivas MA, et al. Transcriptome and genome sequencing uncovers functional variation in humans. Nature 2013;501:506–11.
38. Deschasaux M, Bouter KE, Prodan A, Levin E, Groen AK, Herrema H, et al. Depicting the composition of gut microbiota in a population with varied ethnic origins but shared geography. Nat Med 2018;24: 1526–31.
39. Chaffer CL, Marjanovic ND, Lee T, Bell G, Kleer CG, Reinhardt F, et al. Poised chromatin at the ZEB1 promoter enables breast cancer cell plasticity and enhances tumorigenicity. Cell 2013;154:61–74.
40. Jang MH, Kim HJ, Kim EJ, Chung YR, Park SY. Expression of epithelial-mesenchymal transition-related markers in triple-negative breast cancer: ZEB1 as a potential biomarker for poor clinical outcome. Hum Pathol 2015;46:1267–74.
41. Langer EM, Kendsersky ND, Daniel CJ, Kuziel GM, Pelz C, Murphy KM, et al. ZEB1-repressed microRNAs inhibit autocrine signaling that promotes vascular mimicry of breast cancer cells. Oncogene 2018;37:1005–19.
42. Chua HL, Bhat-Nakshatri P, Clare SE, Morimiya A, Badve S, Nakshatri H. NF-kappaB represses E-cadherin expression and enhances epithelial to mesenchymal transition of mammary epithelial cells: potential involvement of ZEB-1 and ZEB-2. Oncogene 2007;26:711–24.
43. Kumar B, Prasad MS, Bhat-Nakshatri P, Anjanappa M, Kalra M, Marino N, et al. Normal breast-derived epithelial cells with luminal and intrinsic subtype-enriched gene expression document inter-individual differences in their differentiation cascade. Cancer Res 2018;78:5107–23.
44. Sanchez-Tillo E, de Barrios O, Valls E, Darling DS, Castells A, Postigo A. ZEB1 and TCF4 reciprocally modulate their transcriptional activities to regulate Wnt target gene expression. Oncogene 2015;34:5760–70.
45. Larsen JE, Nathan V, Osborne JK, Farrow RK, Deb D, Sullivan JP, et al. ZEB1 drives epithelial-to-mesenchymal transition in lung cancer. J Clin Invest 2016;126:3219–35.
46. Xu J, Bartz TM, Chittoor G, Eiriksdottir G, Manichaikul AW, Sun F, et al. Meta-analysis across Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) consortium provides evidence for an association of serum vitamin D with pulmonary function. Br J Nutr 2018;120: 1159–70.
47. Xu Y, Qin L, Sun T, Wu H, He T, Yang Z, et al. Twist1 promotes breast cancer invasion and metastasis by silencing Foxa1 expression. Oncogene 2017;36: 1157–66.
48. Li CW, Xia W, Huo L, Lim SO, Wu Y, Hsu JL, et al. Epithelial-mesenchymal transition induced by TNF-alpha requires NF-kappaB-mediated transcriptional upregulation of Twist1. Cancer Res 2012;72:1290–300.
49. Gong C, Fujino K, Monteiro LJ, Gomes AR, Drost R, Davidson-Smith H, et al. FOXA1 repression is associated with loss of BRCA1 and increased promoter methylation and chromatin silencing in breast cancer. Oncogene 2015;34:5012–24.
50. Yao S, Hong CC, Ruiz-Narvaez EA, Evans SS, Zhu Q, Schaefer BA, et al. Genetic ancestry and population differences in levels of inflammatory cytokines in women: role for evolutionary selection and environmental factors. PLoS Genet 2018;14:e1007368.

Downloaded from clincancerres.aacrjournals.org on May 5, 2019. © 2019 American Association for Cancer Research.

Published OnlineFirst April 16, 2019; DOI: 10.1158/1541-7786.MCR-19-0165

New Horizons in Cancer Biology

Molecular Cancer Research

Dual TGFβ/BMP Pathway Inhibition Enables Expansion and Characterization of Multiple Epithelial Cell Types of the Normal and Cancerous Breast

Mayuri Prasad[1], Brijesh Kumar[1], Poornima Bhat-Nakshatri[1], Manjushree Anjanappa[1], George Sandusky[2], Kathy D. Miller[3], Anna Maria Storniolo[3], and Harikrishna Nakshatri[1,4,5]

Abstract

Functional modeling of normal breast epithelial hierarchy and stromal–epithelial cell interactions have been difficult due to inability to obtain sufficient stem-progenitor-mature epithelial and stromal cells. Recently reported epithelial reprogramming assay has partially overcome this limitation, but cross-contamination of cells from the feeder layer is a concern. The purpose of this study was to develop a feeder-layer–independent and inexpensive method to propagate multiple cell types from limited tissue resources. Cells obtained after enzymatic digestion of tissues collected at surgery or by core-needle biopsies were plated on tissue culture dishes precoated with laminin-5-rich–conditioned media from the rat bladder tumor cell line 804G and a defined growth media with inhibitors of ROCK, TGFβ, and BMP signaling. Cells were characterized by flow cytometry, mammosphere assay, 3D cultures, and xenograft studies. Cells from the healthy breasts included $CD10^+/EpCAM^-$ basal/myoepithelial, $CD49f^+/EpCAM^+$ luminal progenitor, $CD49f^-/EpCAM^+$ mature luminal, $CD73^+/EpCAM^+/CD90^-$ rare endogenous pluripotent somatic stem, $CD73^+/CD90^+/EpCAM^-$, estrogen receptor alpha–expressing ALCAM $(CD166)^+/EpCAM^+$, and $ALDFLUOR^+$ stem/luminal progenitor subpopulations. Epithelial cells were luminal ($KRT19^+$), basal ($KRT14^+$), or dual-positive luminal/basal hybrid cells. While breast cells derived from BRCA1, BRCA2, and PALB2 mutation carriers did not display unique characteristics, cells from women with breast cancer–protective alleles showed enhanced differentiation. Cells could also be propagated from primary tumors and metastasis of breast, ovarian, and pancreatic cancer–neuroendocrine subtype. Xenograft studies confirmed tumorigenic properties of tumor-derived cells.

Implications: Our method expands the scope of individualized studies of patient-derived cells and provides resources to model epithelial–stromal interactions under normal and pathologic conditions.

Introduction

The breast tissue is composed of the epithelial and stromal cells cushioned by layers of adipose cells (1). The ability to culture and propagate various types of cells that form the breast hierarchy, which include the stem-progenitor-mature epithelial cells and stromal cells, is of vital importance for all functional studies related to *in vitro* characterization of different cell types, epithelial–stromal interaction, elucidation of molecular mechanisms of normal cell differentiation, cancer initiation, and progression. Several protocols have been developed that support the propagation of breast epithelial cells, which, in large part, are biased toward outgrowth of basal cells (summarized in ref. 2). Several of these culture protocols utilize irradiated mouse embryonic fibroblast or human foreskin fibroblasts as feeder cells to maintain the pluripotent state of stem cells (3, 4). Normal breast stromal cells have morphologic features and characteristics of fibroblasts. It is also difficult to distinguish mesenchymal and adipose stem cells from fibroblasts without profiling for cell surface markers. Therefore, an ideal system should allow growth of as many cell types of a tissue as possible and should utilize alternatives to feeder layers such as extracellular matrix proteins and small molecules that enable the maintenance of adult stem cells and their lineage commitment properties. In this respect, vitamin C and inhibitors of Rho-associated coil–coil containing protein kinase (ROCK) have been shown to be effective in stem cell reprogramming and in preventing matrix detachment–induced apoptosis of stem cells, respectively (3, 5).

A recent study reported that plating adult epithelial cells from lungs on tissue culture dishes precoated with laminin-5-rich–conditioned media from the rat bladder cancer cell line 804G and in media containing dual inhibitors of SMAD/BMP pathways permits propagation of diverse epithelial cells (6). We had

[1]Department of Surgery, Indiana University School of Medicine, Indianapolis, Indiana. [2]Department of Pathology and Laboratory Medicine, Indiana University School of Medicine, Indianapolis, Indiana. [3]Department of Medicine, Indiana University School of Medicine, Indianapolis, Indiana. [4]Biochemistry and Molecular Biology, Indiana University School of Medicine, Indianapolis, Indiana. [5]Roudebush VA Medical Center, Indianapolis, Indiana.

Note: Supplementary data for this article are available at Molecular Cancer Research Online (http://mcr.aacrjournals.org/).

M. Prasad and B. Kumar contributed equally to this article.

Corresponding Author: Harikrishna Nakshatri, Indiana University School of Medicine, C218C, 980 West Walnut St., Indianapolis, IN 46202. Phone: 317-278-2238; Fax: 317-274-0396; E-mail: hnakshat@iupui.edu.

Mol Cancer Res 2019;17:1556–70

doi: 10.1158/1541-7786.MCR-19-0165

©2019 American Association for Cancer Research.

Downloaded from mcr.aacrjournals.org on July 2, 2019. © 2019 American Association for Cancer Research.

Published OnlineFirst April 16, 2019; DOI: 10.1158/1541-7786.MCR-19-0165

previously shown that maintaining primary breast epithelial cells in low glucose–containing media and limiting the use of DMSO as a solvent for small molecules enable long-term culture of breast epithelial cells from a core-needle biopsy on a feeder layer and these cultured cells maintain stem-progenitor–differentiated cell hierarchy (7). The goals of this study were to determine whether combining these two methods would permit growth of multiple cell types of the breast without the need for feeder layer and whether the technique can be extended to other cancer types as well as biopsies from metastatic sites. Because there is growing evidence for heterogeneity between primary tumor and metastasis within an individual patient, despite sharing large number of driver mutations (8), there is an urgent need to develop methods that enable characterization of metastatic cells *in vitro* for therapeutic decision-making. We show that the method described here is effective in propagating primary epithelial cells that include but is not limited to (i) normal breast, (ii) carriers of BRCA1, BRCA2, PALB2 mutations, or breast cancer–protective alleles, (iii) primary breast cancer, (iv) pleural effusions of patients with breast cancer, (v) ascites fluids from patients with ovarian cancer, and (vi) liver metastasis of breast cancer and pancreatic cancer of neuroendocrine origin. The majority of cells cultured with this protocol maintained the epithelial phenotype and could be cultured for approximately 10–12 passages and provided >5 $\times 10^6$ cells for cryopreservation, functional studies, and/or to immortalize/transform in case of primary nontransformed cells. Nonepithelial cells could be easily separated from epithelial cells by flow cytometry and cultured for further evaluation of the microenvironment. In addition, tumor cells formed tumors in NSG mice thus allowing simultaneous characterization of patient-derived primary and metastatic tumor cells both *in vitro* and *in vivo*.

Materials and Methods

Creating and propagating primary cells

Fresh or cryopreserved, deidentified tissue samples were obtained from the Indiana University (IU) Simon Cancer Center Tissue Bank or Komen Tissue Bank (KTB) at IU School of Medicine, Indianapolis, after informed consent from the donors. All experiments were carried out in accordance with the approved guidelines of the Indiana University Institutional Review Board. The 804G cells were a generous gift from Dr. Rajagopal (Harvard Medical School, Boston, MA). The 804G cells were grown in RPMI media (10-0400CV, Corning) with 10% FBS and 1% penicillin and streptomycin. Conditioned medium (CM) from the 804G cells was harvested once the plates were confluent. No more than 2–3 batches of CM were collected from one culture. The CM was filtered through 22-μm filters, divided into aliquots in 50 mL tubes and frozen at −20°C for future use. A day before the tissues were processed, 60-mm culture dishes were prepared by adding 5 mL of the thawed 804G CM followed by overnight incubation at 37°C. Just prior to use, the CM was aspirated and the dish was washed with PBS once. This CM is enriched for laminin-related components (9).

Supplementary Table S1 provides details of tissues used in this study. Solid tissues dissociation protocol along with primary cell culture media has been described previously (7). The primary cell media was supplemented with 1 μmol/L of BMP inhibitor DMH-1 (#4126, Tocris), 1 μmol/L of TGFβ inhibitor A-83-01 (#2939, Tocris), 5 μmol/L of ROCK inhibitor Y-27632, and 4 μL/mL of adenine (6 mg/mL stock, A9001, Sigma-Aldrich). Pleural effusion or ascites fluids were centrifuged at 1,000 rpm for 5 minutes to collect cell pellet and the pellet was washed twice in PBS before plating on 804G CM–coated plates. Medium was changed the next day to remove cellular debris and floating cells. The cells could be cultured for 10–12 passages using this method.

Flow cytometry analysis

Adherent cells were collected by trypsinization, stained using antibodies CD49f-APC (FAB13501A), CD140b-FITC (FAB1263F; R&D Systems), PROCR (CD201)-PE (130-105-256), EpCAM−PE (130-091-253), EpCAM−APC (130-091-254; Miltenyi Biotec Inc.), CD271-APC (345108; BioLegend), CD44-APC (559942), CD24-PE (555428), CD73-PE (561014), CD90-APC (559869), CD166-PE (559263), JAM-1-PE (552556), MUC-1-FITC (559774; BD Pharmingen), CD10-PE (340920; BD Biosciences), and CD117-FITC (11-1178-42; eBioscience), and were acquired using a BD LSR II flow cytometer. Data were analyzed using CellQuest or FlowJo software. Forward and side scatter were used to ensure that only live cells were considered in the analysis. Gating was done using appropriate FITC (555573), PE (555749), and APC (555576; BD Pharmingen) isotype control antibodies and only a representative isotype control for two fluorescent markers are shown.

Mammosphere formation assay

Mammosphere assay in 6-well plate at a density of 5,000 cells/mL has been described previously (10). Phase contrast images were taken at day 5. Mammospheres were collected, washed, and trypsinized to obtain single-cell suspensions. Cells were stained with indicated antibodies and subjected to flow cytometry.

Determination of *in vitro* growth of primary cells

We used the KTB200, KTB201, and KTB202 primary cells to characterize the growth kinetics and *in vitro* survival of primary cells in culture. Growth kinetics was evaluated from successive cell counts. A total of 2.5×10^5 cells per cell type were seeded on 60-mm dishes at day zero and incubated under standard conditions. Cells were harvested every 2 days by trypsinization, counted using a hemocytometer, and 2.5×10^5 cells were reseeded for the next culture.

Culturing and characterization of cells in 3D collagen or Matrigel scaffolds

Floating collagen scaffold was prepared on the basis of a protocol from Linnermann and colleagues, with some modifications as detailed in Kumar and colleagues (2, 11). Matrigel basement membrane (354234, Corning) was slowly thawed on ice. Eight-chambered coverglass system (155409, Lab-Tek II) was coated with 40 μL of Matrigel per well and spread evenly to cover the bottom surface. The chamber was incubated for 30 minutes at 37°C for the gel to solidify. A total of 400 μL of overlay media comprising primary cell media as described above with 2% Matrigel and 6,000 cells per well were added. The medium was changed every 3–4 days and the cells were cultured to 10 to 12 days. Immunofluorescence staining of 3D structures has been described previously (2).

RNA isolation and qRT-PCR

Total RNA was isolated using RNeasy Kit from Qiagen and 1 μg of RNA was used to synthesize cDNA with Bio-RAD iScript cDNA

Downloaded from mcr.aacrjournals.org on July 2, 2019. © 2019 American Association for Cancer Research.

Published OnlineFirst April 16, 2019; DOI: 10.1158/1541-7786.MCR-19-0165

Synthesis Kit. qRT-PCR was performed using Taqman universal PCR mix and predesigned Taqman assay primers with best coverage from Applied Biosystems. The following assays were used in our study: *ACTB* (Hs01060665_g1), *ESR1* (Hs01046816_m1), *GATA3* (Hs00231122_m1) *FOXA1* (Hs04187555_m1), and *FOXC1* (Hs00559473_s1). Applied Biosystems StepOneplus real time PCR system was used for PCR and fluorescence detection and the StepOne software was used for analysis.

Immunofluorescence staining of monolayer culture

Cells were cultured in glass-bottomed Microwell Dish (P35G-0-14-C, MatTek Corporation) and stained with KRT14 (ab7800, Abcam), KRT19 (ab52625, Abcam), and KRT17 (ab53707, Abcam) primary antibody according to protocol described previously (2). Cells were stored in PBS in dark and images were taken within 48 hours of staining with Olympus FV1000 MPE inverted confocal microscope.

Xenograft studies

The Indiana University Animal Care and Use Committee approved the use of animals in this study and all procedures were performed as per NIH guidelines. Primary cells created from pleural effusions of breast cancer or triple-negative breast cancers (TNBC; 5×10^5 cells) with 50% Basement Membrane Matrix (BME) Type 3 (100 μL volume) were implanted into the mammary fat pad of 5- to 6-week-old female NSG (NOD/SCID/IL2Rgnull) mice. A 60-day slow-release estradiol (0.72 mg) pellet was implanted at the time of mammary fat pad injection, irrespective of the injected breast cancer subtype. Tumor growth was measured weekly and tumor volume was calculated as described previously (12). After 2–5 months, lungs and primary tumors were collected and processed for hematoxylin and eosin (H&E), ERα, PR, GATA3, and FOXA1 staining.

IHC

H&E, ERα, PR, GATA3, and FOXA1 immunostaining was performed at the CLIA-certified Indiana University Health Pathology Laboratory and the whole-slide digital imaging system of Aperio (ScanScope CS) was used for imaging. The following antibodies were used: ER (clone EP1, Dako IR 084), PR (#RB-9017-P, NeoMarkers), FOXA1 (sc-6553, Santa Cruz Biotechnology), and GATA3 (sc-268, Santa Cruz Biotechnology).

Statistical analysis

Statistical analyses were performed using GraphPad Prism software program (version 6.0). *P* values below 0.05 were considered statistically significant.

Results

We have developed a method for propagating the primary cells from normal, cancerous, and metastatic tissue samples with appropriate modifications to growth media. Cells propagated from more than 30 primary breast tissues (>6 healthy donors, two BRCA1 mutants, three BRCA2 mutants with one right and left breasts of the same donor, three PALB2 high-risk, two protective alleles, tumors with adjacent normal tissues from the same patients, several TNBCs, three breast cancer pleural effusion fluids, four breast cancer metastasis from different organs, one liver metastasis of pancreatic cancer–neuroendocrine subtype, and three ascites fluids of patients with ovarian cancer) were analyzed (Supplementary Table S1). These extensive analyses recognized enormous phenotypic heterogeneity in the normal breast and tumors.

Normal breast contains multiple subpopulations of cells

We subjected primary breast cells to phenotypic analyses using established and recently discovered markers that define the normal and cancer stem cells (CSC), pluripotent somatic stem cells, progenitor, differentiated, luminal, basal/myoepithelial cells, and estrogen receptor alpha (ERα)-positive subpopulation of cells (2, 13, 14). Similar to results obtained with cells propagated using the previous epithelial reprogramming assay (7), this new culturing method enabled us to document enormous phenotypic heterogeneity in the normal breast. The expanded cell populations are described as $CD49f^+/EpCAM^-$ stem/basal–enriched cells, $CD49f^+/EpCAM^+$ luminal progenitor cells, $CD49f^-/EpCAM^+$ mature luminal cells, $CD44^+/CD24^-$ stem/basal cells, $CD44^+/CD24^+$ stem/luminal progenitor cells, $CD44^-/CD24^+$ differentiated cells, $CD201^+/EpCAM^-$ multipotent stem cells, $CD271^+$ rare basal cells in luminal breast cancer (relevant in antiestrogen resistance models), $CD73^+/EpCAM^+/CD90^-$ rare endogenous pluripotent somatic stem cells, $CD73^+/CD90^+/EpCAM^-$ potential mesenchymal stem cells, $CD10^+/EpCAM^-$ basal/myoepithelial cells, ALCAM $(CD166)^+$ cells that are enriched for ERα, and JAM-A^+ cells, which are enriched for cancer stem cell phenotype (15–17). Flow cytometry characterization of breast epithelial cells from six healthy women (KTB200-205) showed remarkable variability in stem/basal, progenitor, and mature/differentiated cell subpopulations (Fig. 1A; Supplementary Fig. S1). All samples were from premenopausal women (20–55 age range, 23–41 BMI range, Supplementary Table S1). Using a combination of CD44 and EpCAM, we were able to document the presence of luminal ($CD44^-/EpCAM^+$ or $CD44^{low}/EpCAM^+$) and basal ($CD44^+/EpCAM^{low}$) cells (18). Proportion of cells that express the cell adhesion molecule MUC1 also varied between samples (Supplementary Fig. S1). MUC1 is expressed predominantly in $CD24^+$ luminal cells and these cells often display higher self-renewal capacity and are metastatic (19, 20). With respect to rare pluripotent somatic stem cells (13), our technique expanded this population, as a significant portion of $CD73^+/CD90^-$ cells was $EpCAM^+$ (Fig. 1B). Cells, propagated using this assay, are free of fibroblasts as they lacked the expression of CD140b, which identifies fibroblasts (Supplementary Fig. S1; ref. 21). We also did not find cells expressing CD117, a putative marker of luminal progenitor cells (Supplementary Fig. S1; ref. 22). These results suggest interindividual heterogeneity among the normal breasts, particularly in number of breast epithelial cells–expressing markers that identify cells as stem, basal, or luminal cells, similar to our previous report (7). However, as we indicated in our previous report, because of limited sample size, the observed heterogeneity cannot be attributed to differences in age, BMI, or time of sample collection (luteal or follicular phase).

We performed serial dilution cell proliferation assay to determine the number of passages these cells can be maintained. Proliferation kinetics varied from samples to samples and declined by around passage 7 (Supplementary Fig. S1). However, by the time cells reached passage 7, couple of million cells could be easily obtained for immortalization and functional studies.

Downloaded from mcr.aacrjournals.org on July 2, 2019. © 2019 American Association for Cancer Research.

Published OnlineFirst April 16, 2019; DOI: 10.1158/1541-7786.MCR-19-0165

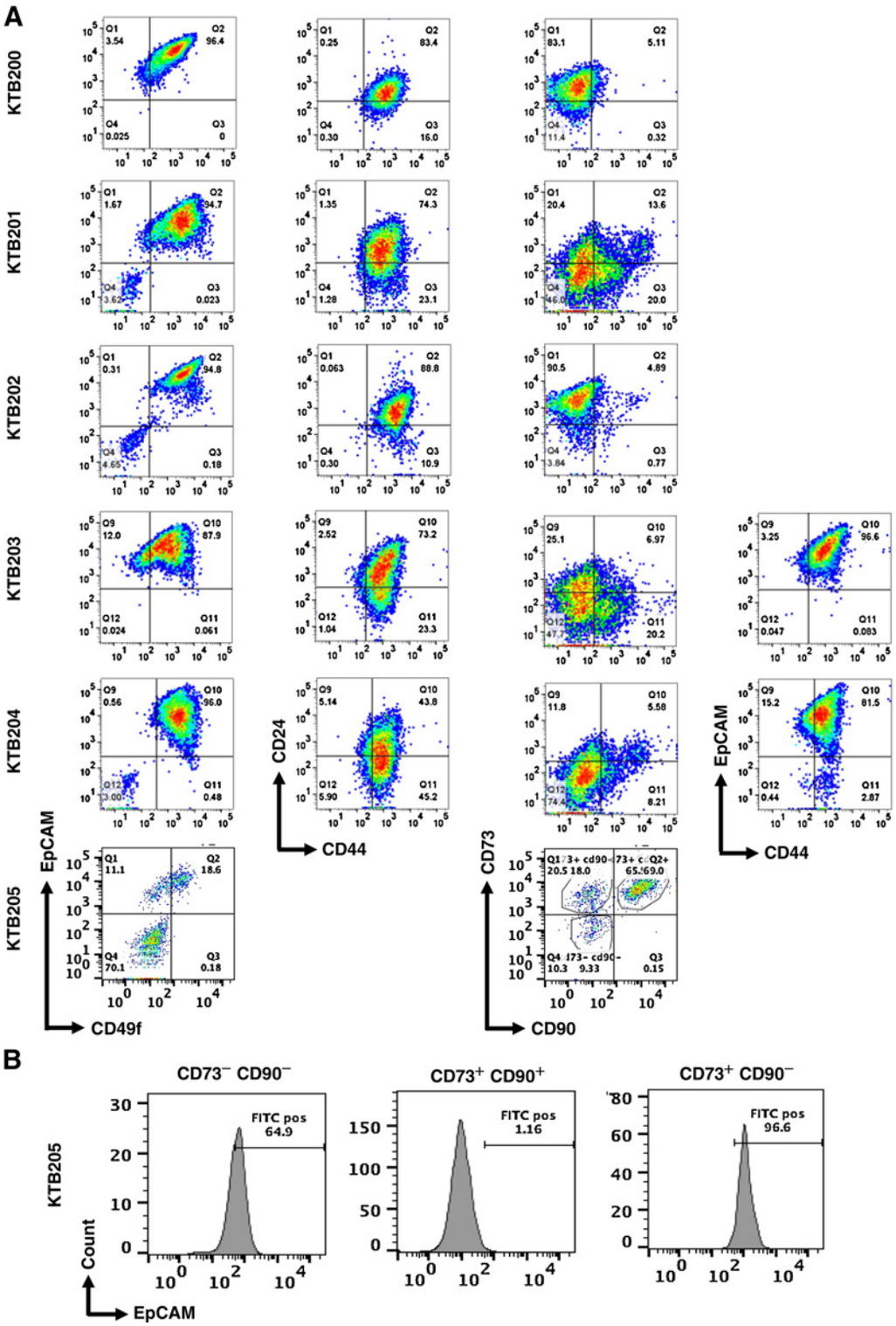


Figure 1.

Breast epithelial cells from the normal breast contain multiple cell types and show interindividual heterogeneity. **A,** Breast epithelial cells isolated from the normal breasts of five clinically healthy women show interindividual differences in staining patterns with CD49f, CD44, CD24, CD90, CD73, and EpCAM antibodies. **B,** The culturing method enriches for rare endogenous pluripotent somatic stem cells based on CD73/CD90/EpCAM staining patterns. EpCAM expression was seen in greater than 96% of $CD73^{+}/CD90^{-}$ cells in KTB205.

Downloaded from mcr.aacrjournals.org on July 2, 2019. © 2019 American Association for Cancer Research.

Published OnlineFirst April 16, 2019; DOI: 10.1158/1541-7786.MCR-19-0165

Breast epithelial cells from women carrying high-risk mutations

To evaluate the suitability of the assay for growing cells from high-risk women, we created primary cells from prophylactic mastectomy or biopsies of women carrying germline mutations in BRCA1 (two cases), BRCA2 (three cases, one with tissues from both breasts), or PALB2 (three cases). Staining with multiple antibody combinations did not show any patterns that are distinct from the cells cultured from healthy breasts (Fig. 2A; Supplementary Fig. S2). It is also remarkable that interindividual differences in staining patterns were noted among high-risk patients analyzed in the study (Supplementary Fig. S2). For example, CD271/EpCAM, CD10/EpCAM, and MUC1/EpCAM staining patterns of two BRCA2 mutants and three PALB2 mutants were different (Supplementary Fig. S2). Thus, it is unlikely that the breasts of high-risk women contain phenotypically distinct population of cells compared with breasts of healthy women. In general, growing epithelial cells from women with PALB2 and CHEK2 mutations was more difficult than others because of outgrowth of cells with fibroblastic appearance or epithelial cells undergoing rapid epithelial-to-mesenchymal transition (EMT).

Cells with BRCA1 mutations, similar to cells from healthy women, formed acini on collagen gel and these acinis were composed of KRT5/6 and KRT17$^+$ cells (Supplementary Fig. S3). We compared the staining pattern of cells from both the breasts of a BRCA2 carrier to ensure that observed staining pattern is not due to collection of biopsy from specific regions of the breast. Randomly collected breast tissues from both the breasts showed similar cell surface marker profiles (Fig. 2B; Supplementary Fig. S3). Immunofluorescence with basal cell marker KRT14 and luminal cell marker KRT19 showed that the majority of cells from high-risk women are hybrid cells that expressed both KRT14 and KRT19 (Fig. 2C), which is a suggested characteristic of cells with luminal progenitor phenotype (23).

We had previously demonstrated that immortalized cell lines derived from healthy breast self-renew as well as undergo differentiation when cultured under mammosphere growth conditions (2). To determine whether cells from high-risk women differ in their capacity to undergo self-renewal and differentiation, we subjected cells from BRCA1 and BRCA2 carriers to mammosphere assay followed by flow cytometry (Fig. 2D and E). Phase contrast images of these mammospheres showed similarity between samples. However, as noted previously with cells from healthy breasts (2), interindividual variability in differentiation capacity was noted with cells from high-risk women. For example, the proportion of cells with differentiated luminal phenotype (CD49f$^-$/EpCAM$^+$ or CD44$^-$/EpCAM$^+$ cells) increased under mammosphere condition compared with 2D growth condition but the degree of increase varied between samples (Fig. 2A and E). Staining patterns of mammospheres with additional markers are shown in Supplementary Fig. S3.

Cells grown from women with protective alleles show enhanced differentiation capacity

Although epidemiologic and genome-wide association studies have identified various risk, predisposition, and protective alleles including new 65 risk loci for breast cancer (24, 25), functional analyses of these risk loci have been difficult because of limited cell resources. To determine whether culture system developed here is useful in this regard, we focused on the recently described breast cancer–protective variant (rs140068132) in Latinas on chromosome 6q25 located 5′ of *Estrogen Receptor 1 (ESR1)* gene, which is limited to Indigenous American Ancestry (26). We examined DNA from self-reported 43 Hispanic women from whom cryopreserved normal breast tissues were available at the Komen Tissue Bank (KTB). Among these women, three carried the protective allele (Fig. 3A) but viable cells could be generated from two samples. These cells were less proliferative than cells generated from healthy or high-risk women and only limited number of experiments could be performed. Cell surface marker profiles showed clear enrichment of cells with differentiated/luminal characteristics (Fig. 3B; Supplementary Fig. S4A and S4B). For example, unlike in cases of healthy or high-risk women, the majority of cells from women with protective alleles were CD49f$^-$/EpCAM$^+$, CD271$^-$/EpCAM$^+$, and CD201$^-$/EpCAM$^+$. In addition, these cells had higher proportion of cells that were KRT14$^-$/KRT19$^+$ compared with cells from noncarriers of protective alleles or high-risk individuals (Fig. 3C), and were KRT17$^+$ (Supplementary Fig. S4B). Furthermore, cells from one protective allele carrier, which we were able to grow for additional analyses, expressed higher levels of luminal cell markers *ESR1*, *FOXA1*, and *GATA3* compared with cells from normal (KTB51) and high-risk individuals (Fig. 3D). In mammosphere assays, cells with protective alleles generated very small spheres compared with cells generated from two Hispanic women not carrying the protective allele (Fig. 3E). Thus, the assay established here can be used to characterize cells from women carrying distinct risk loci.

Individualizing tumor characterization through propagation of tumor and adjacent normal from the same patient

Because gene expression in normal tissues shows interindividual variations due to SNP in the gene regulatory regions (27), we had recently proposed that normal cells from the same patient need to be used to determine cancer-specific gene expression changes (28). Although the use of cells from normal tissues adjacent to tumor is far from perfect because of cancer-induced field defects on adjoining cells (29, 30), it is still better to use those cells or from the unaffected contralateral breast than normal cells from unrelated individuals as controls to identify cancer-specific gene expression changes. Toward achieving this goal, we established culture conditions to grow tumor and tumor adjacent–normal from the same patient. Representative flow cytometry profiles with various markers are shown in Fig. 4A; Supplementary Fig. S4C and S4D. Both tumor and the adjacent normal showed similar marker profiles except that tumors had a minor CD49f^{high}/EpCAM$^+$ population. Cells derived from not all tumors that were characterized showed similar profiles as samples from TNBCs showed a distinct profile than the sample described above (Fig. 4B; Supplementary Fig. S4D). Tumor cells were KRT14$^+$/KRT19$^+$ suggesting their luminal progenitor characteristics, similar to flow cytometry stain pattern (Fig. 4C). We have created a bank of tumor cells from at least 15 patients with few cases of paired tumor and tumor adjacent–normal, which could be utilized for functional studies.

Tumor cells form ductal and acinar-like structures on collagen gels

Because organoid culture system is gaining traction for various types of studies including screening for anticancer therapies (31), we determined whether tumor cells form organoid-like structures in 3D. On Matrigel, tumor cells formed spheres, whereas on collagen gels, cells formed elongated structures that resembled

Downloaded from mcr.aacrjournals.org on July 2, 2019. © 2019 American Association for Cancer Research.

Published OnlineFirst April 16, 2019; DOI: 10.1158/1541-7786.MCR-19-0165

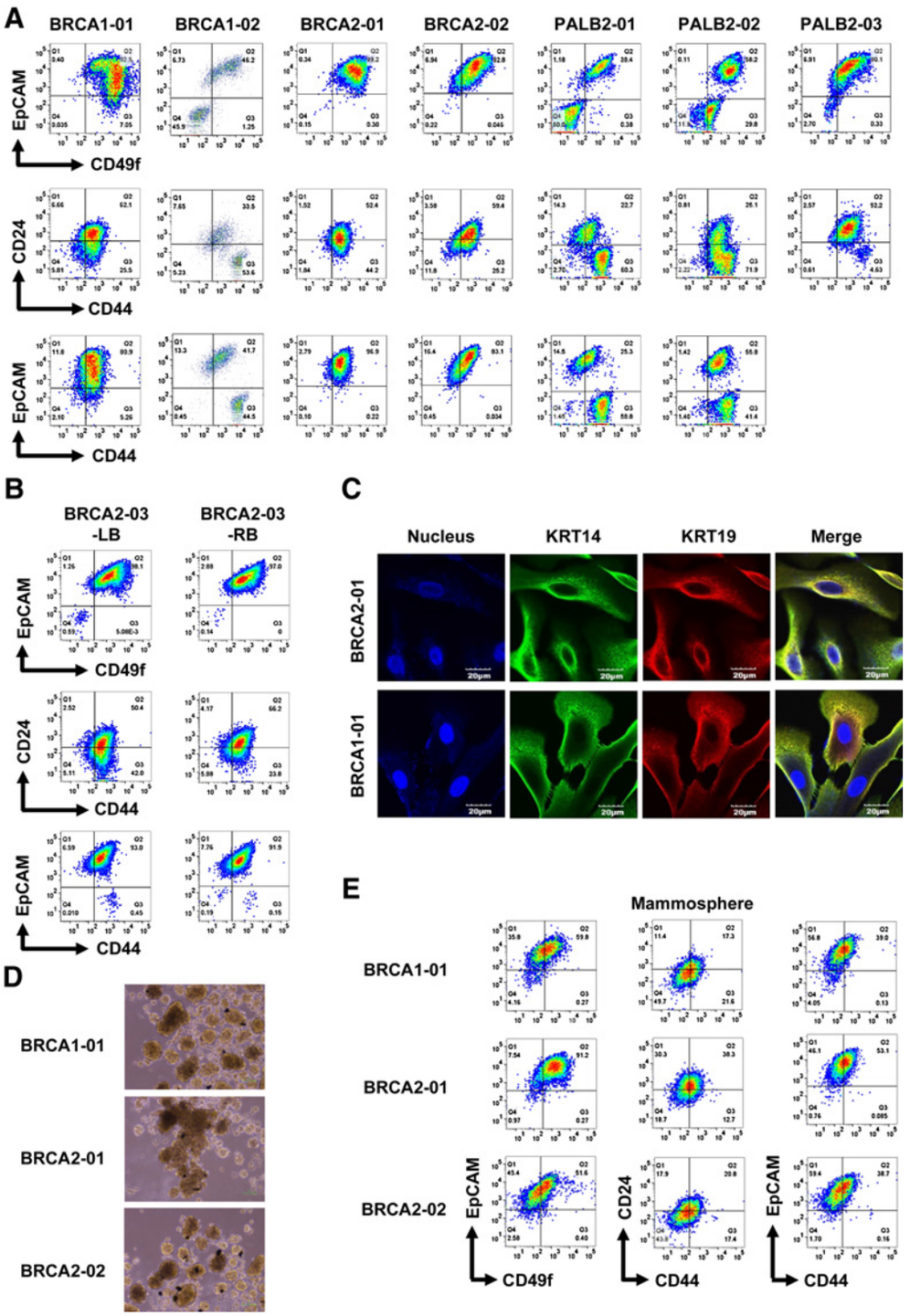


Figure 2.

Breast epithelial cells with luminal progenitor characteristics can be grown from breast cancer high-risk carriers. **A,** Breast epithelial cells from BRCA1, BRCA2, and PALB2 mutation carriers show variable subpopulation of cells based on CD49f/EpCAM, CD44/CD24, and CD44/EpCAM staining patterns. **B,** Breast epithelial cells propagated from randomly selected regions from two breasts of a BRCA2 mutation carrier show similar phenotypes based on CD49f/EpCAM, CD44/CD24, and CD44/EpCAM staining patterns. **C,** BRCA1 and BRCA2-mutant breast epithelial cells are enriched for KRT14^{+}/KRT19^{+} luminal progenitor cells as determined by immunofluorescence. **D,** BRCA1 and BRCA2-mutant cells form mammospheres. **E,** Cells in mammospheres show different degree of differentiation compared with cells grown in 2D culture. For example, CD49f^{-}/EpCAM^{+} cells that are considered differentiated luminal cells increased under mammosphere condition compared with 2D culture as shown in **A**.

Downloaded from mcr.aacrjournals.org on July 2, 2019. © 2019 American Association for Cancer Research.

Published OnlineFirst April 16, 2019; DOI: 10.1158/1541-7786.MCR-19-0165

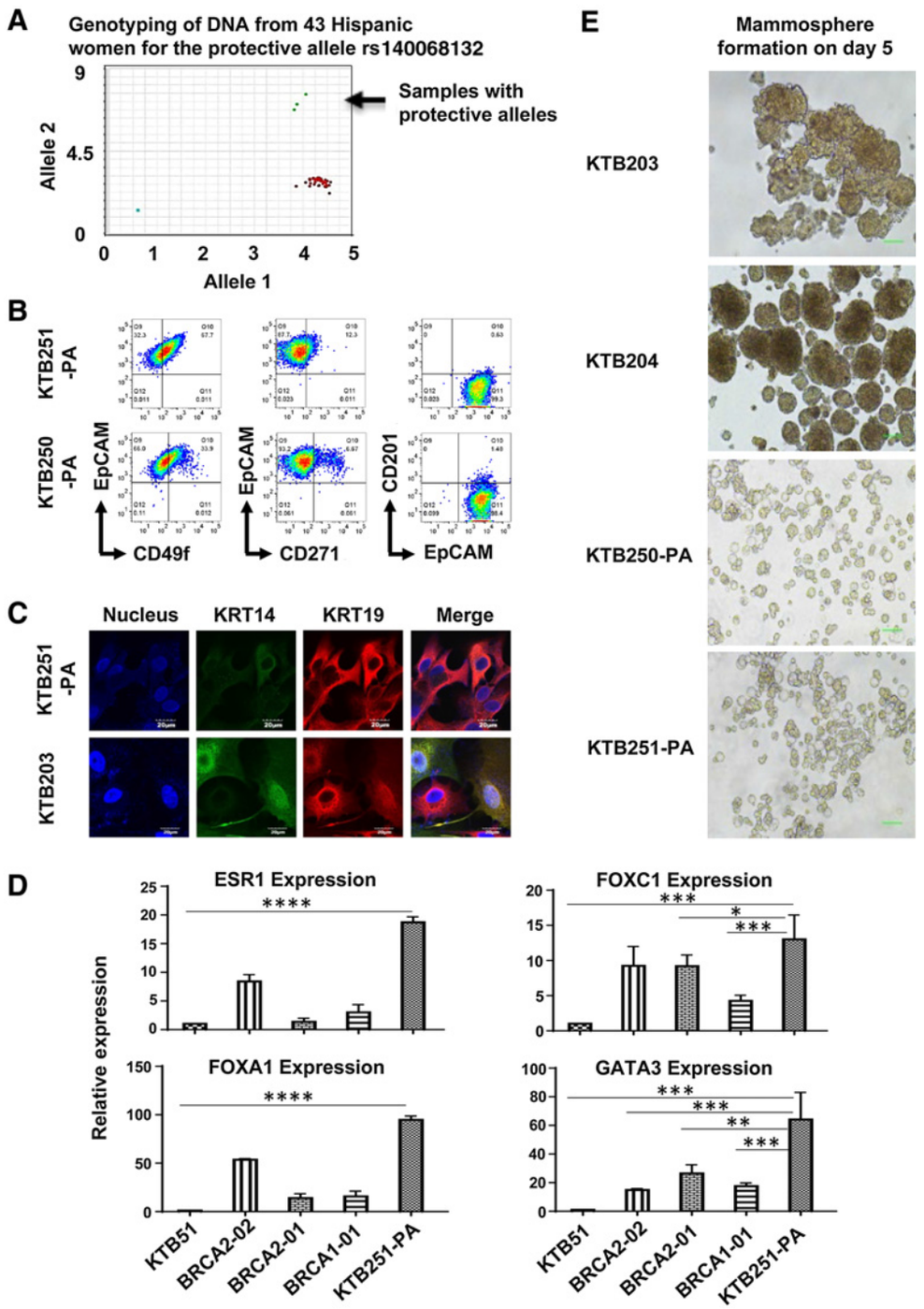


Figure 3.

Breast epithelial cells propagated from women with breast cancer–protective alleles show differentiated characteristics. **A,** Distribution patterns of the breast cancer–protective SNP rs140068132 in 43 self-reported Hispanic women. **B,** Breast epithelial cells from women with protective alleles show higher number of CD49f$^-$/EpCAM$^+$, CD271$^-$/EpCAM$^+$, and CD201$^-$/EpCAM$^+$ differentiated cells. **C,** Cells with protective alleles are predominantly KRT14$^-$/KRT19$^+$ differentiated cells compared with cells from another Hispanic women not carrying protective allele (KTB203). KTB203 contained both KRT14$^+$/KRT19$^+$ and KRT14$^-$/KRT19$^+$ cells. **D,** Cells with protective alleles express higher levels of luminal genes such as *ESR1, FOXA1*, and *GATA3* compared with noncarriers (KTB51) as well as BRCA1 and BRCA2-mutant carriers. mRNA levels measured through qRT-PCR are shown. Differences in expression between noncarrier (KTB51) and protective allele carrier (KTB251-PA) with all genes are statistically significant ($P < 0.05$, indicated by asterisks). Similarly, differences in expression between high-risk carriers and protective allele carrier are also significant. **E,** Breast epithelial cells from protective carriers generate small mammospheres compared with breast epithelial cells from Hispanic women without the protective allele.

Downloaded from mcr.aacrjournals.org on July 2, 2019. © 2019 American Association for Cancer Research.

Published OnlineFirst April 16, 2019; DOI: 10.1158/1541-7786.MCR-19-0165

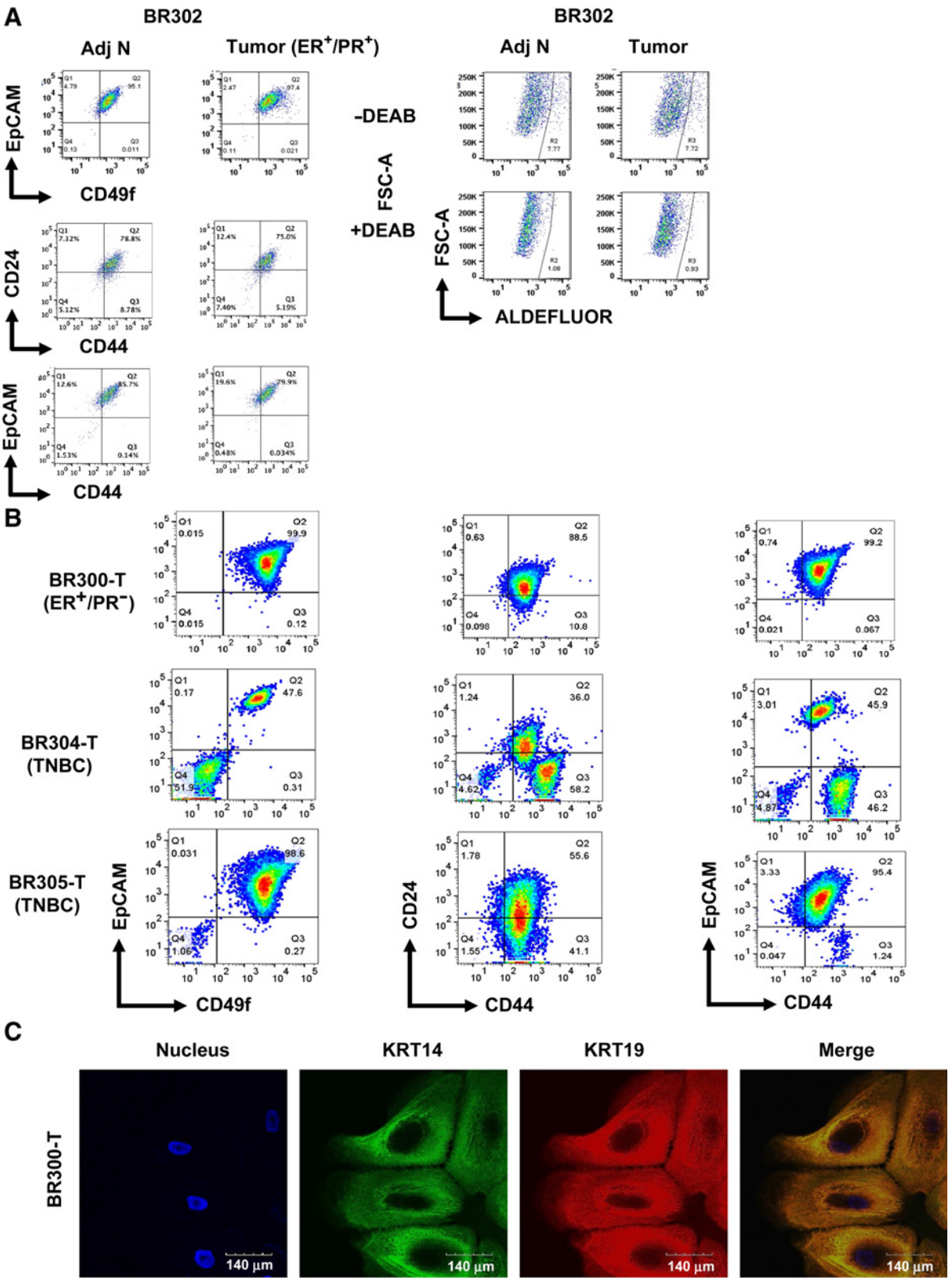


Figure 4.

Characteristics of cells propagated from tumors and tumor adjacent–normal cells. **A,** ER^+/PR^+ tumor cells and tumor adjacent–normal cells of the same patient show limited heterogeneity. **B,** Distinct profiles of tumor cells propagated from an ER^+/PR^- tumor and two TNBCs. **C,** KRT14 and KRT19 staining patterns of ER^+/PR^- tumor cells. On the basis of merged staining pattern, KRT19 staining appears dominant over KRT14.

Downloaded from mcr.aacrjournals.org on July 2, 2019. © 2019 American Association for Cancer Research.

Published OnlineFirst April 16, 2019; DOI: 10.1158/1541-7786.MCR-19-0165

breast ducts (Fig. 5A). Immunofluorescence of collagen gel structures with KRT14, KRT17, and KRT19 showed these ducts are KRT17$^+$ and are lined by KRT14$^+$/KRT19$^+$ double-positive cells (Fig. 5B). Most breast cancers are believed to originate from luminal progenitor cells, which are typically KRT14$^+$/KRT19$^+$ (23).

Breast cancer cells propagated from pleural effusions develop ERα-positive tumors in NSG mice

Although it has proven extremely difficult to cultivate breast epithelial cells that express ERα, there have been few recent advances in culture methods that enabled propagation of ERα-positive cells (2, 14). Similarly, patient-derived tumor xenograft (PDX) models heavily favor ERα-negative breast cancers relative to ERα$^+$ PDXs (32). In addition, very rarely, tumor cells and PDXs have been created from the same tumors. Because we had >90% success rate in generating cells from tumors, we next investigated whether cells derived from breast tumors or pleural effusions can create tumors in NSG mice. Detailed characteristics of tumor cells derived from pleural effusions developed from two ERα$^+$ tumors are described in Fig. 6A; Supplementary Fig. S5. Flow cytometry patterns of cells grown from two pleural effusions were different with pleural effusion #2 containing higher proportion of EpCAM$^+$ cells. Cells from both pleural effusions contained significant levels of ALDEFLUOR$^+$ cancer stem cells (33). Western blotting analysis showed expression of ERα in effusion #2, although very low levels of ERα were detectable in the effusion #1 (Fig. 6B). Cells derived from effusion #2 generated tumors in the mammary fat pad of NSG mice by two months postimplantation (Fig. 6C), which expressed all three markers of luminal breast cancer—ERα, FOXA1, and GATA3 but lacked progesterone receptor (PR), indicating luminal B characteristics (Fig. 6D). A lung metastatic nodule, which expressed GATA3, was also detected (Fig. 6E). We have generated tumors from primary TNBC cells, which could be reimplanted into additional animals similar to PDX models (Fig. 6F). Whole-genome sequencing of these TNBC cells showed mutation patterns observed in breast tumor samples (data not shown). Thus, the method described here provides resources of patient tumor–derived primary cells and tumors from these cells for functional studies.

Propagation of cells derived from liver metastasis and from other cancers

Metastasis is the main cause of cancer-related death and several studies have shown distinct differences between primary tumors and metastasis (34, 35). We had previously shown that epithelial reprogramming assay could be used to grow and perform genomics of liver metastasis of various cancers (36). However, because mouse embryonic feeder layer cells were needed to grow cancer cells, sorting of cancer cells by flow cytometry was required before sequencing, and bioinformatics analyses of genomic data required additional steps to ensure that sequence reads are not from residual contaminating mouse DNA. To determine whether metastatic cells grown under reprogramming growth conditions can adapt to the current method, we transferred cells from a liver metastasis of a breast cancer grown initially under reprogramming assay to the new method. Indeed, cells readily adapted to the new method and a significant number of cells could be obtained for phenotypic characterization (Fig. 7A; Supplementary Fig. S5). The majority of cells were CD49f$^+$/EpCAM$^+$, Jam-A$^+$/EpCAM$^+$, and CD271$^+$/EpCAM$^+$.

We extended this method to propagate cells from other cancer types. Cells from ascites fluid of patients with ovarian cancer (OV351-MT to OV353-MT) and liver metastasis of pancreatic cancer–neuroendocrine subtype were generated and characterized with various cell surface markers. Cells from ascites fluid of patients with ovarian cancer displayed the phenotypic heterogeneity and exhibited predominant CD49f$^+$/EpCAM$^-$, CD44$^+$/CD24-, CD44$^+$/EpCAM$^-$, CD73$^+$/CD90$^+$, CD201$^+$/EpCAM$^-$, CD166$^+$/EpCAM, and CD10$^+$/EpCAM$^-$ phenotypes (Fig. 7B; Supplementary Fig. S6). However, it is likely that few of the cell types that do not express EpCAM are noncancer stromal cells. Nonetheless, the method permitted propagation of EpCAM$^+$ cells, which can be sorted by flow cytometry for functional studies.

The cells from liver metastasis of pancreatic cancer–neuroendocrine subtype contained predominantly CD49f$^+$/EpCAM$^+$, CD44$^+$/CD24^{low}, CD44$^+$/EpCAM$^+$, CD73$^+$/CD90$^-$, CD201^{low}/EpCAM$^+$, CD271$^-$/EpCAM$^+$, CD10$^-$/EpCAM$^+$, CD166$^+$/EpCAM$^+$, CD117$^-$/EpCAM$^+$, MUC-1$^-$/EpCAM$^+$, JAM-1$^+$/EpCAM$^+$, and CD140b$^-$/EpCAM$^+$ subpopulations (Fig. 7C; Supplementary Fig. S6). A fraction of these cells were CXCR4-positive, which is expressed only in metastatic cancer stem cells of pancreatic cancer and is required for metastasis (37). Taken together, detailed characterization of various cells from different sources documents that our propagating method is suitable for growing primary cells, establishing heterogeneity of cancer cells in tumors, and provides a replenishable source of live primary cells for functional studies.

Discussion

Functional modeling of breast epithelial hierarchy and stromal–epithelial cell interactions have been difficult due to inability to obtain sufficient stem-progenitor-mature epithelial cells and stromal cells. Establishment of such techniques for modeling of the normal breast *in vitro* is critical if progress made in identifying various risk and protective alleles of cancers to have an impact in clinic. Similarly, further advancement is needed to translate knowledge gained through single-cell sequencing technologies, which are currently descriptive cataloguing of different cell types in primary and metastatic tumors without allowing functional evaluation of different cell types (35). In this study, we report *in vitro* propagation method for primary cells from different sources including breast, ovary, and metastasis. These cells are suitable for "omics" studies as TNBC cells from one of the patients have already been sequenced and found to have mutations enriched in breast cancers and additional samples are currently being sequenced (data not shown). Most importantly, sequence information can be obtained when primary tissue is very limited. We expect these tumor-derived cells to be also compatible for proteomics and metabolomics studies with and without prior exposure to chemotherapeutics. Other advantages of the method are that it is relatively inexpensive, adaptable in any labs, and cancer cells can be grown from cryopreserved tissues. We have found that cryopreserving tissues with ROCK inhibitor yields higher number of live cells (7). Although continuous passage leads to drift in proportion of different cell types, the original cell types remain after passage. Samples with disproportionately higher number of fibroblast-like cells compared with epithelial-like cells within a week of initiating culture, which can be identified through phase contrast microscopy, show a higher

Downloaded from mcr.aacrjournals.org on July 2, 2019. © 2019 American Association for Cancer Research.

Published OnlineFirst April 16, 2019; DOI: 10.1158/1541-7786.MCR-19-0165

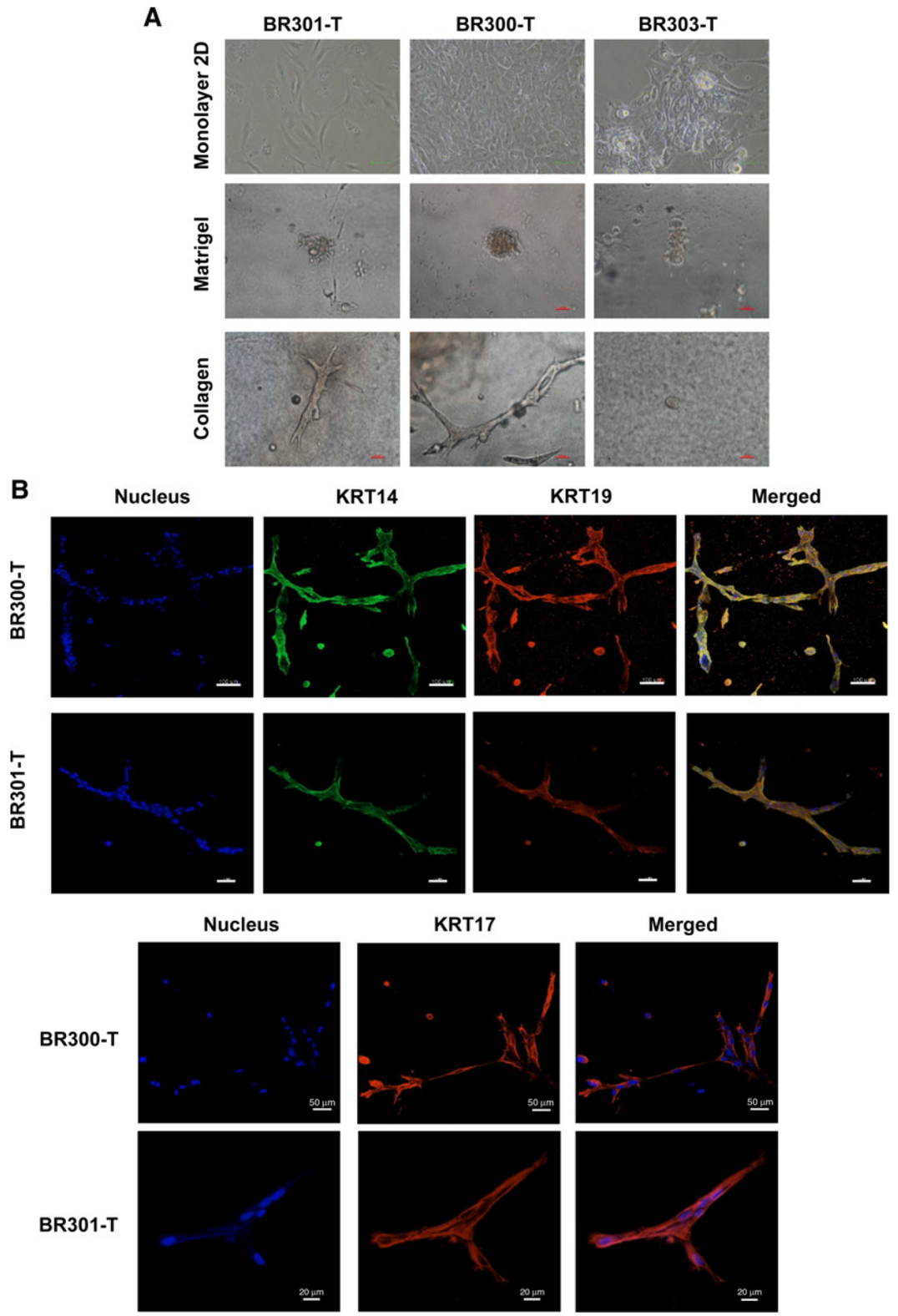


Figure 5.

Tumor cells form acinar structures on matrigel and ductal structures on collagen gel. **A,** Phase contrast images of tumor cells on monolayer, matrigel, and collagen gel. **B,** KRT14, KRT17, and KRT19 staining patterns of ductal structures on collagen gel.

Downloaded from mcr.aacrjournals.org on July 2, 2019. © 2019 American Association for Cancer Research.

Published OnlineFirst April 16, 2019; DOI: 10.1158/1541-7786.MCR-19-0165

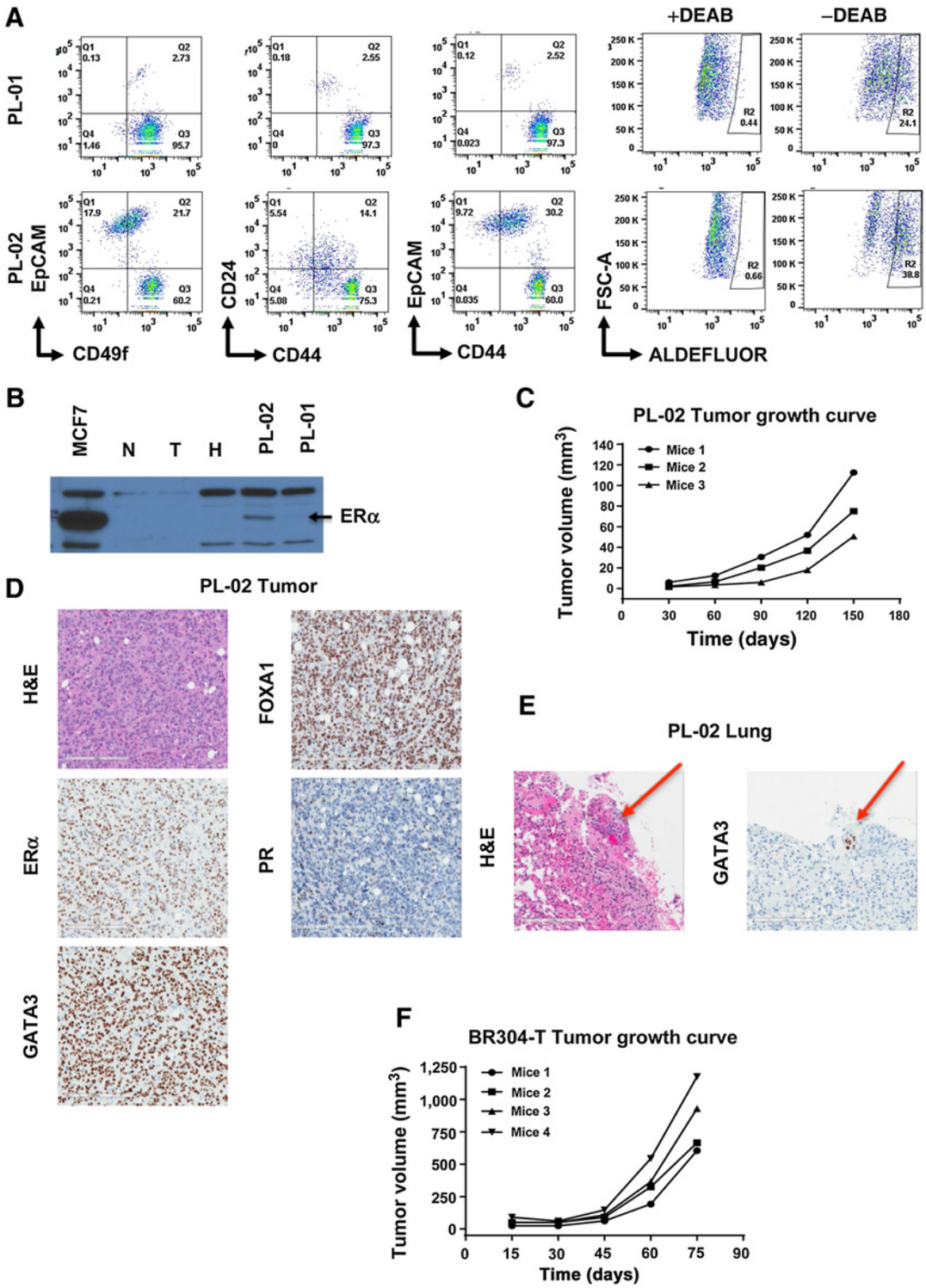


Figure 6.

Tumorigenic properties of tumor cells. **A,** Flow cytometry staining patterns of cells propagated from pleural effusions of patients who had ER^{+} breast cancer. **B,** ERα expression pattern in MCF7 cells (positive control), and cells from tumor adjacent–normal (N), tumor (T), KTB normal (H), and two pleural effusions. **C,** Growth characteristics of tumors developed in the mammary fat pad of NSG mice injected with cells from pleural effusion 2 (PL-02). **D,** H&E staining and luminal marker expression patterns in PL-02. **E,** PL-02–derived tumor metastasized to lungs (red arrow), which expressed GATA3. **F,** Growth characteristics of tumors derived from TNBC cells.

Downloaded from mcr.aacrjournals.org on July 2, 2019. © 2019 American Association for Cancer Research.

Published OnlineFirst April 16, 2019; DOI: 10.1158/1541-7786.MCR-19-0165

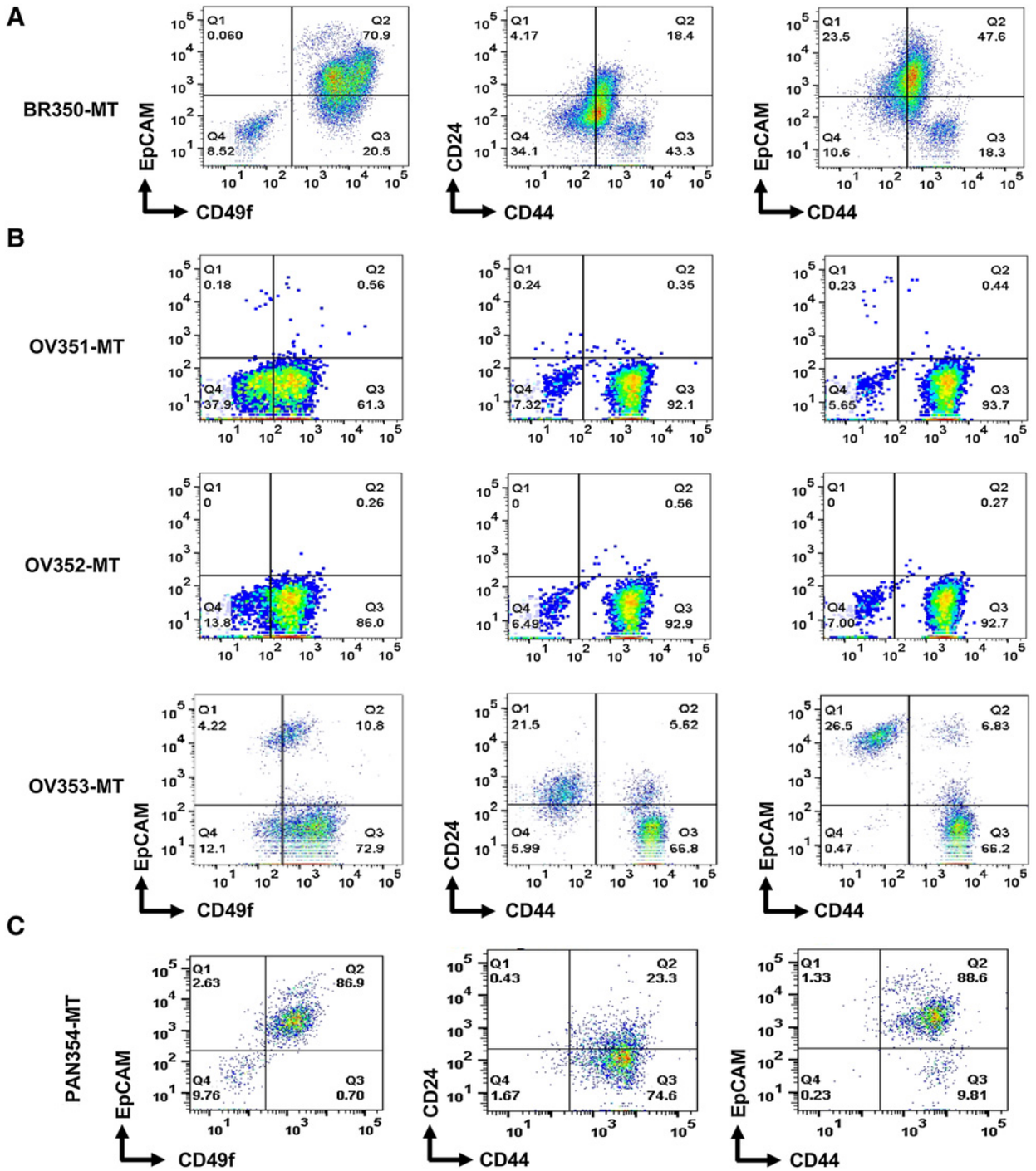


Figure 7.

Characterization of cells propagated from sites of metastasis of different cancers. **A,** Flow cytometry profiles of liver metastasis of a breast cancer. **B,** Flow cytometry profiles of cells propagated from ascites fluids of 3 patients with ovarian cancer. **C,** Flow cytometry profiles of liver metastasis of a pancreatic cancer (neuroendocrine).

tendency of drift to nonepithelial (EpCAM$^-$) cells with continuous passage. Differential trypsinization (two minutes at room temperature) or flow cytometry sorting can be used to enrich epithelial cells in culture. The only caveat of this modification is that cancer cells that have undergone EMT would get excluded from functional studies.

Single-cell studies are gaining popularity for better understanding of heterogeneity of cell types in an adult organ. Fifteen and

Downloaded from mcr.aacrjournals.org on July 2, 2019. © 2019 American Association for Cancer Research.

Published OnlineFirst April 16, 2019; DOI: 10.1158/1541-7786.MCR-19-0165

three epithelial cell types have been described in the mouse mammary gland and the human breast, respectively (38, 39). Because the assay described here allows propagation of cells from a core biopsy and cells form distinct structures on a 3D matrix, functional studies of cells from different clusters can now be performed with and without addition of stromal cells. Additional advantage of this system is the availability of cells that carry breast cancer risk alleles. Depending on the type of risk alleles, these cells can be easily immortalized with human telomerase. Although we could easily immortalize cells with BRCA1 or BRCA2 mutations, generating cell lines from PALB2 and CHEK2 mutations carriers as well as those with protective alleles has so far proved difficult. Phenotypic analyses using approximately 10 cell surface marker combinations did not reveal unique features of cells obtained from high-risk individuals, which is bit surprising considering a previous study, which demonstrated enrichment of cells with luminal progenitor properties in the breast of BRCA1-mutant carriers (17). However, we were able to document enhanced differentiation properties of cells obtained from women who carry breast cancer–protective alleles, which clearly indicates the role of proper differentiation in protecting against breast cancer.

We had previously demonstrated that short-term culturing followed by sequencing of cultured tumor cells could detect functionally important mutations that are otherwise undetectable upon direct sequencing of tumors (36). Culturing for short duration as such did not introduce mutations. Culturing method used at that time required propagation of tumor cells on a mouse fibroblast feeder layer and purifying tumor cells by flow cytometry before sequencing. The new method described here eliminates those requirements. Despite TNBCs displaying higher rate of mutations than other breast cancer subtypes (40) and data from many samples already available, very few are detected across multiple TNBCs and very rarely the impact of these mutations using the same tumor material could be studied. In general, cancers believed to contain 2–8 "driver" mutations and 30–60 protein-coding changes in passenger genes that alter cellular functions (41). We hope that the assay described here would not only enhance detection of additional mutations but also evaluation of presumed driver and passenger mutations in future. Furthermore, new mutations detected from these cell lines can be incorporated for screening of distant metastasis using ctDNA technology (42). In this respect, tumor cells generated from our culturing method form unique structures under 3D growth conditions and develop tumors in NSG mice, similar to a direct PDX model. Availability of millions of cancer cells from patients would permit further refinement of these models including characterizing growth properties of specific subclones with or without coculture with stromal cells, *in vitro* high-throughput drug screening on an individual patient basis and further validation of the effective drugs in xenograft models. Sufficient cells can be obtained within a month of initiating culture for rapid *in vitro* screening for drug sensitivity. In this respect, a recent study that utilized single-cell genomics and TNBC samples from patients receiving neoadjuvant therapy concluded that drug-resistant genotypes are preexisting in tumors and adaptively selected during neoadjuvant therapy (43). Relatively inexpensive way to generate cells from TNBCs described here would enable identification and characterization of such preexisting drug-resistant clones. While breast cancer research community has placed significant emphasis on finding cure for TNBCs, similar degree of attention is also needed for ERα-positive breast cancer because of higher number of patients with this disease type and its late recurrence (44). Recent studies have identified set of genes that predict early and late recurrence (45). However, developing *in vitro* models with patient-derived cells to functionally characterize genes in the signature has proven to be a challenge. Our assay has overcome this challenge as the cultured cells derived from recurrent tumors maintain ERα and one of our future goals is to establish cultures from metastasis that carry mutations in ERα, which are frequently observed in metastasis but not in primary tumors (46). We do acknowledge that, similar to PDX model (47), clonal selection of different tumor cells during culturing is one of the limitations of this assay. Nonetheless, simplicity of the assay allows propagation of cells from multiple cancers and serves as a resource for validation of pan-cancer genomic signatures including DNA methylation and chromatin accessibility identified through bioinformatics analyses (48–50).

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Authors' Contributions

Conception and design: K.D. Miller, H. Nakshatri
Development of methodology: M. Prasad, B. Kumar, M. Anjanappa, G. Sandusky, H. Nakshatri
Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): M. Prasad, B. Kumar, M. Anjanappa, G. Sandusky, A.M. Storniolo, H. Nakshatri
Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): M. Prasad, B. Kumar, M. Anjanappa, G. Sandusky, H. Nakshatri
Writing, review, and/or revision of the manuscript: M. Prasad, B. Kumar, K.D. Miller, H. Nakshatri
Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): P. Bhat-Nakshatri, K.D. Miller, H. Nakshatri
Study supervision: H. Nakshatri

Acknowledgments

We thank Dr. Rajagopal from Harvard Medical School for 804G cell line. We thank the members of IU Simon Cancer Center tissue procurement facility, flow cytometry core, clinical trial office, IU School of Medicine IHC core, Komen Tissue Bank for their service, and a countless number of women who donated their breast tissue for research purposes as well as volunteers who facilitated tissue collection. Department of Defense DOD-W81XWH-15-1-0707, Susan G. Komen for the Cure SAC110025, 100 Voices of Hope, and R03CA195250-01A1 (to H. Nakshatri) funded this study. Susan G. Komen for the Cure, Breast Cancer Research Foundation, Vera Bradley Foundation for Breast Cancer Research, and the Catherine Peachey Fund provided funding support to the Komen Tissue Bank.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received February 8, 2019; revised March 18, 2019; accepted April 12, 2019; published first April 16, 2019.

References

1. Visvader JE, Stingl J. Mammary stem cells and the differentiation hierarchy: current status and perspectives. Genes Dev 2014;28:1143–58.
2. Kumar B, Prasad MS, Bhat-Nakshatri P, Anjanappa M, Kalra M, Marino N, et al. Normal breast-derived epithelial cells with luminal

Downloaded from mcr.aacrjournals.org on July 2, 2019. © 2019 American Association for Cancer Research.

Published OnlineFirst April 16, 2019; DOI: 10.1158/1541-7786.MCR-19-0165

and intrinsic subtype-enriched gene expression document inter-individual differences in their differentiation cascade. Cancer Res 2018;78:5107–23.
3. Liu X, Ory V, Chapman S, Yuan H, Albanese C, Kallakury B, et al. ROCK inhibitor and feeder cells induce the conditional reprogramming of epithelial cells. Am J Pathol 2012;180:599–607.
4. Yang H, Qiu Y, Zeng X, Ding Y, Zeng J, Lu K, et al. Effect of a feeder layer composed of mouse embryonic and human foreskin fibroblasts on the proliferation of human embryonic stem cells. Exp Ther Med 2016;11:2321–8.
5. Cimmino L, Neel BG, Aifantis I. Vitamin C in stem cell reprogramming and cancer. Trends Cell Biol 2018;28:698–708.
6. Mou H, Vinarsky V, Tata PR, Brazauskas K, Choi SH, Crooke AK, et al. Dual SMAD signaling inhibition enables long-term expansion of diverse epithelial basal cells. Cell Stem Cell 2016;19:217–31.
7. Nakshatri H, Anjanappa M, Bhat-Nakshatri P. Ethnicity-dependent and -independent heterogeneity in healthy normal breast hierarchy impacts tumor characterization. Sci Rep 2015;5:13526.
8. Reiter JG, Makohon-Moore AP, Gerold JM, Heyde A, Attiyeh MA, Kohutek ZA, et al. Minimal functional driver gene heterogeneity among untreated metastases. Science 2018;361:1033–37.
9. Langhofer M, Hopkinson SB, Jones JC. The matrix secreted by 804G cells contains laminin-related components that participate in hemidesmosome assembly *in vitro*. J Cell Sci 1993;105:753–64.
10. Bhat-Nakshatri P, Goswami CP, Badve S, Sledge GW Jr, Nakshatri H. Identification of FDA-approved drugs targeting breast cancer stem cells along with biomarkers of sensitivity. Sci Rep 2013;3:2530.
11. Linnemann JR, Miura H, Meixner LK, Irmler M, Kloos UJ, Hirschi B, et al. Quantification of regenerative potential in primary human mammary epithelial cells. Development 2015;142:3239–51.
12. Kumar S, Kishimoto H, Chua HL, Badve S, Miller KD, Bigsby RM, et al. Interleukin-1 alpha promotes tumor growth and cachexia in MCF-7 xenograft model of breast cancer. Am J Pathol 2003;163:2531–41.
13. Roy S, Gascard P, Dumont N, Zhao J, Pan D, Petrie S, et al. Rare somatic cells from human breast tissue exhibit extensive lineage plasticity. Proc Natl Acad Sci U S A 2013;110:4598–603.
14. Fridriksdottir AJ, Kim J, Villadsen R, Klitgaard MC, Hopkinson BM, Petersen OW, et al. Propagation of oestrogen receptor-positive and oestrogen-responsive normal human breast cells in culture. Nat Commun 2015;6:8786.
15. Lathia JD, Li M, Sinyuk M, Alvarado AG, Flavahan WA, Stoltz K, et al. High-throughput flow cytometry screening reveals a role for junctional adhesion molecule as a cancer stem cell maintenance factor. Cell Rep 2014;6:117–29.
16. Chang CH, Hale SJ, Cox CV, Blair A, Kronsteiner B, Grabowska R, et al. Junctional adhesion molecule-A is highly expressed on human hematopoietic repopulating cells and associates with the key hematopoietic chemokine receptor CXCR4. Stem Cells 2016;34:1664–78.
17. Lim E, Vaillant F, Wu D, Forrest NC, Pal B, Hart AH, et al. Aberrant luminal progenitors as the candidate target population for basal tumor development in BRCA1 mutation carriers. Nat Med 2009;15:907–13.
18. Skibinski A, Breindel JL, Prat A, Galvan P, Smith E, Rolfs A, et al. The Hippo transducer TAZ interacts with the SWI/SNF complex to regulate breast epithelial lineage commitment. Cell Rep 2014;6:1059–72.
19. Park SY, Lee HE, Li H, Shipitsin M, Gelman R, Polyak K. Heterogeneity for stem cell-related markers according to tumor subtype and histologic stage in breast cancer. Clin Cancer Res 2010;16:876–87.
20. Alam M, Rajabi H, Ahmad R, Jin C, Kufe D. Targeting the MUC1-C oncoprotein inhibits self-renewal capacity of breast cancer cells. Oncotarget 2014;5:2622–34.
21. Hewitt KJ, Shamis Y, Knight E, Smith A, Maione A, Alt-Holland A, et al. PDGFRbeta expression and function in fibroblasts derived from pluripotent cells is linked to DNA demethylation. J Cell Sci 2012;125:2276–87.
22. Kim J, Villadsen R. Expression of luminal progenitor marker CD117 in the human breast gland. J Histochem Cytochem 2018;66:879–88.
23. Villadsen R, Fridriksdottir AJ, Ronnov-Jessen L, Gudjonsson T, Rank F, LaBarge MA, et al. Evidence for a stem cell hierarchy in the adult human breast. J Cell Biol 2007;177:87–101.
24. Rahman N. Realizing the promise of cancer predisposition genes. Nature 2014;505:302–8.
25. Michaloglou C, Vredeveld LC, Soengas MS, Denoyelle C, Kuilman T, van der Horst CM, et al. BRAFE600-associated senescence-like cell cycle arrest of human naevi. Nature 2005;436:720–4.
26. Fejerman L, Ahmadiyeh N, Hu D, Huntsman S, Beckman KB, Caswell JL, et al. Genome-wide association study of breast cancer in Latinas identifies novel protective variants on 6q25. Nat Commun 2014;5:5260.
27. Lappalainen T, Sammeth M, Friedlander MR, t Hoen PA, Monlong J, Rivas MA, et al. Transcriptome and genome sequencing uncovers functional variation in humans. Nature 2013;501:506–11.
28. Anjanappa M, Cardoso A, Cheng L, Mohamad S, Gunawan A, Rice S, et al. Individualized breast cancer characterization through single-cell analysis of tumor and adjacent normal cells. Cancer Res 2017;77:2759–69.
29. Teschendorff AE, Gao Y, Jones A, Ruebner M, Beckmann MW, Wachter DL, et al. DNA methylation outliers in normal breast tissue identify field defects that are enriched in cancer. Nat Commun 2016;7:10478.
30. Nakshatri H, Kumar B, Burney HN, Cox ML, Jacobsen M, Sandusky GE, et al. Genetic ancestry-dependent differences in breast cancer-induced field defects in the tumor-adjacent normal breast. Clin Cancer Res 2019;25:2848–59.
31. Takebe T, Wells JM, Helmrath MA, Zorn AM. Organoid center strategies for accelerating clinical translation. Cell Stem Cell 2018;22:806–9.
32. Whittle JR, Lewis MT, Lindeman GJ, Visvader JE. Patient-derived xenograft models of breast cancer and their predictive power. Breast Cancer Res 2015;17:17.
33. Ginestier C, Hur MH, Charafe-Jauffret E, Monville F, Dutcher J, Brown M, et al. ALDH1 is a marker of normal and malignant human mammary stem cells and a predictor of poor clinical outcome. Cell Stem Cell 2007;1:555–67.
34. Hunter KW, Amin R, Deasy S, Ha NH, Wakefield L. Genetic insights into the morass of metastatic heterogeneity. Nat Rev Cancer 2018;18:211–23.
35. Lawson DA, Kessenbrock K, Davis RT, Pervolarakis N, Werb Z. Tumour heterogeneity and metastasis at single-cell resolution. Nat Cell Biol 2018;20:1349–60.
36. Anjanappa M, Hao Y, Simpson ER, Bhat-Nakshatri P, Nelson JB, Tersey SA, et al. A system for detecting high impact-low frequency mutations in primary tumors and metastases. Oncogene 2018;37:185–96.
37. Hermann PC, Huber SL, Herrler T, Aicher A, Ellwart JW, Guba M, et al. Distinct populations of cancer stem cells determine tumor growth and metastatic activity in human pancreatic cancer. Cell Stem Cell 2007;1:313–23.
38. Bach K, Pensa S, Grzelak M, Hadfield J, Adams DJ, Marioni JC, et al. Differentiation dynamics of mammary epithelial cells revealed by single-cell RNA sequencing. Nat Commun 2017;8:2128.
39. Nguyen QH, Pervolarakis N, Blake K, Ma D, Davis RT, James N, et al. Profiling human breast epithelial cells using single cell RNA sequencing identifies cell diversity. Nat Commun 2018;9:2028.
40. Cancer Genome Atlas Network. Comprehensive molecular portraits of human breast tumours. Nature 2012;490:61–70.
41. Vogelstein B, Papadopoulos N, Velculescu VE, Zhou S, Diaz LA Jr, Kinzler KW. Cancer genome landscapes. Science 2013;339:1546–58.
42. Bardelli A, Pantel K. Liquid biopsies, what we do not know (yet). Cancer Cell 2017;31:172–79.
43. Kim C, Gao R, Sei E, Brandt R, Hartman J, Hatschek T, et al. Chemoresistance evolution in triple-negative breast cancer delineated by single-cell sequencing. Cell 2018;173:879–93.
44. Pan H, Gray R, Braybrooke J, Davies C, Taylor C, McGale P, et al. 20-year risks of breast-cancer recurrence after stopping endocrine therapy at 5 years. N Engl J Med 2017;377:1836–46.
45. Buus R, Yeo B, Brentnall AR, Klintman M, Cheang MCU, Khabra K, et al. Novel 18-gene signature for predicting relapse in ER-positive, HER2-negative breast cancer. Breast Cancer Res 2018;20:103.
46. Toy W, Weir H, Razavi P, Lawson M, Goeppert AU, Mazzola AM, et al. Activating ESR1 mutations differentially affect the efficacy of ER antagonists. Cancer Discov 2017;7:277–87.

Downloaded from mcr.aacrjournals.org on July 2, 2019. © 2019 American Association for Cancer Research.

Published OnlineFirst April 16, 2019; DOI: 10.1158/1541-7786.MCR-19-0165

47. Eirew P, Steif A, Khattra J, Ha G, Yap D, Farahani H, et al. Dynamics of genomic clones in breast cancer patient xenografts at single-cell resolution. Nature 2015;518:422–6.
48. Saghafinia S, Mina M, Riggi N, Hanahan D, Ciriello G. Pan-cancer landscape of aberrant DNA methylation across human tumors. Cell Rep 2018; 25:1066–80.
49. Corces MR, Granja JM, Shams S, Louie BH, Seoane JA, Zhou W, et al. The chromatin accessibility landscape of primary human cancers. Science 2018; 362:pii: eaav1898.
50. Dhawan A, Scott JG, Harris AL, Buffa FM. Pan-cancer characterisation of microRNA across cancer hallmarks reveals microRNA-mediated downregulation of tumour suppressors. Nat Commun 2018;9:5228.

Downloaded from mcr.aacrjournals.org on July 2, 2019. © 2019 American Association for Cancer Research.

CANCER DETECTION UTILIZING NORMAL TISSUE ADJACENT TO BREAST TUMORS WITH GENETIC ANCESTRY-MAPPING

CROSS-REFERENCE TO RELATED APPLICATIONS

This application claims priority to U.S. Provisional Patent Application Nos. 62//773,457, filed November 30, 2018, and 62/795218, filed January 22, 2019, the contents of each are incorporated by reference in their entirety into the present application.

STATEMENT OF GOVERNMENT SUPPORT

This invention was made with government support under W81XWH-15-1-0707 awarded by U.S. Army Medical Research & Material Command. The government has certain rights in the invention.

BACKGROUND OF THE DISCLOSURE

Recent data demonstrating a correlation between lymph node positivity at the time of detection, and the probability of disease recurrence even decades post detection, only solidifies the principle that the detection of breast cancer prior to lymph node metastasis can appreciably improve clinical outcomes.

Although the last decade witnessed significant improvements in imaging technologies including 3D-mammography, false negatives remain a significant concern. One way to overcome these false negatives is to complement radiologic techniques with molecular assays that measure the “transcriptomic and epigenetic field effect” or the “cancer induced field effect” of tumors on “normal” adjacent tissues (NATs). Teschendorff et al (Nature Communications (2016) 7:10478) demonstrated tumor-induced epigenetic field defects in the NATs and specifically, the targeting of transcription factor binding sites specifying chromatin architecture and stem cell differentiation pathways including Wnt and FGF signaling networks. Unfortunately, the Tumor Genome Atlas (TCGA) of breast cancer utilized reduction mammoplasty or NATs as their controls in transcriptome analyses, which are often used as “normal” controls for comparative analyses of tumors. This limitation was highlighted in another study, which compared TCGA “normal” breast transcriptome with the transcriptome of epithelial cells from normal breast tissues from healthy women. Significant differences were noted between these two sources of normal

tissues. Reduction mammoplasty samples are also histologically abnormal compared to breast tissues from healthy women.

While molecular markers of cancers, particularly gene expression signatures, were traditionally developed by comparing gene expression between healthy and cancer tissues, the effects of genetic ancestry on tumor evolution and gene expression are just beginning to be recognized. This observation is highly relevant in the context of known differences in cancer incidence and outcome based on genetic ancestry. For example, women of African Ancestry (AA) suffer higher mortality from the aggressive breast cancer subtype, triple negative breast cancer (TNBC), than women of European Ancestry (EA). Furthermore, basal-like and mesenchymal stem-like TNBCs are significantly more common in AA women. By contrast, breast cancer in Hispanic and Native American women is less prevalent and these women have better outcomes. Whether the worse outcome in AA women is due to an increased incidence of TNBC or unique biological factors that promote aggressive biology is an important but unresolved challenge in cancer disparities research.

Dietze et al. (Nat Rev Cancer (2015) 5(4):248-54) recently highlighted that key molecular pathways, including Aurora A-PLK, EZH2, and Wnt-stem cell signaling networks, are significantly up-regulated in TNBCs of AA women compared to TNBCs of women of EA. Dietze et al. further emphasized that it remains unknown whether genomic aberrations unique to AA women TNBCs result in activation of these signaling pathways or whether basal activity of these pathways in normal breast tissue of AA women is inherently different compared to EA women breast tissue. It remains possible that normal breast tissue biology varies based on genetic ancestry. Evidence for this possibility comes from the recent discovery of breast cancer protective alleles in Latinas. Single nucleotide polymorphisms (SNPs) in the protective allele are located on gene regulatory regions affecting the expression of genes linked to differentiation. Studies have discovered enrichment of a unique population of cells in the normal breast tissue of AA women (Nakshatri et al., Scientific Reports (2015) 5:13526). Furthermore, a breast cancer susceptibility loci in AA women, potentially altering the expression levels of microRNA miR-3065, has recently been discovered (Bensen et al., Breast Cancer Res (2018) 20(1):45).

Breast cancer diagnosis prior to lymph node metastasis can appreciably improve clinical outcomes. While radiologic techniques have significantly improved early diagnosis, molecular markers that can complement radiologic techniques are

needed to reduce false positives or negatives. A need also exists to develop tests specific to the genetic ancestry of the patient.

SUMMARY

The present disclosure is directed to methods and compositions for detecting and/or determining a patient's risk of cancer. The methods utilize categorizing the source of biopsy samples based on patient ancestry mapping, in combination with cancer markers to identify patients at higher risk of cancer. More particularly, the present disclosure shows specific upregulation of zinc-finger E-box-binding homeobox 1 ("ZEB1") + cells in cancerous breast tissue in European Ancestry (EA) women but not in African Ancestry (AA) women. Accordingly, an increase in the number of ZEB1+ cells in the breast tissue can be an indicator of the presence of breast cancer in EA women. Breast tissue of AA women naturally contains higher numbers of ZEB1+ cells, which may contribute to the aggressive behavior of AA women's breast cancer. In addition, cancerous breast tissue in AA women shows reduced levels of forkhead box A1 ("FOXA1") + cells. Accordingly, staining for the presence of ZEB1+ and FOXA1+ cells by immunohistochemistry coupled with ancestry mapping allow in conjunction with radiological techniques can provide for an improved method of early detection of breast cancer in EA and AA women.

In accordance with one embodiment breast biopsies from the affected breast as well as from contralateral breast are assessed for the presence of ZEB1+ and FOXA1+ cells. DNA from the patient can be analyzed for ancestry markers using commercially available high-discriminatory 41-single nucleotide ancestry markers to ensure that self-reported ethnicity matches at the genomic level.

As disclosed herein both genetic ancestry and the use of breast tissue from clinically healthy women as controls influence the detection of cancer-induced changes in the breast tissue. Applicant shows that alterations in ZEB1+ cells in normal adjacent tissue (NATs) to tumors are observed predominantly in EA women, whereas FOXA1+ cells were altered in the NATs of AA women. Immune cell activation in the tumors, as well as, surrounding tissue show genetic ancestry dependent variations as evident from differences in PD1+ and PDL1+ cells in the NATs of AA women and EA women. Thus, biomarker discovery needs to consider not only sample size and statistical methods, but also genetic ancestry and relevant true normal control tissues.

In some embodiments, a method of reducing the number of cancer false positives detected in patients by screening an initial set of potential cancer or pre-cancer biopsy samples to identify samples associated with higher risk of cancer is provided. The method comprises the steps of: obtaining a set of potential cancer or pre-cancer biopsy samples recovered from patient tissues identified by radiology techniques; categorizing the source of said biopsy samples based on patient ancestry mapping; analyzing a breast tissue biopsy sample to determine the numbers of ZEB1+ cells and FOXA1+ cells; identifying patients of European Ancestry having upregulation of ZEB1+ cells relative to true normal breast tissue; identifying patients of African Ancestry reduced levels of FOXA1+ cells relative to true normal breast tissue, wherein said identified patient of European Ancestry and women of African Ancestry are assessed with a higher risk of cancer or higher risk of developing cancer than patients of European Ancestry having normal/average levels of ZEB1+ cells and patients of African Ancestry having normal/average levels of FOXA1+ cells.

In some embodiments, a method of analyzing potential cancer or pre-cancer biopsy samples of patients of European Ancestry or African Ancestry is provided. The method comprises obtaining a set of potential cancer or pre-cancer biopsy samples recovered from patient tissues identified by radiology techniques; categorizing said sample based on ancestry mapping of the patient source of said sample; and quantifying the number of cell that comprise a cancer associated biomarker selected from a group consisting of ZEB1, FOXA1, and GATA Binding Factor 3 ("GATA3") or a combination thereof. In one embodiment the method comprises quantifying the number of ZEB1+ cells and FOXA1+ cells in a tissue biopsy sample identified as potentially being cancerous by radiological techniques. In some embodiments, the method comprises obtaining a set of potential cancer or pre-cancer biopsy samples recovered from patient tissue identified by radiology techniques; quantifying the expression level of Proto-Oncogene C-Met ("c-MET") in the biopsy sample, wherein the patient is duffy-null or duffy-heterozygous.

In some embodiments, an improved method of treating cancer in a patient of European Ancestry or African Ancestry is provided. The method comprises the steps of: obtaining a tissue biopsy sample identified as potentially being cancerous or pre-cancerous by radiological techniques and categorized as being from a patient of European Ancestry or African Ancestry; analyzing said sample for upregulation of ZEB1+ cells and reduced levels of FOXA1+ cells compared to a relevant true normal

sample or relevant true normal population data; identifying patients of European Ancestry and having specific upregulation of ZEB1+ cells; identifying patients of African Ancestry and having reduced levels of FOXA1+ cells; and treating said identified patients of European Ancestry and patients of African Ancestry using anti-cancer therapies.

In some embodiments, an improved method of analyzing a pre-cancer or cancerous biopsy in a patient who is duffy-null is provided. The method comprises the steps of: analyzing a tissue biopsy sample obtained from said patient and identified as cancerous or pre-cancerous using radiology techniques; and quantifying the levels of phosphorylated c-MET in the patient sample.

In some embodiments, an improved method of analyzing a pre-cancerous or cancerous biopsy sample from a patient of European Ancestry is provided. The method comprises the steps of: analyzing the biopsy sample identified as pre-cancerous or cancerous using radiology techniques; quantifying the levels of GATA3+ cells; and comparing the level of GATA3+ cells to a relevant true normal breast tissue sample or a relevant true normal population data.

In some embodiments, a kit for detecting and quantitating the relative number of ZEB1+ cells and FOXA1+ cells in a tissue sample recovered from a patient is provided. The kit comprises a reagent for the detection of ZEB1+ cells; and a reagent for the detection of FOXA1+ cells. In one embodiment the reagent is an antibody specific for the respective ZEB1 and FOXA1 markers. In one embodiment, the kit further comprises reagents for the detection of GATA3+ cells, optionally wherein the reagent is an antibody specific for the GATA3 marker.

In accordance with one embodiment an improved method of treating cancer in a patient is provided. The method comprises the steps of categorizing said patient based on ancestry mapping and analyzing a tissue biopsy sample recovered from the patient for upregulation of ZEB1+ cells and reduced levels of FOXA1+ cells relative to normal breast tissue. Patients that are identified as being of European Ancestry and having specific upregulation of ZEB1+ cells as well as patients identified as being of African Ancestry and having reduced levels of FOXA1+ cells are then provided with anti-cancer therapies known to those skilled in the art. Such therapies include radiation and/or chemotherapy. In one embodiment the tissue biopsy sample is recovered from human female breast tissue.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1A-1D: shows ZEB1 expression pattern in KTB-normal, normal adjacent tissues (NATs), and breast tumors. Fig. 1A) shows a representative immunohistochemistry in KTB normal, NATs, and tumors of women of African Ancestry (AA) and European Ancestry (EA). Fig. 1B) are graphs presenting data that demonstrates the differences in ZEB1 expression (positivity and H-score) between KTB-normal of AA and EA women. Fig. 1C) are graphs presenting data that demonstrates the differences between KTB-normal and NATs in AA and EA women. Fig. 1D) are graphs presenting data that demonstrates the differences between NATs and tumors in AA and EA women.

FIG. 2A-2D: shows FOXA1 expression pattern in KTB-normal, normal adjacent to tumor (NATs) and in breast tumors. Fig. 2A) shows representative immunohistochemistry in KTB normal, NATs, and tumors of AA and EA women. Fig. 2B) are graphs presenting data that demonstrates the differences in FOXA1 expression (positivity and H-score) between KTB-normal of AA and EA women. Fig. 2C) are graphs presenting data that demonstrates the differences between KTB-normal and NATs in AA and EA women. Fig. 2D) are graphs presenting data that demonstrates the differences between NATs and tumors in AA and EA women.

FIG. 3A- 3D: shows ERα expression pattern in KTB-normal, normal adjacent to tumor (NATs) and in breast tumors. Fig. 3A) shows representative immunohistochemistry in KTB normal, NATs, and tumors of AA and EA women. Fig. 3B) are graphs presenting data that demonstrates the differences in ERα expression (positivity and H-score) between KTB-normal of AA and EA women. Fig. 3C) are graphs presenting data that demonstrates the differences between KTB-normal and NATs in AA and EA women. Fig. 3D) shows differences between NATs and tumors in AA and EA women.

FIG. 4: are graphs presenting data that present the statistical analyses of CD8 and CD68 positivity in NATs and tumors (T) of AA and EA women. All statistically significant differences are indicated with p values.

FIG. 5: are graphs presenting data that presents the statistical analyses of PD1 and PDL1 TPS scores in NATs and tumors (T) of AA and EA women. All statistically significant differences are indicated with p values.

FIG. 6 are bar graphs showing the ancestry marker distribution patterns of African American and Caucasian women who donated their breast tissue to KTB and were part of the TMA.

FIG. 7A – 7D shows the distribution of MSRB3 within normal, NATs, and tumor cells of AA women and EA women. Fig. 7A shows MSRB3 immunohistochemistry of normal tissue, NATs, and tumors of AA and EA women; Fig. 7B is a graph showing the difference in MSRB3 positivity between normal samples from AA and EA women; Fig. 7C provides two graphs showing the differences between MSRB3 expression in normal tissue and NATs in AA and EA women; and Fig. 7D shows two graphs that show the differences between GATA3 expression in NATs and tumors of AA and EA women.

FIG. 8A-8C are graphs presenting data regarding the expression levels of GATA3; wherein Fig. 8A shows the differences in GATA3 expression between normal tissue of AA and EA women; Fig. 8B shows differences between GATA3 expression in normal tissue and NATs of AA and EA women; and Fig. 8C shows the differences between GATA3 expression in NATs and tumors in AA and EA women.

FIG. 9 presents dot blot data demonstrating duffy-null epithelial cells display higher levels of tyrosine phosphorylated c-MET, IGFR1, and Insulin receptor.

FIG. 10 are photographs of expression data demonstrating duffy-null and duffy-heterozygous epithelial cells contain increased levels of phospho-c-MET and phosphor-Tyr307 GAB1.

DETAILED DESCRIPTION

ABBREVIATIONS:

The following abbreviations are used throughout the specification:

NATs: “normal” adjacent tissues;

TCGA: Tumor Genome Atlas;

AA: African Ancestry;

EA: European Ancestry;

TNBC: triple negative breast cancer;

TMA: tissue microarray.

DEFINITIONS

In describing and claiming the invention, the following terminology will be used in accordance with the definitions set forth below.

As used herein, the term "treating" includes prophylaxis of the specific disorder or condition, or alleviation of the symptoms associated with a specific disorder or condition and/or preventing or eliminating said symptoms.

As used herein the term "patient" without further designation is intended to encompass any warm blooded vertebrate domesticated animal (including for example, but not limited to livestock, horses, mice, cats, dogs and other pets) and humans.

As used herein, the term “cancer induced field effect” refers to the effect cancer cells have on surrounding phenotypically normal appearing tissues.

As used herein, the term “GATA3” represents GATA Binding Protein 3 or GATA Binding Factor 3. This protein belongs to a family of GATA transcription factors. A GATA3 + cell refers to a cell that expresses GATA3.

As used herein, the term “ERα” is an abbreviation for estrogen receptor alpha.

As used herein the term “ZEB1” is an abbreviation for zinc-finger E-box-binding homeobox 1. A ZEB1+ cell refers to a cell that expresses ZEB1.

As used herein the term “FOXA1” is an abbreviation for forkhead box A1 and is a transcriptional activator protein. FOXA1+ cells refer to cells that express FOXA1.

As used herein the term “c-MET” is an abbreviation for Tyrosine-Protein Kinase Met or Proto-Oncogene C-Met.

As used herein the term “NATs” defines normal adjacent tissue that is located near or next to a cancerous cell or tumor. A “matching NATs or matching NAT” refers to a normal adjacent tissue(s) and the tumor or cancerous cell coming from the same patient.

As used herein the term “Ancestry Mapping” defines a methodology for conducting a molecular analysis of nuclear and organelle genomic nucleic acid sequences recovered from an individual to determine the geographic origin of the ancestors of that individual.

As used herein the term “African Ancestry” is used to define an individual identified based on Ancestry Mapping as having ancestors having a geographic origin of Africa as described in Nievergelt CM et al., Investigative Genetics 2013, 4:13, the contents of which are incorporated herein by reference in its entirety. African Ancestry excludes a combination of genetic markers that would identify a person as predominately Latin Ancestry, Native American Ancestry, Indigent Ancestry, Asian Ancestry, or European Ancestry.

As used herein the term “European Ancestry” is used to define an individual identified based on Ancestry Mapping as having ancestors having a geographic origin of Europe as described in Nievergelt CM et al., Investigative Genetics 2013, 4:13. European Ancestry excludes a combination of genetic markers that would identify a person as predominately Latin Ancestry, Native American Ancestry, Indigent Ancestry, Asian Ancestry, or African Ancestry.

As used herein the term “duffy-null” defines a genetic mutation where the person is lacking expression of protein from both copies of the duffy gene. This mutation was originally discovered in a subset of people with Sub-Saharan African Ancestry.

As used herein the term “duffy-heterozygous” defines a genetic mutation where the person is lacking expression of protein from one copy of the duffy gene.

As used herein the term “TPS” describes a tumor proportion score, which is the percentage of viable tumor cells showing partial or complete membrane staining.

As used herein the term “true normal breast tissue sample” defines a breast tissue sample free of cancer or pre-cancer cells and free of cancer induced field effects. Further the sample is from another healthy individual of at least the same

species, gender, ancestry, and age. Additional criteria may be used to narrow down the scope of a true normal breast tissue sample including whether both the sample and the healthy individual were ever pregnant, breast-fed, or undergone any kind of mammoplasty as an example.

As used herein the term “true normal population data” defines an averaged data generated from samples recovered from numerous healthy individuals, wherein each sample obtained from said individuals is a true normal breast tissue sample. Further, the numerous healthy individuals would all be at least the same species, gender, ancestry, and age as the patient sample. Additional criteria may be used to narrow down the scope of a true normal population data including whether both the sample and the numerous individuals were ever pregnant, breast-fed, or undergone any kind of mammoplasty as an example.

As used herein the term “biopsy” defines any solid tissue sample recovered form a patient.

As used herein the term “increased levels” in reference to a biological marker, encompasses increased numbers of cells expressing a biomarker and/or increased total expression of a biomarker. As an illustrative example, increased levels include increased numbers of cells expressing ZEB1 and/or increased total expression of ZEB1 compared to the true normal sample or true normal population data.

As used here the term “decreased levels” in reference to a biological marker, encompasses a decrease in the number of cells expressing a biomarker and/or a decrease in the total expression of a biomarker. As an illustrative example, decreased levels include decreased numbers of cells expressing FOXA1 or GATA3 and/or decreased total expression of FOXA1 or GATA3 compared to the true normal sample or true normal population data.

EMBODIMENTS

In one embodiment the present disclosure is directed to an improved method for detecting cancer and reducing the number of false positives associated with traditional radiological techniques for detecting cancer.

The method combines quantitative assessment of cancer associated markers with ancestry information to more accurately assess the impact of altered expression levels relative to normal tissues recovered from healthy individuals having the same or similar ancestry. In one embodiment the method is used to enhance the accuracy of

known techniques for detecting solid tumors. Expression levels of the relevant cancer associated markers can be quantitatively measured by methods known by those skilled in the art such as, for example, immunohistochemistry, immunofluorescence, northern blotting, amplification, polymerase chain reaction, microarray analysis, tag-based technologies (e.g., serial analysis of gene expression and next generation sequencing such as whole transcriptome shotgun sequencing or RNA-Seq), Western blotting, enzyme linked immunosorbent assay (ELISA), *in situ* hybridization, and combinations thereof.

In some embodiments, the pre-cancer or cancer is breast cancer. In some embodiments, the pre-cancer or cancer is triple negative breast cancer, wherein the cancer cells test negative for estrogen receptors, progesterone receptors, and HER2 protein. In some embodiments, the pre-cancer or cancer is HER2+ breast cancer. In some embodiments, the pre-cancer or cancer is HER2- breast cancer. In some embodiments, the pre-cancer or cancerous cells are EpCAM -. In some embodiments, the pre-cancer or cancerous cells are ERα +. In some embodiments, the pre-cancer or cancerous cells are ERα-.

In some embodiments, the patient is male or female. In some embodiments, the patient is a female. In some embodiments, the patient is a woman.

In some embodiments, the method comprises detecting the presence, and optionally the relative concentration, of a cancer associated biomarker. In some embodiments, the cancer associated biomarker or "biomarker" is selected from a group consisting of ZEB1, GATA3, FOXA1, c-MET, ERα, PDLI, PD1, CD68, and CD8.

In one embodiment a method of analyzing potential cancer or pre-cancer biopsy samples of patients of European Ancestry or African Ancestry is provided. The method comprises obtaining a set of potential cancer or pre-cancer biopsy samples recovered from patient tissues identified by radiology techniques or other screening techniques; categorizing said sample based on ancestry mapping of the patient source of said sample; and quantifying the number of cells comprising a cancer associated biomarker selected from a group consisting of ZEB1, FOXA1, and GATA3, or a combination thereof. In one embodiment, the final step of the method comprises quantifying

i) the number of cells comprising the biomarker ZEB1; and the number of cells comprising the biomarker FOXA1; or

ii) the number of cells comprising the biomarker ZEB1; and the number of cells comprising the biomarker GATA3; or

iii) the number of cells comprising the biomarker FOXA1 and the number of cells comprising the biomarker GATA3; or

iv) the number of cells comprising the biomarker ZEB1; the number of cells comprising the biomarker FOXA1; and the number of cells comprising the biomarker GATA3. In one embodiment the method comprises quantifying the number of cells comprising the biomarker ZEB1; and the number of cells comprising the biomarker FOXA1.

In some embodiments, a method of reducing the number of cancer false positives detected in patients by screening an initial set of potential cancer or pre-cancer biopsy samples to identify samples associated with higher risk of cancer is provided. The method comprises the steps of: obtaining a potential cancer or pre-cancer biopsy sample from patient tissues identified as potentially cancerous or pre-cancerous by radiology techniques or use of associated cancer markers; categorizing the source of said biopsy samples based on patient ancestry mapping; analyzing the tissue biopsy sample to determine the numbers of ZEB1+ cells and FOXA1+ cells; identifying patients of European Ancestry having upregulation of ZEB1+ cells relative to true normal breast tissue; identifying patients of African Ancestry having reduced levels of FOXA1+ cells relative to true normal breast tissue, wherein said identified patient of European Ancestry and women of African Ancestry are assessed with a higher risk of cancer or higher risk of developing cancer than patients of European Ancestry having normal levels of ZEB1+ cells and patients of African Ancestry having normal levels of FOXA1+ cells. As used herein "normal" is defined as the range of values found in comparable healthy tissues of individuals free of disease. In one embodiment the normal range for a given test is based on the results that are seen in 95% of the healthy population. Identified individuals are those that have detected levels of the relevant marker outside the normal range of comparable (i.e. ancestry matched) sourced tissues. In one embodiment patients of European Ancestry having upregulation of ZEB1+ cells by more than 1%, 2%, 5%, 10% or more from the highest normal value for true normal breast tissue; and patients of African Ancestry having reduced levels of FOXA1+ cells by more than 1%, 2%, 5%, 10% or more from the lowest normal value for true normal breast tissue are identified as the identified individuals having an higher risk of having or developing cancer. In

some embodiments, the tissue biopsy sample is recovered from breast tissue. In some embodiments, the patient is a woman.

In some embodiments, a method of analyzing potential cancer or pre-cancer biopsy samples of patients of European Ancestry or African Ancestry is provided. The method comprises obtaining a set of potential cancer or pre-cancer biopsy samples recovered from patient tissues identified by radiology techniques; categorizing said sample based on ancestry mapping of the patient source of said sample; and quantifying the number of cells expressing a cancer associated biomarker selected from a group consisting of ZEB1, FOXA1, GATA3, and phospho-c-MET or a combination thereof.

In some embodiments the method further comprises comparing the number of cells expressing the biomarker in said sample with a relevant true normal sample or relevant true normal population data. In some embodiments, the biomarker is ZEB1. In some embodiments, the biomarker is FOXA1. In some embodiments, the biomarker is GATA3. In some embodiments, the biomarker is phospho-c-MET. In some embodiments the biomarker is ZEB1 and FOXA1. In some embodiments, the biomarker is selected from a group consisting of ZEB1, FOXA1, GATA3, and phospho-c-MET, or a combination thereof, and further includes at least one other PDL1, PD1, CD68, CD8, ERα, or a combination thereof.

In some embodiments, the relevant true normal sample is provided by a woman of African Ancestry. In some embodiments, a relevant true normal population data comprising samples from African Ancestry (AA) women is used.

In some embodiments, the relevant true normal sample is provided by a woman of European Ancestry. In some embodiments, a relevant true normal population data comprising samples from European Ancestry (EA) women is used.

In some embodiments, the method further comprises screening said biopsy sample for duffy-null or duffy-heterozygous alleles. In some embodiments, the method further comprises screening said biopsy sample for duffy-null alleles. In some embodiments, the method further comprises quantifying the expression level of phospho-c-MET in said biopsy sample. In some embodiments, the expression level of phospho-c-MET is compared to a matching NATs or a relevant true normal sample or relevant true normal population data.

In some embodiments, the biopsy sample is recovered from human female breast tissue.

In some embodiments, the biopsy sample is categorized as being either European Ancestry or African Ancestry.

In some embodiments, the number of ZEB1+ cells and FOXA1+ cells are compared to a relevant true normal sample or a relevant true normal population data. In some embodiments, the number of GATA3+ cells is compared to a relevant true normal sample or a relevant true normal population data. In some embodiments, the level of phospho-c-MET expression in the sample biopsy is compared with a relevant true normal sample or a relevant true normal population data.

In some embodiments, the method further comprises, comparing the number of ZEB1+ cells and FOXA1+ cells to a matching normal adjacent tissues (NATs).

In some embodiments, an improved method of treating cancer in a patient of European Ancestry or African Ancestry is provided. The method comprises the steps of: obtaining a tissue biopsy sample identified as being cancer by radiological techniques and categorized as being from a patient of European Ancestry or African Ancestry; analyzing said sample for upregulation of ZEB1+ cells and reduced levels of FOXA1+ cells compared to a relevant true normal sample or relevant true normal population data; identifying patients of European Ancestry and having specific upregulation of ZEB1+ cells; identifying patients of African Ancestry and having reduced levels of FOXA1+ cells; and treating said identified patients of European Ancestry and patients of African Ancestry using anti-cancer therapies. In some embodiments, the patient is a woman and the biopsy sample is recovered from breast tissue.

In some embodiments, a patient is identified as having, or at risk of developing, cancer, when a patient of European Ancestry has an increase of 1, 5, 10, 15, 20, 25, or 50% in the numbers of ZEB1+ cells relative to true normal breast tissue and when a patient of African Ancestry has a reduction of 1, 5, 10, 15, 20, 25, or 50% of the number of FOXA1+ cells relative to true normal breast tissue. In some embodiments, the ZEB1+ cells or FOXA1+ cells comprise the patient's normal adjacent tissue (NAT).

In some embodiments, an improved method of analyzing a pre-cancer or cancerous biopsy in a patient who is duffy-null is provided. The method comprises the steps of: analyzing a tissue biopsy sample obtained from said patient and identified as cancerous or pre-cancerous using radiology techniques; and quantifying the levels of phospho-c-MET in the patient sample. In some embodiments, the

method further comprises quantifying the levels of ZEB1+ cells in said biopsy sample. In some embodiments, the method further comprises comparing the levels of phospho-c-MET to a relevant true normal sample or a relevant true normal population data.

In some embodiments, a kit for detecting and quantitating the relative number of ZEB1+ cells and FOXA1+ cells in a tissue sample recovered from a patient is provided. In one embodiment the kit comprises a reagent for the detection of ZEB1+ cells; and a reagent for the detection of FOXA1+ cells. In some embodiments, the reagent for detecting ZEB1+ cells is an antibody specific for ZEB1. In some embodiments, the reagent for detecting FOXA1+ cells is an antibody specific for FOXA1.

In one embodiment, the kit comprises reagents for conducting the ancestry mapping as well as reagents for detecting and quantitating the presence of ZEB1+ cells and FOXA1+ cells in a tissue sample recovered from a patient. In accordance with one embodiment, a kit for detecting and quantitating the relative number of ZEB1+ cells and FOXA1+ cells in a tissue sample recovered from a patient comprises a reagent for the detection of ZEB1+ cells, and a reagent for the detection of FOXA1+ cells. In one embodiment, the reagents are antibodies. In one embodiment, the kit comprises one or more of the following antibodies: FOXA1 (Santa Cruz sc-6553), PD1 (Cell Marque 315M-98), PDL1 (Keytruda) (clone 22c3, Dako IHC 22c3) and ZEB1 (3G6, cat no 14-9741-82, eBioscience). The kit may further comprise reagents for conducting ancestry mapping. In some embodiments, the kit further comprises a reagent for detecting GATA3, including for example an antibody specific for detecting GATA3. In one embodiment the kit further comprises a reagent for detecting phospho-c-MET. In some embodiments, the reagent is an antibody for detecting phospho-c-MET. The kit typically further comprises instructional materials.

In some embodiments, an improved method of analyzing a pre-cancerous or cancerous biopsy sample from a patient of European Ancestry is provided. The method comprises the steps of analyzing the biopsy sample identified as pre-cancerous or cancerous using radiology techniques; quantifying the levels of GATA3+ cells; and comparing the level of GATA3+ cells to a relevant true normal breast tissue sample or a relevant true normal population data.

In some embodiments, identifying patients of European Ancestry having a reduction of 1, 5, 10, 15, 20, 25, or 50% of the number of GATA3+ cells is an indication of pre-cancer or cancer.

In an illustrative embodiment, decreased levels of GATA3+ cells is an indication of pre-cancer or cancer. In some embodiments, the method further comprises quantifying the levels of ZEB1+ cells and comparing the ZEB1+ cell level to a relevant true normal breast tissue sample.

Various functions and advantages of these and other embodiments of the present disclosure will be more fully understood from the examples shown below. The examples are intended to illustrate the benefits of the present disclosure, but do not exemplify the full scope of the disclosure.

CLAUSE LIST

Clause 1. A method of reducing the number of cancer false positives detected in patients by screening an initial set of potential cancer or pre-cancer biopsy samples to identify samples associated with higher risk of cancer, said method comprising the steps of

obtaining a set of potential cancer or pre-cancer biopsy samples recovered from patient tissues, optionally wherein the potential cancer or pre-cancer biopsy samples were identified by radiology techniques or other cancer detection techniques;

categorizing the source of said biopsy samples based on patient ancestry mapping;

analyzing the tissue biopsy sample to determine the numbers of ZEB1+ cells and FOXA1+ cells;

identifying patients of European Ancestry having upregulation of ZEB1+ cells relative to true normal tissue;

identifying patients of African Ancestry reduced levels of FOXA1+ cells relative to true normal tissue, wherein said identified patient of European Ancestry and women of African Ancestry are assessed with a higher risk of cancer or higher risk of developing cancer than patients of European Ancestry having normal levels of ZEB1+ cells and patients of African Ancestry having normal levels of FOXA1+ cells.

Clause 2. The method of clause 1 wherein a patient is identified as being of African Ancestry or European Ancestry based on analysis using a discriminative 41-ancestry marker profiles (see Nievergelt CM et al., Investigative Genetics 2013, 4:13).

Clause 3. The method of clause 1 or 2 wherein said tissue biopsy sample is recovered from breast tissue and said true normal tissue is true normal breast tissue.

Clause 4. The method of any one of clauses 1-3 wherein the patient is a woman.

Clause 5. A method of analyzing potential cancer or pre-cancer biopsy samples of patients of European Ancestry or African Ancestry, said method comprising:

obtaining a set of potential cancer or pre-cancer biopsy samples recovered from patient tissues identified by radiology techniques or other cancer associated marker;

categorizing said sample based on ancestry mapping of the patient source of said sample; and

quantifying the number of cells expressing ZEB1 and the number of cells expressing FOXA1.

Clause 6. The method of clause 5 further comprising a step of quantifying the number of cells expressing GATA3.

Clause 7. The method of clause 5 or 6 further comprising a step of quantifying the number of cells expressing phospho-c-MET.

Clause 8. A method of analyzing potential cancer or pre-cancer biopsy samples of patients of European Ancestry or African Ancestry, said method comprising:

obtaining a set of potential cancer or pre-cancer biopsy samples recovered from patient tissues identified by radiology techniques or other cancer associated marker;

categorizing said sample based on ancestry mapping of the patient source of said sample; and

quantifying the number of first cells expressing a cancer associated biomarker selected from a group consisting of ZEB1, FOXA1, GATA3, and phospho-c-MET and

quantifying the number of second cells expressing a cancer associated biomarker selected from a group consisting of ZEB1, FOXA1, GATA3, and phospho-c-MET, wherein the first and second cells express different markers selected from the group consisting of ZEB1, FOXA1, GATA3, and phospho-c-MET.

Clause 8.5 The method of clause 8 wherein said first cells express a marker selected from the group consisting of ZEB1, FOXA1, and GATA3 and said second cells express markers selected from the group consisting of ZEB1, FOXA1, and GATA3.

Clause 9. The method of any one of clause 5 to 8 further comprising comparing the biomarker in said sample with a relevant true normal sample or relevant true normal population data.

Clause 10. The method of clause 9, wherein the relevant true normal sample is provided by a woman of African Ancestry.

Clause 11. The method of clause 9, wherein the relevant true normal sample is provided by a woman of European Ancestry.

Clause 12. The method of any one of clause 5 to 11 further comprising screening said biopsy sample for duffy-null alleles.

Clause 13. The method of any one of clause 5 to 12 wherein the biopsy sample is recovered from human female breast tissue.

Clause 14. The method of clause 12 or 13 wherein the biopsy sample is categorized as being either European Ancestry or African Ancestry.

Clause 15. The method of clause 5 wherein the number of ZEB1+ cells and FOXA1+ cells are compared to a relevant true normal sample.

Clause 16. The method of clause 15 further comprising, comparing the number of ZEB1+ cells and FOXA1+ cells to a matching normal adjacent tissues (NATs).

Clause 17. An improved method of treating cancer in a patient of European Ancestry or African Ancestry, said method comprising the steps of

obtaining a tissue biopsy sample identified as being cancer by radiological techniques and categorized as being from a patient of European Ancestry or African Ancestry;

analyzing said sample for increased number of ZEB1+ cells and reduced levels of FOXA1+ cells compared to a relevant true normal sample or relevant true normal population data;

identifying patients of European Ancestry and having specific increase in the number of ZEB1+ cells;

identifying patients of African Ancestry and having reduced number of FOXA1+ cells; and

treating said identified patients of European Ancestry and patients of African Ancestry using anti-cancer therapies.

Clause 18. The method of clause 17 wherein the patient is a woman and the biopsy sample is recovered from breast tissue.

Clause 19. An improved method of analyzing a pre-cancer or cancerous biopsy in a patient who is duffy-null, said method comprising the steps of

analyzing a tissue biopsy sample obtained from said patient and identified as cancerous or pre-cancerous using radiology techniques; and

quantifying the levels of c-MET in the patient sample.

Clause 20. The method of clause 19, further comprising quantifying the levels of ZEB1+ cells in said biopsy sample.

Clause 21. The method of clause 19, further comprising comparing the levels of phospho-c-MET to a relevant true normal sample or a relevant true normal population data.

Clause 22. A kit for detecting and quantitating the relative number of ZEB1+ cells and FOXA1+ cells in a tissue sample recovered from a patient, said kit comprising

a reagent for the detection of ZEB1+ cells; and

a reagent for the detection of FOXA1+ cells.

Clause 23. The kit of clause 22 wherein the reagent is an antibody specific for the respective protein.

Clause 24. The kit clause of clause 22 or 23 further comprising a reagent for detecting GATA3.

Clause 25. The kit clause of any one of clauses 22-24 further comprising a reagent for detecting phospho-c-MET.

Clause 26. The kit clause of 24 and 25, wherein the reagent is an antibody specific for the respective protein.

Clause 27. An improved method of analyzing a pre-cancerous or cancerous biopsy sample from a patient of European Ancestry, said method comprising the steps of

analyzing the biopsy sample identified as pre-cancerous or cancerous using radiology techniques;

quantifying the levels of GATA3 + cells; and

comparing the level of GATA3+ cells to a relevant true normal breast tissue sample or a relevant true normal population data.

Clause 28. The method of clause 27, wherein lowered levels of GATA3+ cells is an indication of pre-cancer or cancer.

Clause 29. The method of clause 27 further comprising quantifying the levels of ZEB1+ cells and comparing the ZEB1+ cell level to a relevant true normal breast tissue sample.

Clause 30. The method of clause 29, wherein an increased level of ZEB1+ cells is an indication of pre-cancer or cancerous breast tissue.

Clause 31. The method of any of the preceding clauses wherein the biopsy comprises NATs.

Clause 32. The method of any of the preceding clauses wherein the biopsy comprises a tumor.

EXAMPLE 1

Genetic ancestry influences evolutionary pathways of cancers. For a discussion see Cancer Cell 34:529), the contents of which are incorporated herein by reference. However, whether ancestry influences cancer-induced field effects is unknown. Here the ancestry-mapped relevant true normal breast tissues were utilized as controls to identify cancer-induced field defects in NATs in AA and EA women. As disclosed herein, applicants have taken advantage of genetic ancestry mapped relevant true normal breast tissues to identify differences between relevant true normal and NATs, which can potentially develop into earlier markers of breast cancer initiation. A tissue microarray (TMA) comprising breast tissues from clinically normal breast tissue, NATs, and tumors was analyzed for markers that are expressed in cells with stem and mature luminal cell properties. Applicants also examined the TMA for CD8+ T cells, CD68+ macrophages, PD1+ immune cells and PDL1+ epithelial cells to determine whether immune cell composition of tumors and NATs in AA women differ from that of EA women with the goal of identifying cancer-induced field defects in normal tissue adjacent to breast tumors (NATs) in women of African Ancestry (AA) and European ancestry (EA).

This disclosure will help in early detection of breast cancer. It will also help in resolving ambiguity in breast cancer diagnosis by radiology techniques such as mammography. Biopsy of sites with ambiguity and staining for FOXA1 and ZEB1+ cells provide additional evidence for the presence of cancer. We propose that cancer-induced field defects cause increase in ZEB1+ cells in women of European Ancestry and decrease in FOXA1+ cells in women of African Ancestry.

Materials and Methods:

Generation of a tissue microarray (TMA). Breast core biopsies from healthy women donated to Komen Tissue Bank (KTB) and surgical material left over after

pathologic assessments as part of a treatment protocol were obtained after informed written consent from the subjects. Applicant created a tissue microarray (TMA) comprising: 1) healthy breast tissue from the KTB (KTB-normal), 2) matched NATs, and 3) tumor tissue. About 50 samples from AA women and about 50 samples from EA women were provided for a total of about 300 samples. KTB-normal tissues were age and ancestry mapping matched to NATs/Tumors. BMI of AA women who donated tissues to KTB was 32.3±9, and 28.3±8.5 in case of EA women. Each tumor sample and matching NAT were spotted in duplicate.

EXAMPLE 2

Immunohistochemistry and statistical analyses. The TMA was analyzed for ZEB1, methione sulfoxide reductase B3 ("MSRB3"), estrogen receptor alpha (ERα), GATA3, and FOXA1 expression. All immunohistochemistry (IHC) was done in a CLIA-certified histopathology lab and evaluated by three pathologists in a blinded manner and quantitative measurements were done using automated Aperio Imaging system and analysis was done using an FDA approved algorithm. Positivity and H-scores were scored and statistically analyzed as described previously (Breast Cancer Res Treat (2015) 152(3):519-31). With respect to PD1 and PDL1, a tumor proportion score (TPS) was created. The TPS describes the ratio of positive viable tumor cells against all viable tumor cells. The PDL1 TPS followed the prescribed FDA reading: a negative score having <1% positive staining on the cell membrane, a positive score having 1%-49% cells partially or completely expressing PDL1 on the cell membrane, and a strong positive having ≥ 50% tumor cells partially or completely expressing PDL1 on the cell membrane at a stronger intensity (Modern Pathol 29:1165-72; NSCLC-Keytruda PDL1 IHC 22C3 pharmDX Interpretation manual). Data were analyzed in three different ways: 1) Expression differences between AA and EA KTB-normal; 2) Expression differences between KTB-normal and NATs; and 3) Expression differences between NATs and tumors. Non-parametric Wilcoxon rank-sum tests were used for unpaired analyses, as positivity and H scores were not normally distributed, whereas non-parametric Wilcoxon signed-rank tests were used for paired analyses. The following antibodies were used: CD8 (Dako IR623), CD68 KP1 (Dako IR609), ER clone:EP1 (Dako IR 084), FOXA1 (Santa Cruz sc-6553), GATA3 (Santa Cruz sc-268), MSRB3 (HPA014432, rabbit polyclonal,

Sigma), PD1 (Cell Marque 315M-98), PDL1 (Keytruda) (clone 22c3, Dako IHC 22c3) and ZEB1 (3G6, cat no 14-9741-82, eBioscience).

EXAMPLE 3

ZEB1+ cells are enriched in the normal breast tissue of AA compared to EA women. In the mouse mammary gland, PROCR+/EpCAM- cells are suggested to function as multi-potent stem cells. In a previous study focused on evaluating ethnicity-dependent differences in the normal breast tissue, specific enrichment of PROCR+/EpCAM- in cultured normal breast epithelial cells from biopsies of healthy AA women compared to EA women was observed (Nakshatri et al., Scientific Reports (2015) 5:13526). These cells are enriched for the expression of stemness-related transcription factor ZEB1 and have enhanced Wnt pathway activity compared to PROCR±/EpCAM+ cells. ZEB1 has recently been demonstrated to limit onco-suppressive p53-driven DNA damage response in stem cells and thus increase the stem cells' intrinsic susceptibility to malignant transformation. ZEB1+ cells co-express the methionine sulfoxidase reductase (MSRB3), which protects against DNA damage. These observations raised the possibility that PROCR+/ZEB1+ cells are naturally present at higher levels in the normal breast tissue of AA women and failure to consider natural variation in gene expression pattern, influenced at least partially by genetic ancestry, could have an impact on identifying cancer-induced field effect on adjacent normal breast tissue. Measuring PROCR itself in the breast tissue is complicated because there are four haplotypes of PROCR due to SNPs and only one among them is a cell surface protein. Since ZEB1 expression is enriched in PROCR+/EpCAM- cells, applicant used ZEB1 as a surrogate marker for PROCR+/EpCAM- cells in un-manipulated breast tissues.

Representative IHC staining patterns of ZEB1 in KTB-normal, NATs, and tumors from AA and EA women were conducted and statistical analyses are presented in Figure 1B-1D and in Table 1. Descriptive statistics of ethnicity, age, menstrual status, pregnancy, breastfeeding, hormone replacement therapy and family history of breast cancer for the KTB-normal cohort is shown in Table 2. Highly discriminative 41-ancestry marker profiles of KTB-normal showed >75% African ancestry markers in samples from African American women and >80% European ancestry markers in Caucasian women for the generation of a relevant true normal sample or relevant true normal population data (Figure 6). Characteristics of breast cancer in the tumor cohort

are shown in Table 3. ZEB1 expressing cells are localized outside the ductal structures of the normal breast tissue and in the stromal part of the tumors. KTB-normal breast tissue of AA women contained significantly higher levels of ZEB1-positive cells compared to KTB-normal breast tissue of EA women (Figure 1B). NATs of AA women showed a modest increase in ZEB1+ cells compared to KTB-normal (Figure 1C and 1D). The situation was completely different in EA women; both NATs and tumors contained significantly higher levels of ZEB1+ cells compared to KTB-normal (Figure 1C and D). NAT to tumor differences were noted only in EA women where an increase in ZEB1+ cells was noted predominantly in ERα+ tumors (Table 4). Thus, ZEB1+ cells are intrinsically higher in the normal breast tissue of AA women, whereas remarkably elevated ZEB1+ cells in the breast tissue of EA women were observed only in the context of breast cancer.

EXAMPLE 4

Increase in ZEB1+ cells in KTB-normal of AA women compared to EA women is less likely related to BMI differences, as when analysis was done by subdividing women above and below BMI of 30, irrespective of genetic ancestry, ZEB1 H-score but not positivity showed a marginal relationship ($p=0.04$) to BMI above and below 30 (Table 5).

EXAMPLE 5

MSRB3 has recently been shown to be one of the downstream transcriptional targets of ZEB1 and it cooperates with ZEB1 during transformation of stem-like cells. To correlate ZEB1 expression with its activity, we measured the levels of MSRB3 using the same antibody used in the above study. We could measure positivity but not H-score because of low-level expression. The expression pattern was similar to that of ZEB1, as cells surrounding the ducts showed expression (Figure 7A. However, KTB-normal tissues of AA and EA women expressed similar levels of MSRB3 (Table 1 and Figure 7B, which could be due to regulation by other transcription factors or to the low expression levels, making data interpretation difficult. Furthermore, except for a modest change in expression in NATs compared to KTB normal tissues, no other differences were noted (Figures 7C and 7D).

EXAMPLE 6

FOXA1 expression is lower in NATs of only AA women: FOXA1 serves as a pioneer factor that controls chromatin access of various nuclear receptors including ERα and controls luminal gene expression. FOXA1 along with another pioneer factor GATA3 and ERα form a lineage restricted hormone-responsive signaling network in the normal breast tissue. While higher expression of FOXA1 in the primary tumor is associated with better outcome, its overexpression in metastatic and/or anti-estrogen resistant tumors is associated with rewiring of ERα signaling and poor outcome. In addition, it is suggested that FOXA1 gene is preferentially methylated in tumors of AA women. Because of its relative importance in breast cancer, we assessed our TMA for FOXA1 expression. Staining patterns of FOXA1 were evaluated and numerical values are presented in Table 1. While FOXA1 levels in KTB-normal of AA women was modestly higher than EA women (Figure 2B), NATs of AA women had lower FOXA1 compared to KTB-normal. Thus, tumors through their field effect may decrease FOXA1 in the NATs of AA women.

EXAMPLE 7

GATA3 levels are higher in KTB-normal of EA compared to AA women: We also examined expression levels of GATA3 to determine whether hormonal signaling networks show genetic ancestry-dependent variation. Consistent with this possibility, GATA3 H-score and positivity were higher in KTB-normal tissues of EA women compared to those of AA women (Table 1 and Figure 8A. Furthermore, GATA3 is a likely candidate for cancer-induced field effects in EA women as its levels were significantly lower in NATs of EA but not AA women compared to their KTB-normal counterparts (Figure 8B).

EXAMPLE 8

ERα+ cells remain stable: ERα-positive cells in the normal breast tissue are considered to be highly differentiated non-proliferative cells and control proliferation of ERα-negative cells through paracrine mechanisms. Representative ERα staining pattern and statistical analyses are presented in Figure 3B-3D and Table 1. There were no differences between AA and EA women in ERα-positivity in KTB-normal and NATs. Neither KTB normal tissues nor NATs showed genetic ancestry-dependent differences in ERα levels. The results are not only relevant, but also reassure that the TMA detects only specific changes.

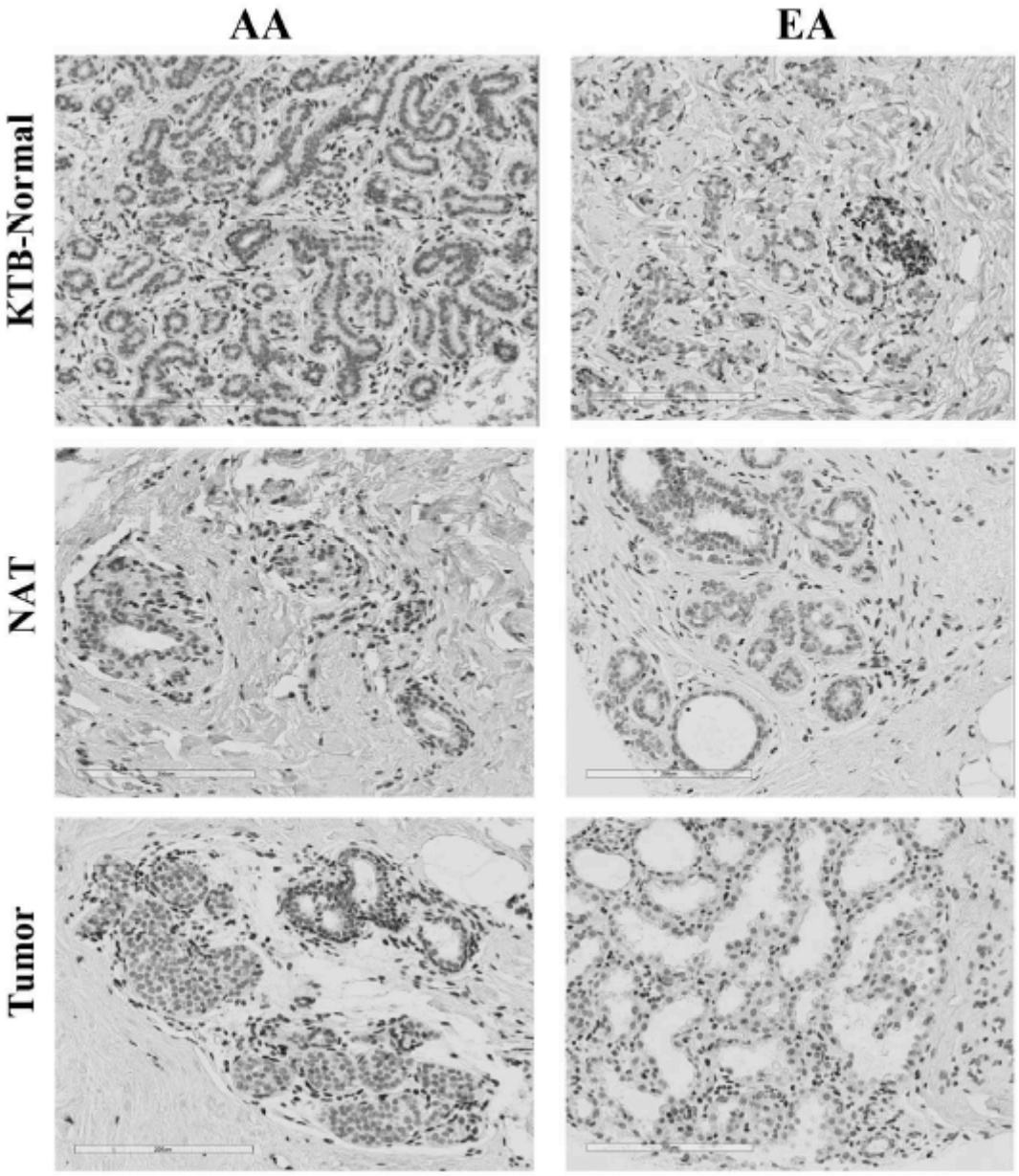


FIG. 1A

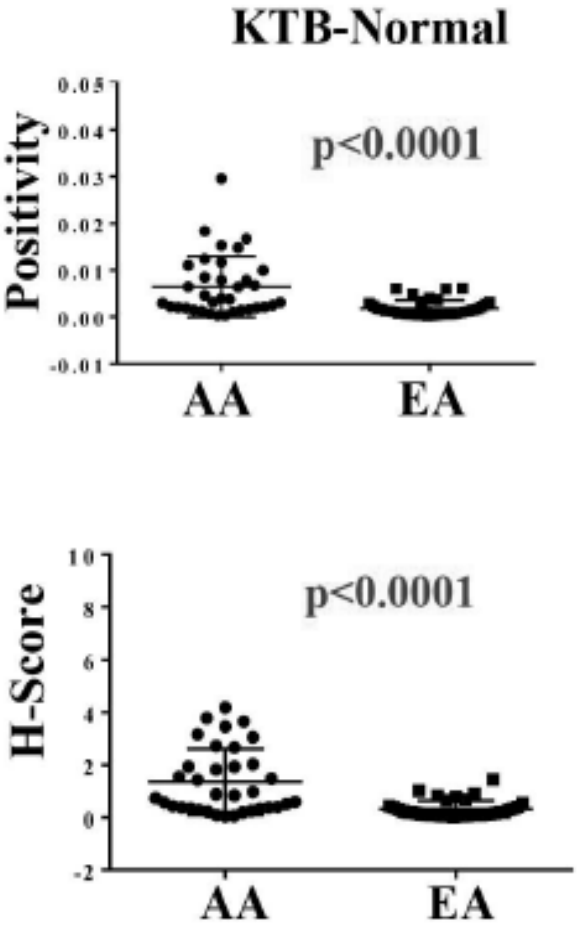

KTB-Normal
Positivity
0.05
0.04
0.03
0.02
0.01
0.00
-0.01
p<0.0001
AA
EA
H-Score
10
8
6
4
2
0
-2
p<0.0001
AA
EA

FIG. 1B

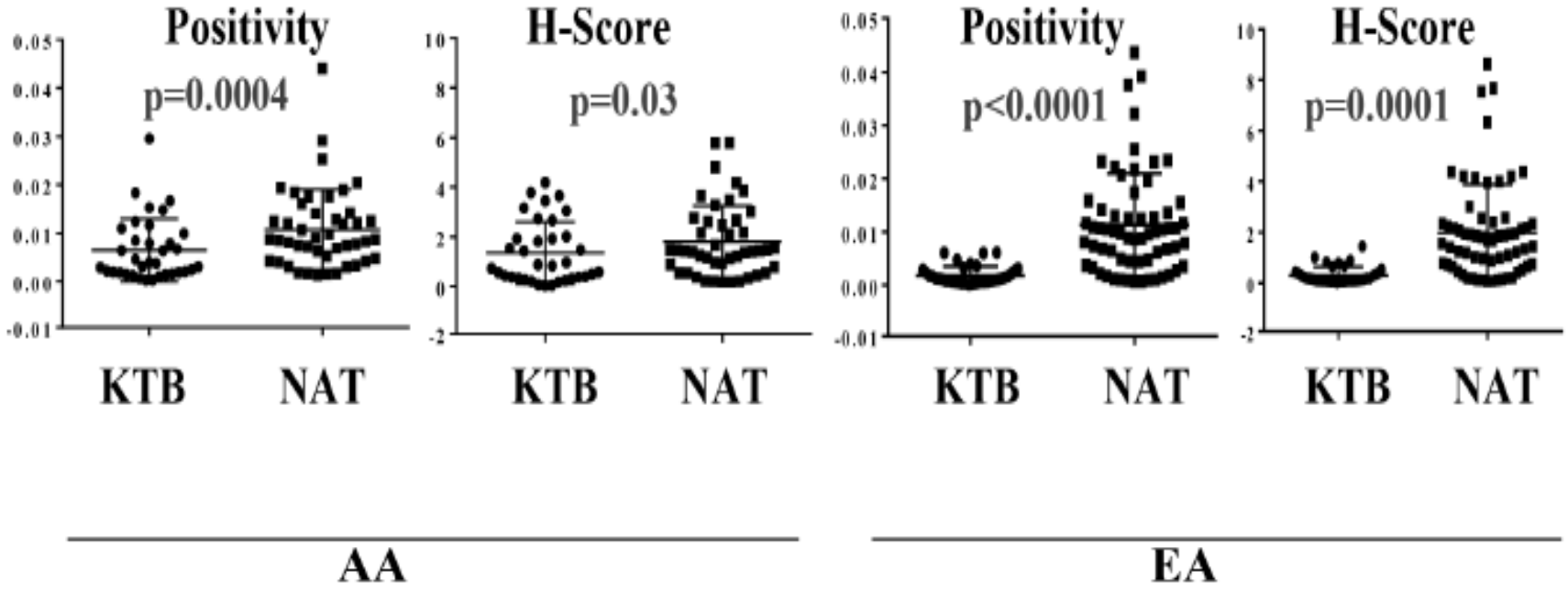

Positivity
p=0.0004
H-Score
p=0.03
Positivity
p<0.0001
H-Score
p=0.0001
KTB
NAT
KTB
NAT
KTB
NAT
KTB
NAT
AA
EA

FIG. 1C

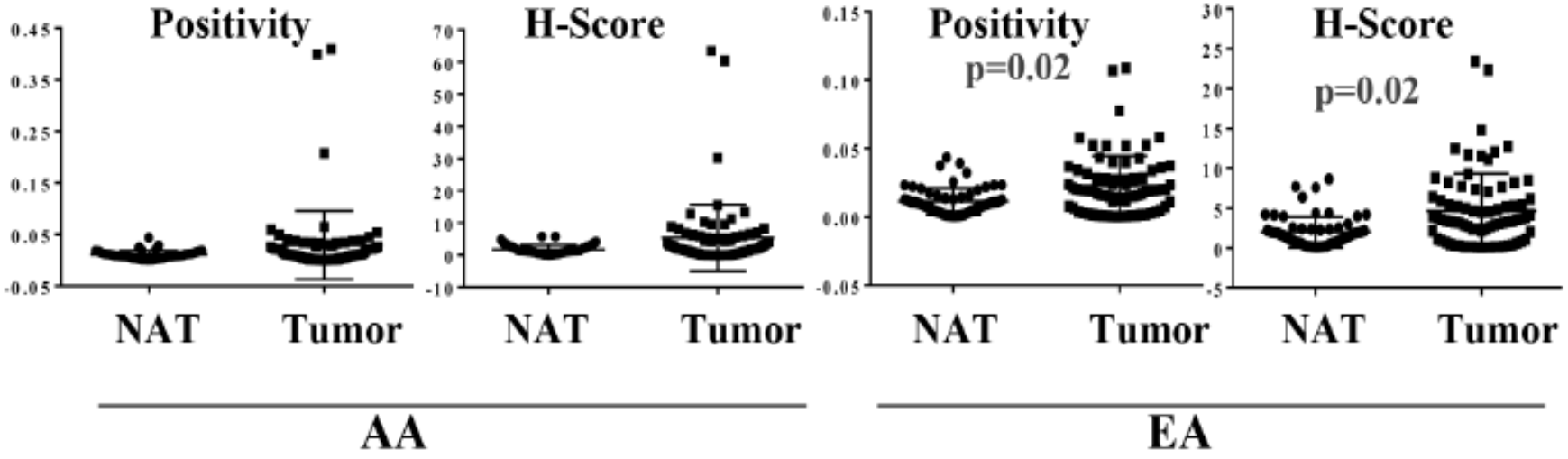

Positivity
H-Score
Positivity
p=0.02
H-Score
p=0.02
0.45
0.35
0.25
0.15
0.05
-0.05
70
60
50
40
30
20
10
0
-10
0.15
0.10
0.05
0.00
-0.05
30
25
20
15
10
5
0
-5
NAT
Tumor
NAT
Tumor
NAT
Tumor
NAT
Tumor
AA
EA

FIG. 1D

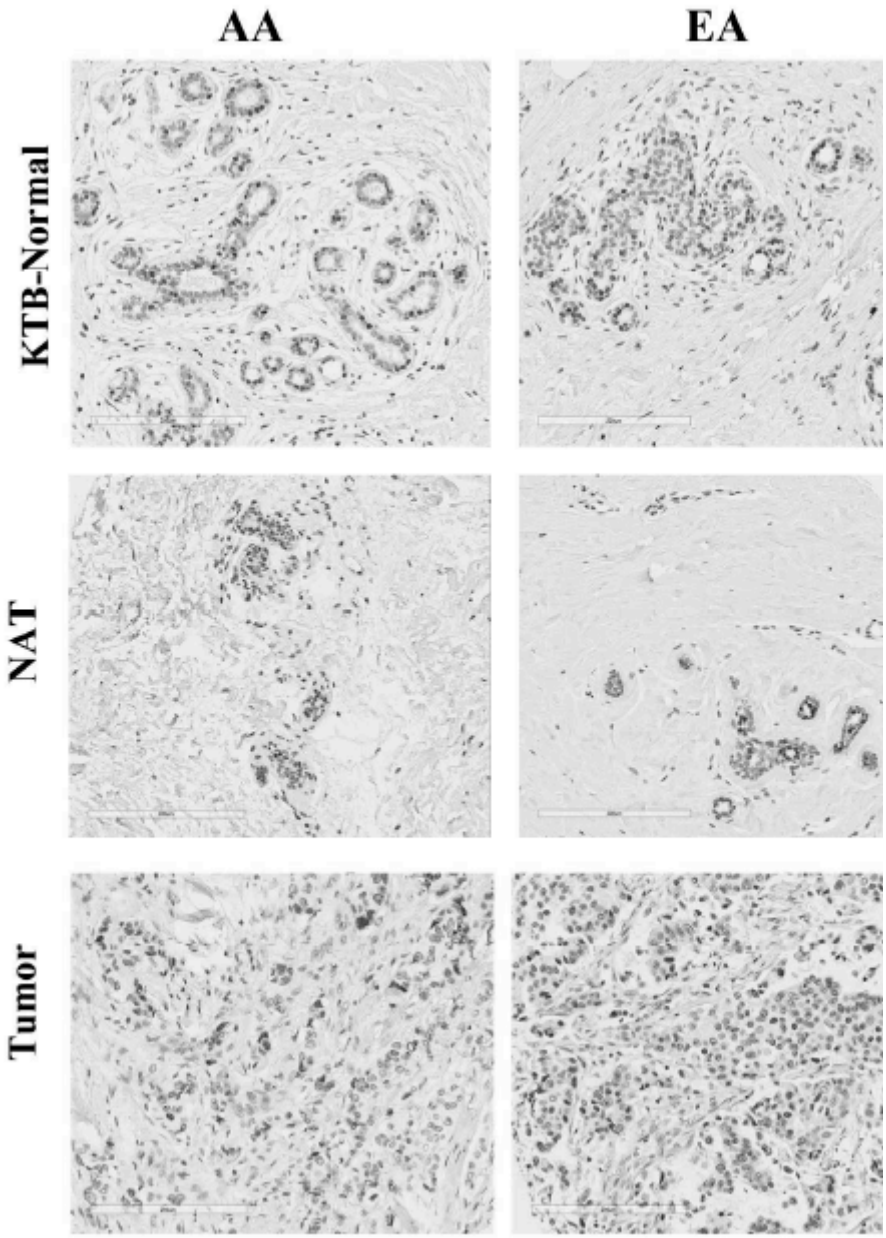


FIG. 2A

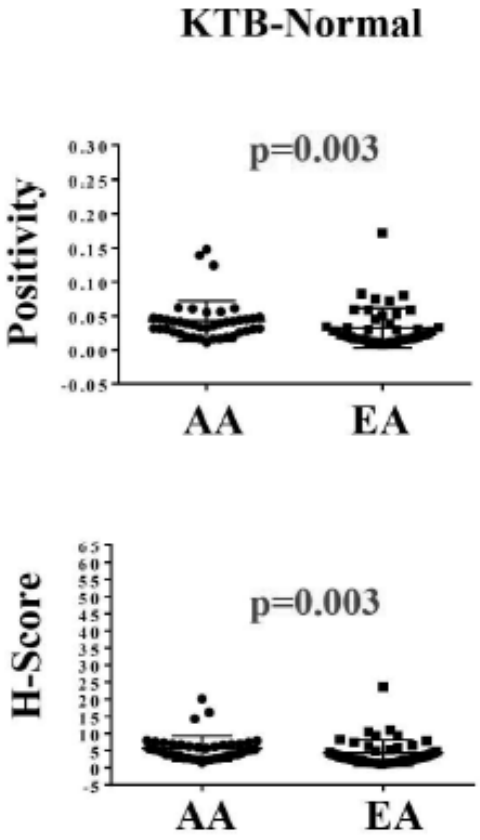

KTB-Normal
p=0.003
Positivity
0.30
0.25
0.20
0.15
0.10
0.05
0.00
-0.05
AA
EA
p=0.003
H-Score
65
60
55
50
45
40
35
30
25
20
15
10
5
0
-5
AA
EA

FIG. 2B

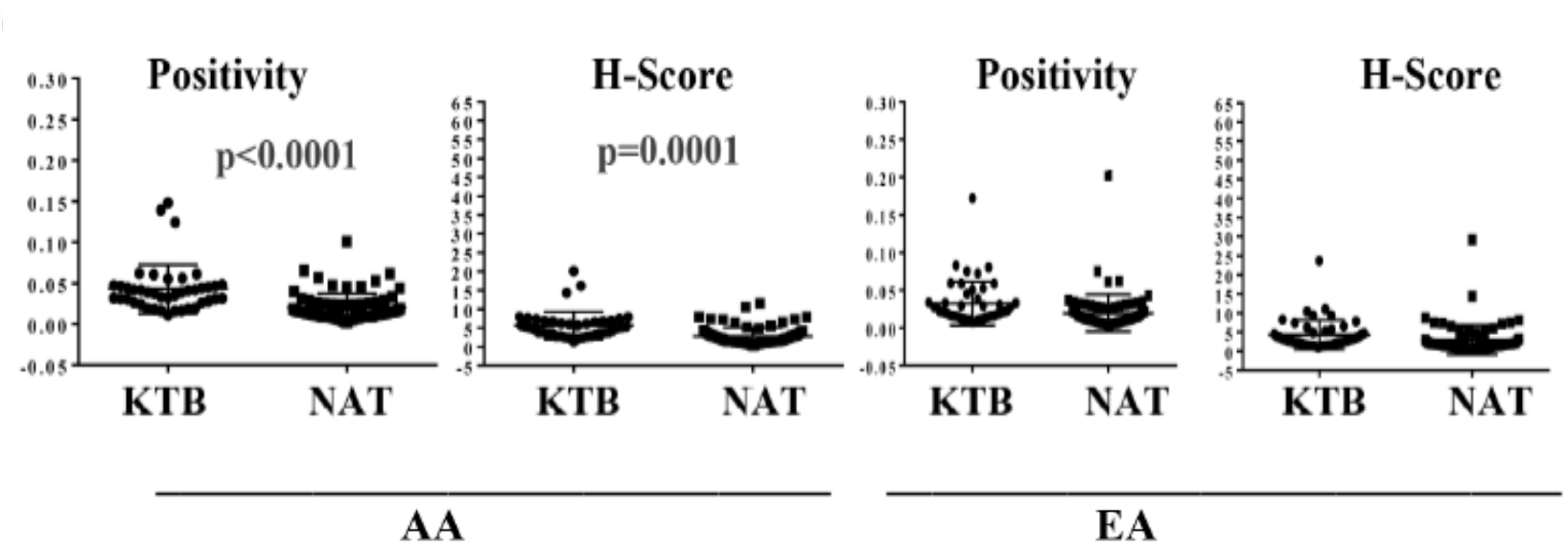

Positivity
p<0.0001
H-Score
p=0.0001
Positivity
H-Score
KTB
NAT
KTB
NAT
KTB
NAT
KTB
NAT
AA
EA

FIG. 2C

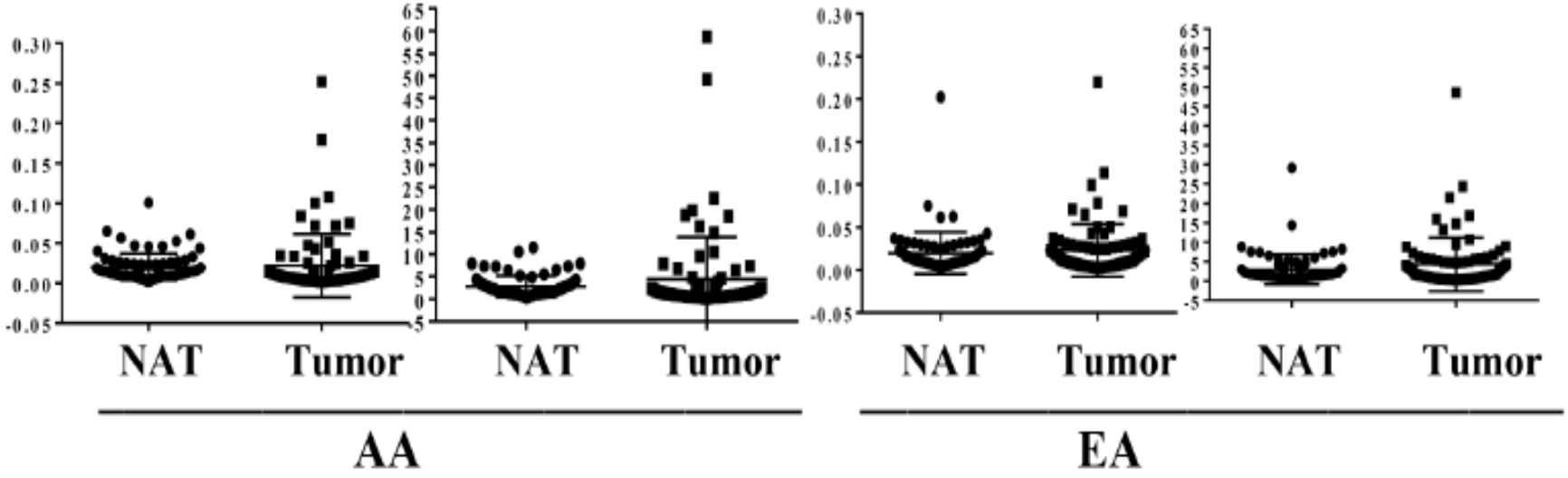

0.30
0.25
0.20
0.15
0.10
0.05
0.00
-0.05
65
60
55
50
45
40
35
30
25
20
15
10
5
0
-5
NAT
Tumor
NAT
Tumor
AA
0.30
0.25
0.20
0.15
0.10
0.05
0.00
-0.05
65
60
55
50
45
40
35
30
25
20
15
10
5
0
-5
NAT
Tumor
NAT
Tumor
EA

FIG. 2D

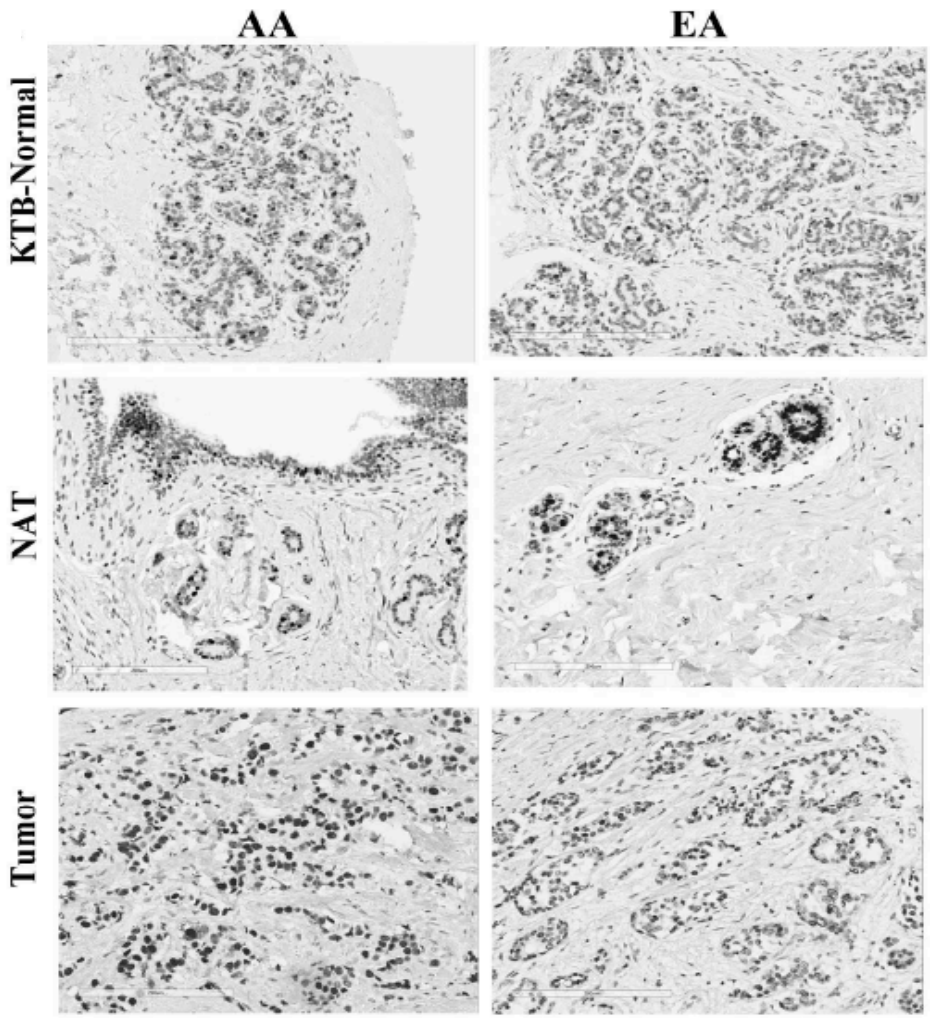


FIG. 3A

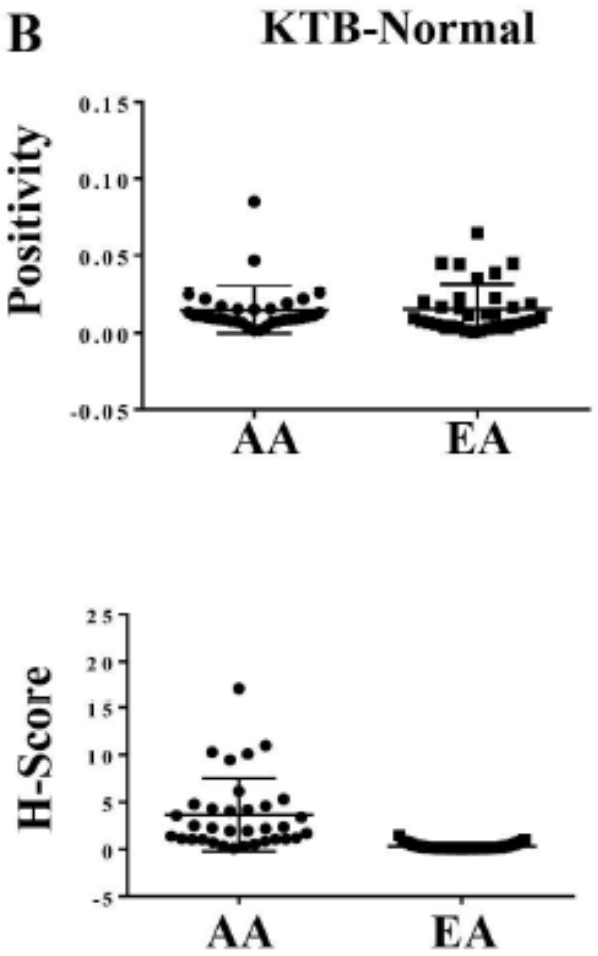

B
KTB-Normal
Positivity
0.15
0.10
0.05
0.00
-0.05
AA
EA
H-Score
25
20
15
10
5
0
-5
AA
EA

FIG. 3B

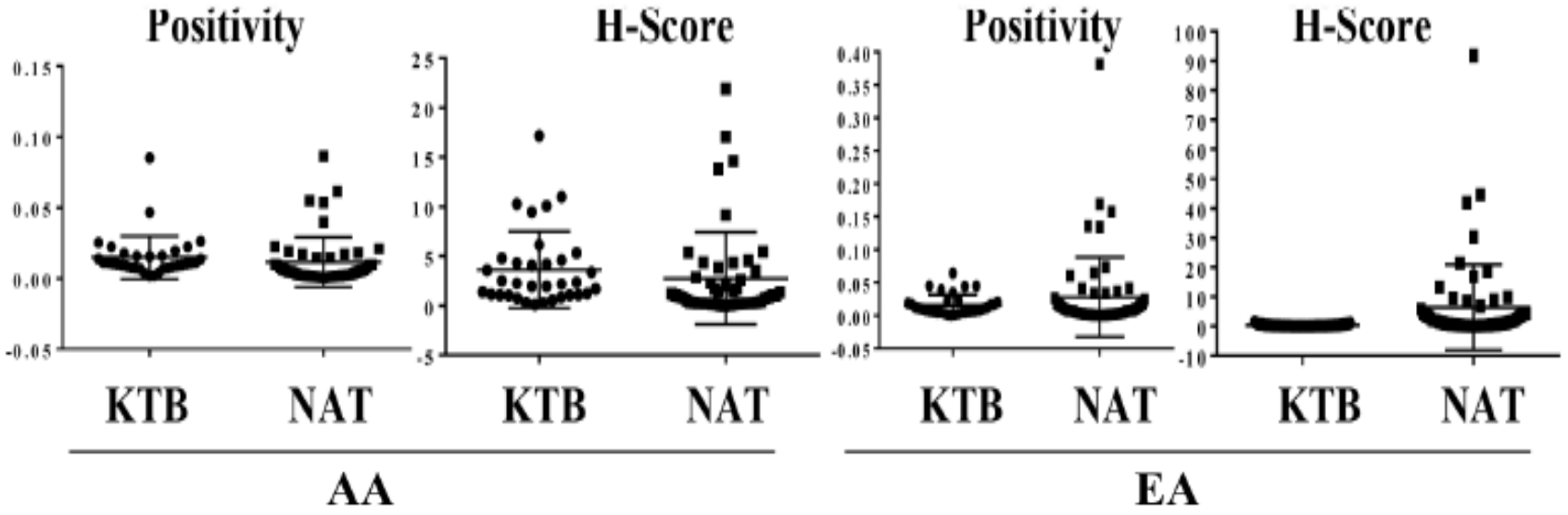

Positivity
H-Score
Positivity
H-Score
0.15
0.10
0.05
0.00
-0.05
25
20
15
10
5
0
-5
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
-0.05
100
90
80
70
60
50
40
30
20
10
0
-10
KTB
NAT
KTB
NAT
KTB
NAT
KTB
NAT
AA
EA

FIG. 3C

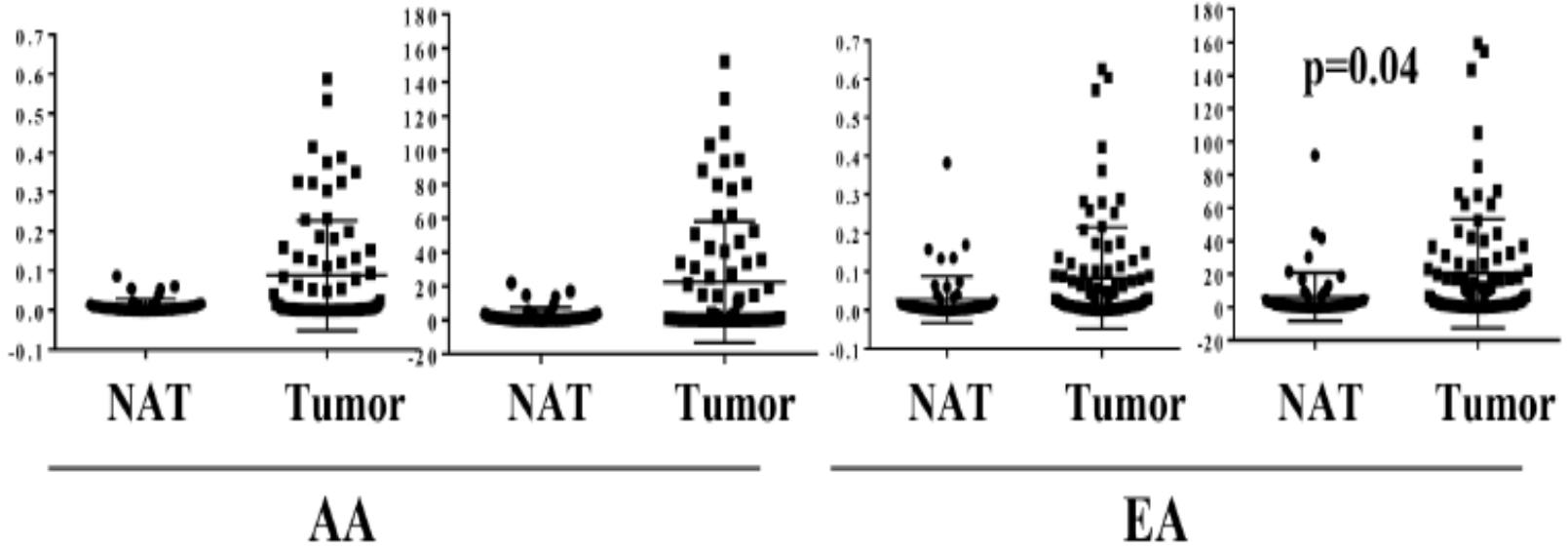

0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
-0.1
180
160
140
120
100
80
60
40
20
0
-20
p=0.04
NAT
Tumor
NAT
Tumor
NAT
Tumor
NAT
Tumor
AA
EA

FIG. 3D

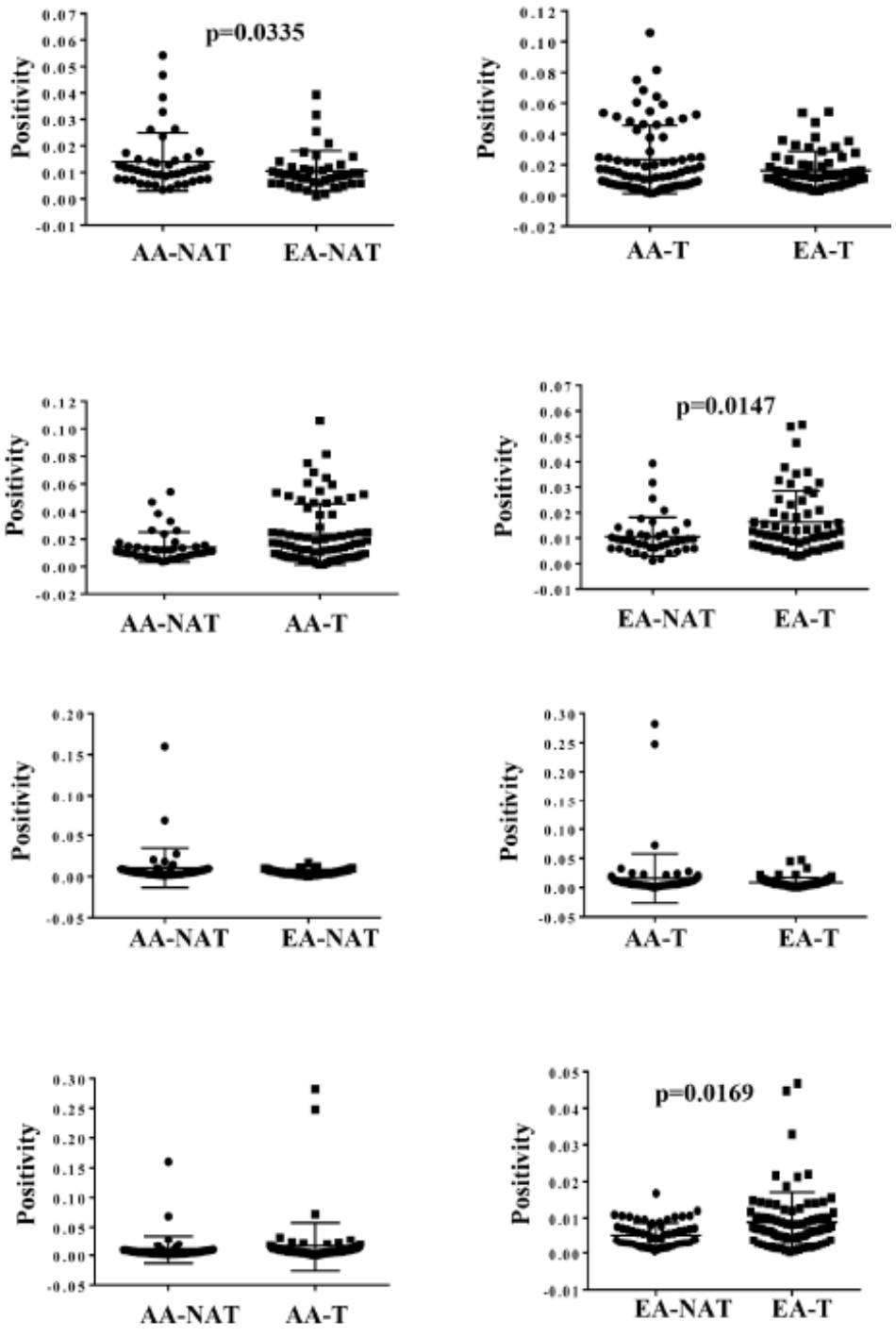

p=0.0335
Positivity
AA-NAT
EA-NAT
Positivity
AA-T
EA-T
Positivity
AA-NAT
AA-T
p=0.0147
Positivity
EA-NAT
EA-T
Positivity
AA-NAT
EA-NAT
Positivity
AA-T
EA-T
Positivity
AA-NAT
AA-T
p=0.0169
Positivity
EA-NAT
EA-T

FIG. 4

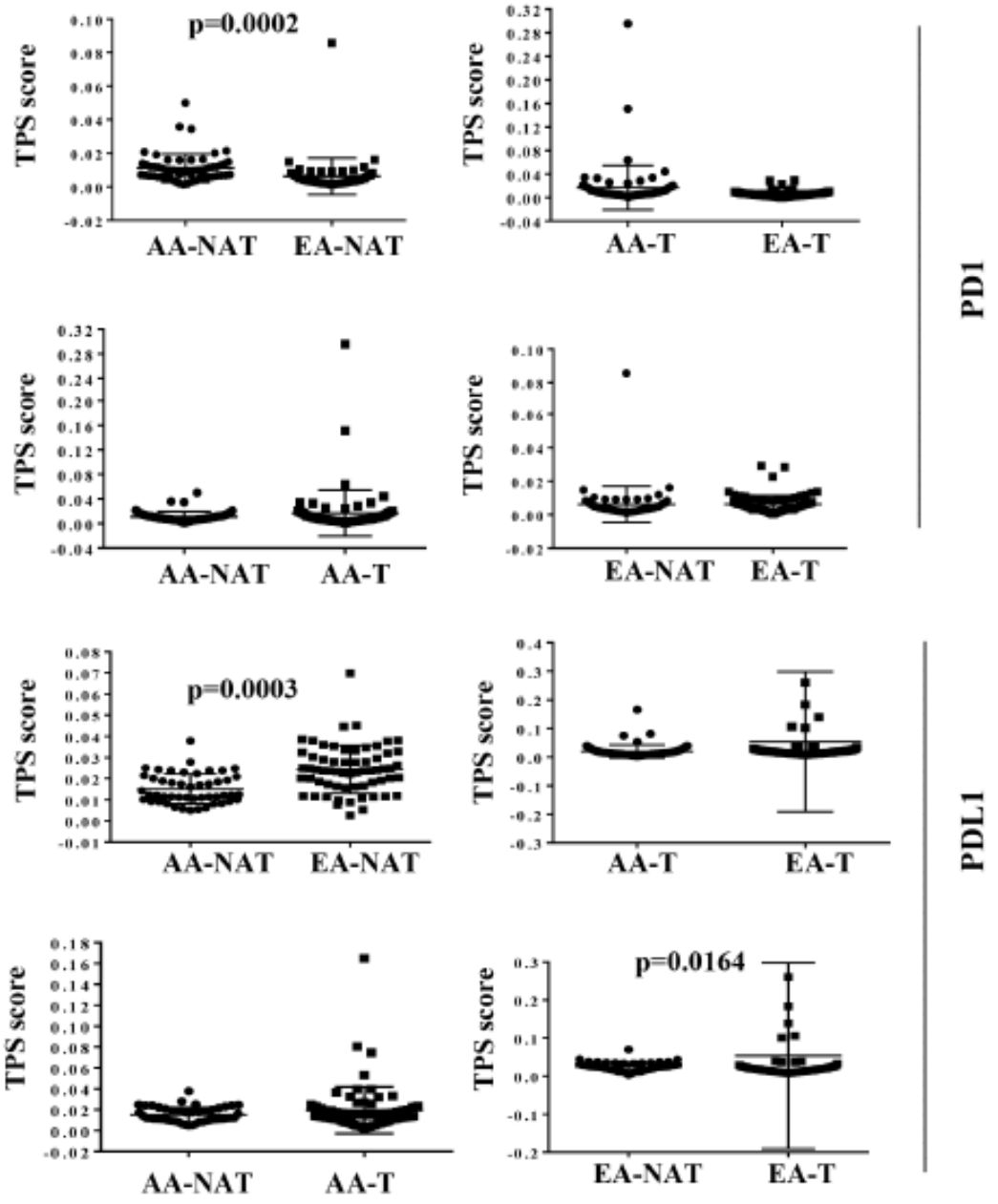

p=0.0002
TPS score
AA-NAT
EA-NAT
AA-T
EA-T
AA-NAT
AA-T
EA-NAT
EA-T
PD1
p=0.0003
AA-NAT
EA-NAT
AA-T
EA-T
p=0.0164
AA-NAT
AA-T
EA-NAT
EA-T
PDL1

FIG. 5

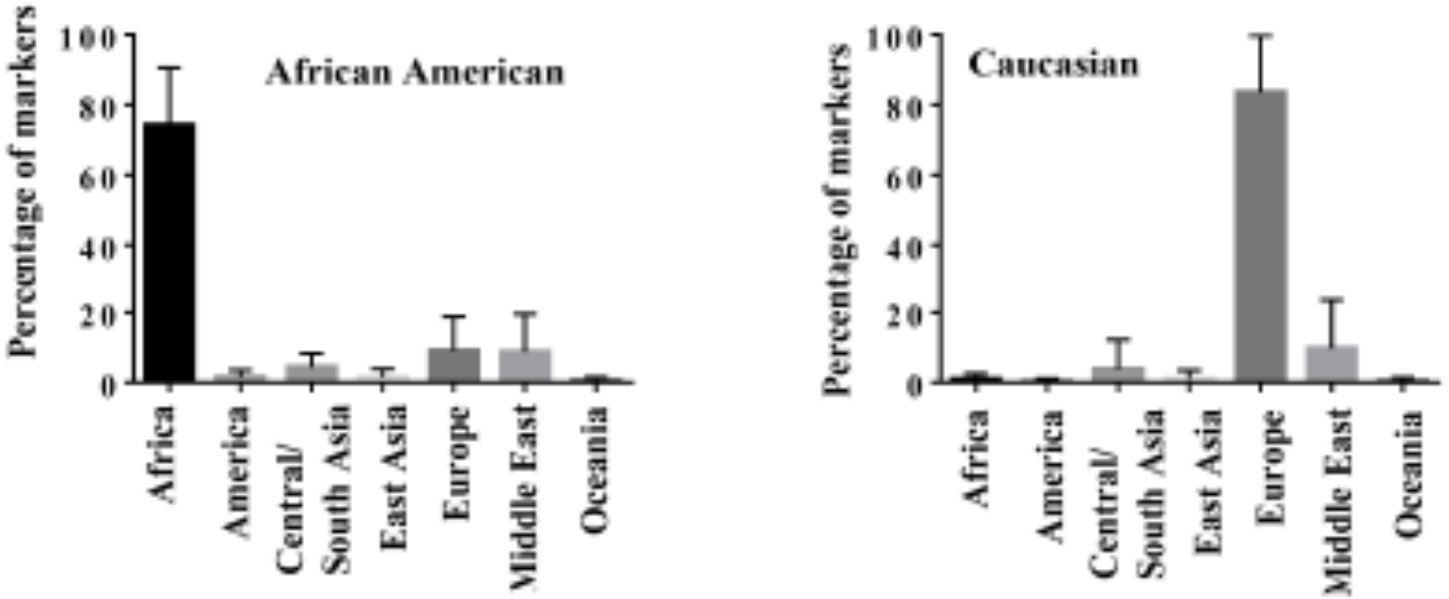

African American
Percentage of markers
100
80
60
40
20
0
Africa
America
Central/ South Asia
East Asia
Europe
Middle East
Oceania
Caucasian
Percentage of markers
100
80
60
40
20
0
Africa
America
Central/ South Asia
East Asia
Europe
Middle East
Oceania

FIG. 6

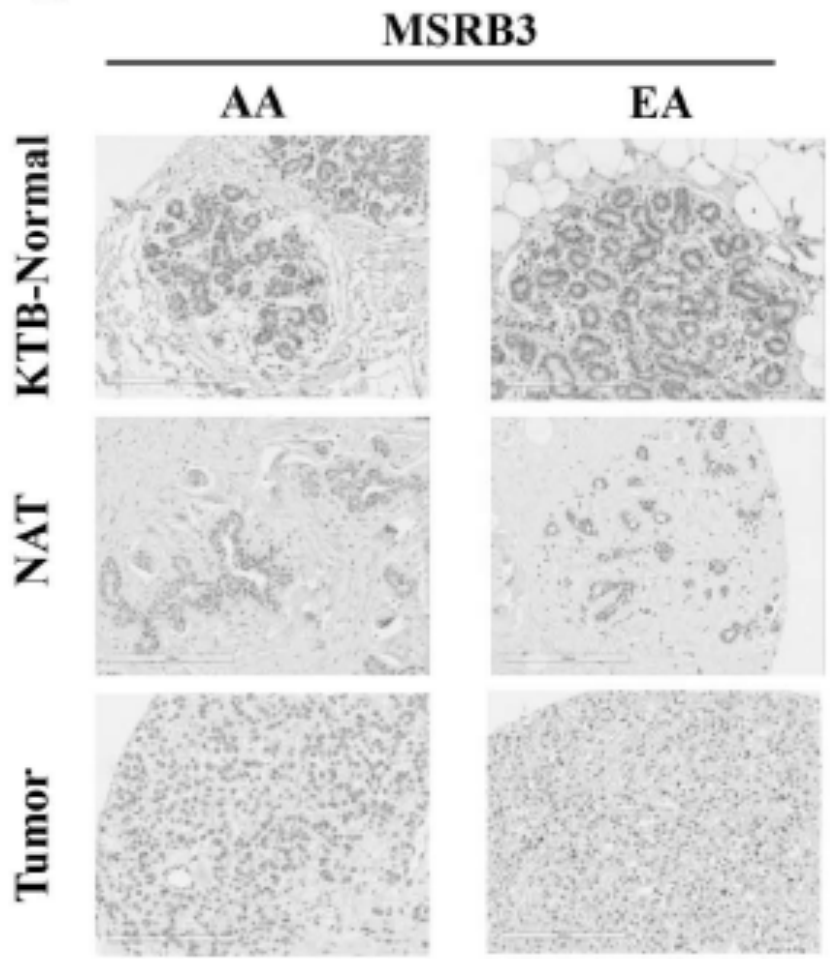


FIG. 7A

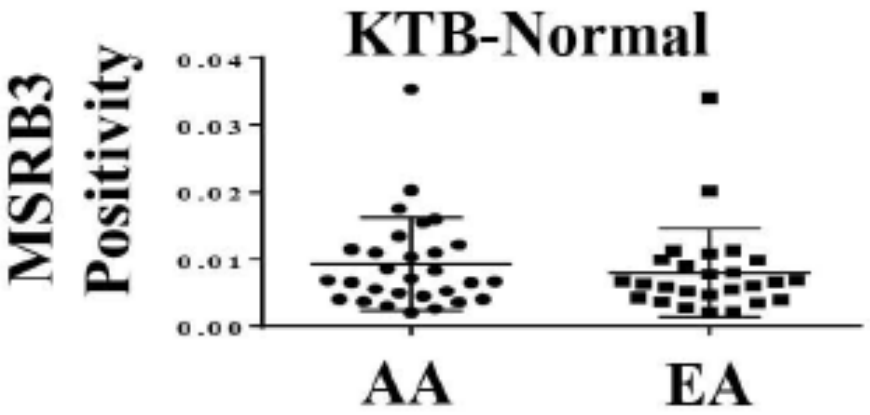


FIG. 7B

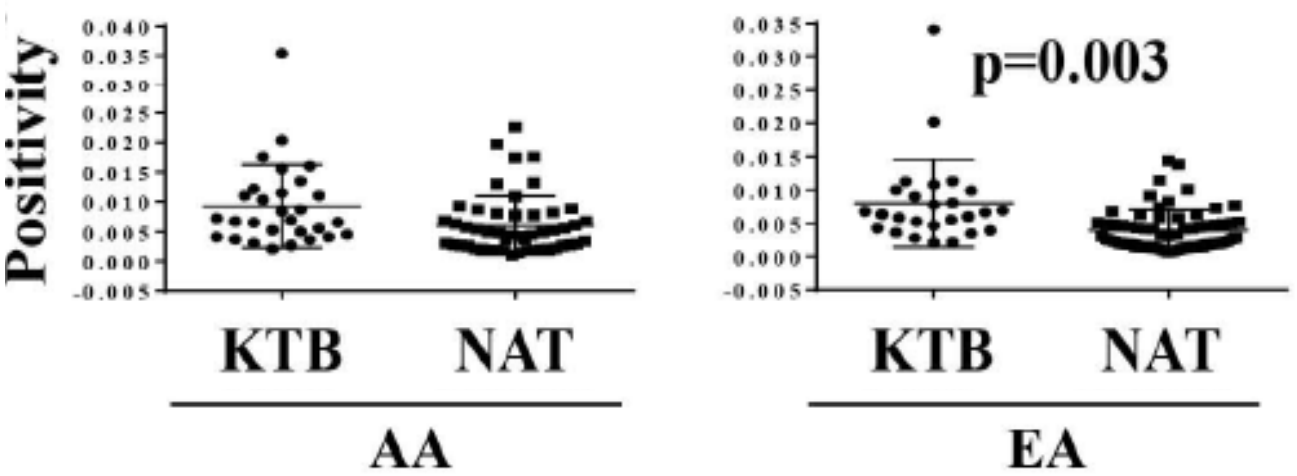

Positivity
0.040
0.035
0.030
0.025
0.020
0.015
0.010
0.005
0.000
-0.005
KTB
NAT
AA
0.035
0.030
0.025
0.020
0.015
0.010
0.005
0.000
-0.005
p=0.003
KTB
NAT
EA

FIG. 7C

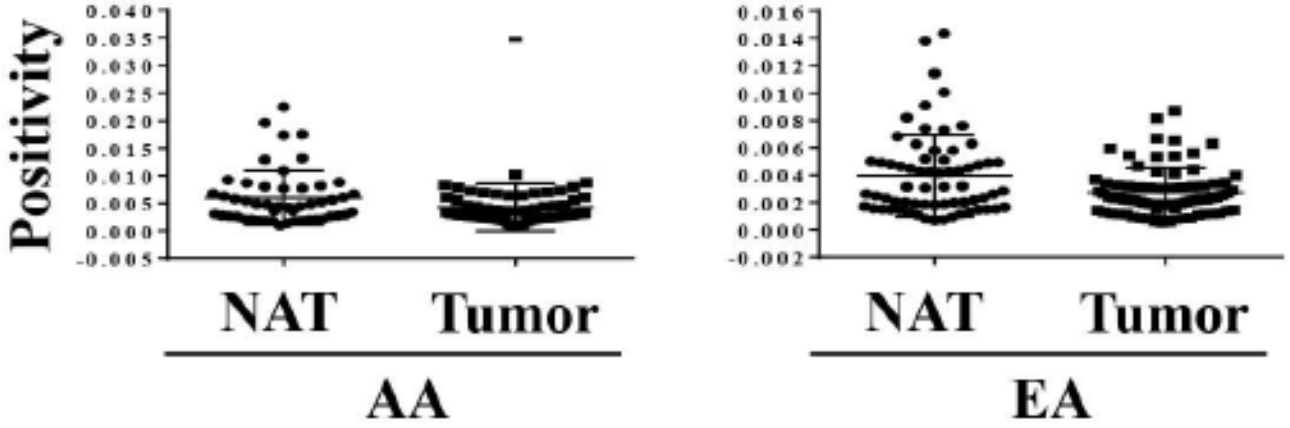

Positivity
0.040
0.035
0.030
0.025
0.020
0.015
0.010
0.005
0.000
-0.005
NAT
Tumor
AA
0.016
0.014
0.012
0.010
0.008
0.006
0.004
0.002
0.000
-0.002
NAT
Tumor
EA

FIG. 7D

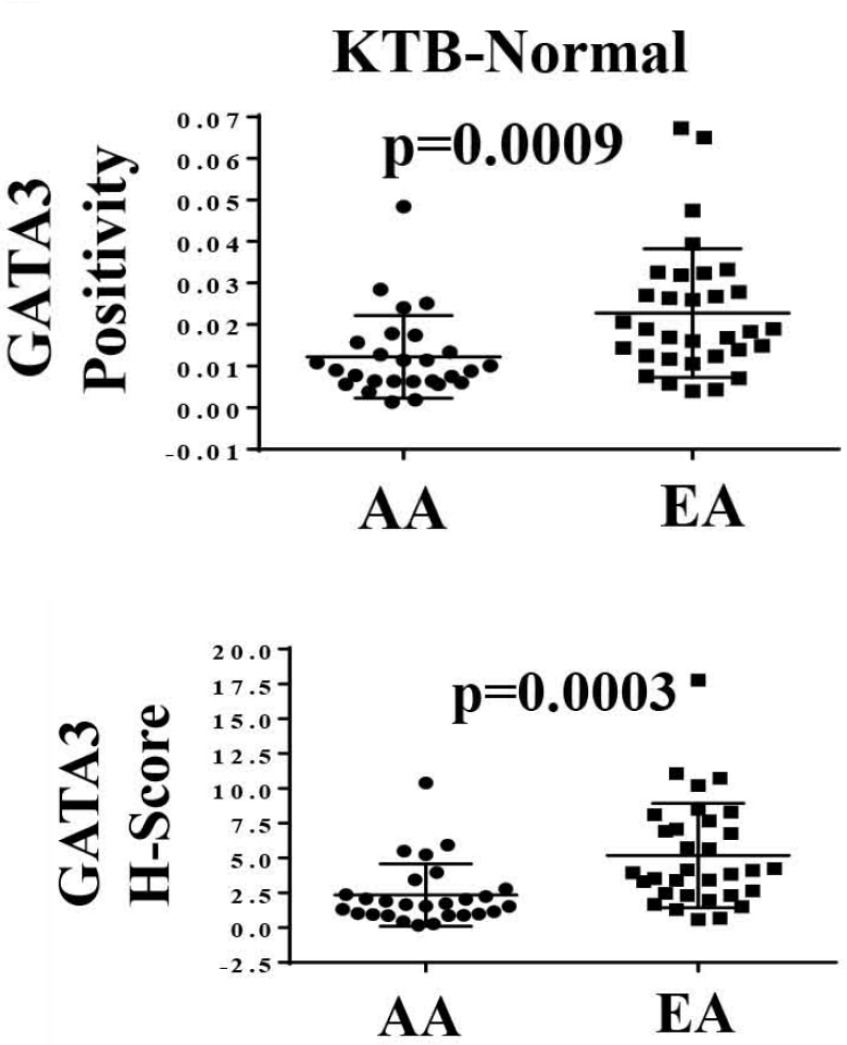

KTB-Normal
GATA3 Positivity
p=0.0009
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0.00
-0.01
AA
EA
GATA3 H-Score
p=0.0003
20.0
17.5
15.0
12.5
10.0
7.5
5.0
2.5
0.0
-2.5
AA
EA

FIG. 8A

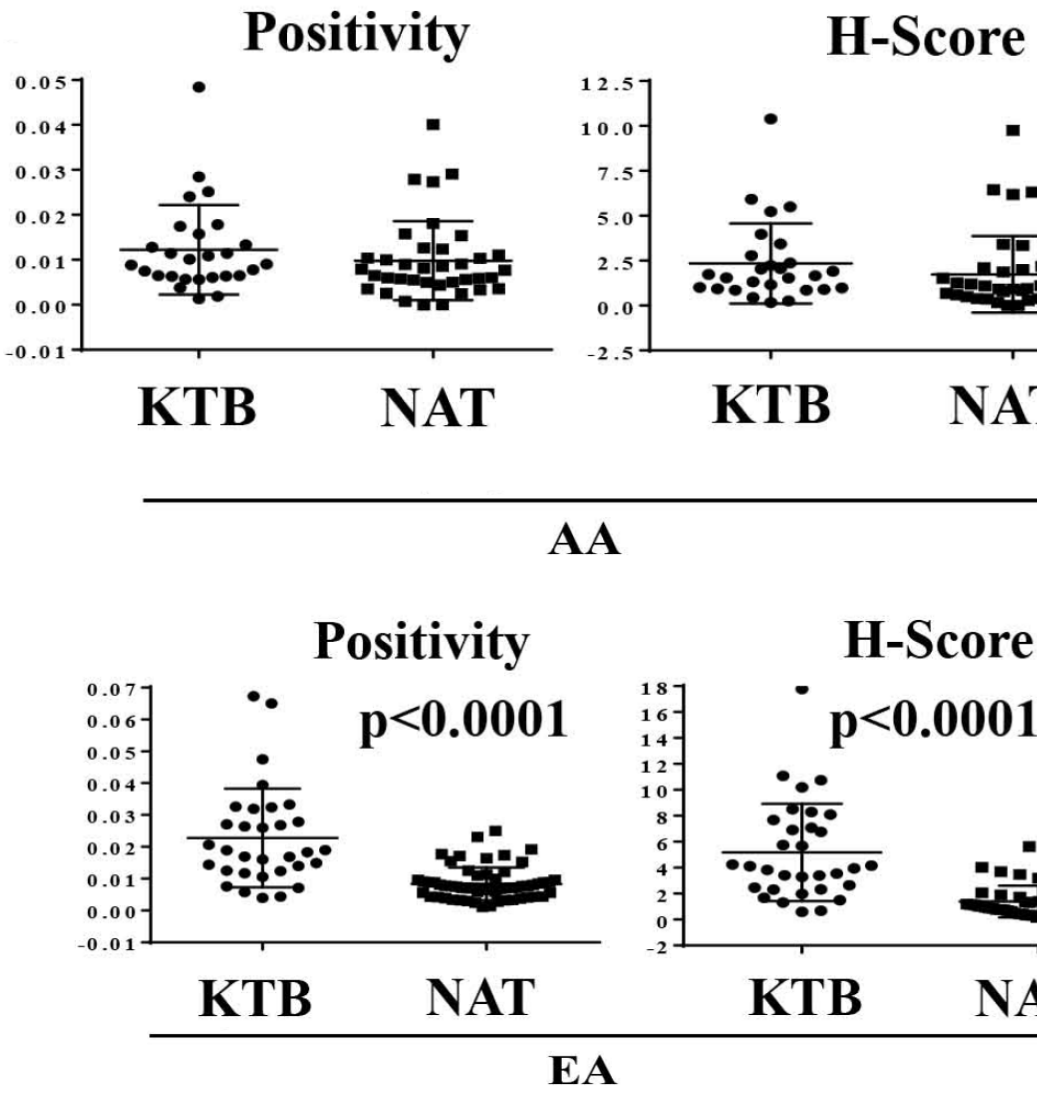

Positivity
H-Score
0.05
0.04
0.03
0.02
0.01
0.00
-0.01
12.5
10.0
7.5
5.0
2.5
0.0
-2.5
KTB
NAT
KTB
NA
AA
Positivity
H-Score
p<0.0001
p<0.0001
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0.00
-0.01
18
16
14
12
10
8
6
4
2
0
-2
KTB
NAT
KTB
NA
EA

FIG. 8B

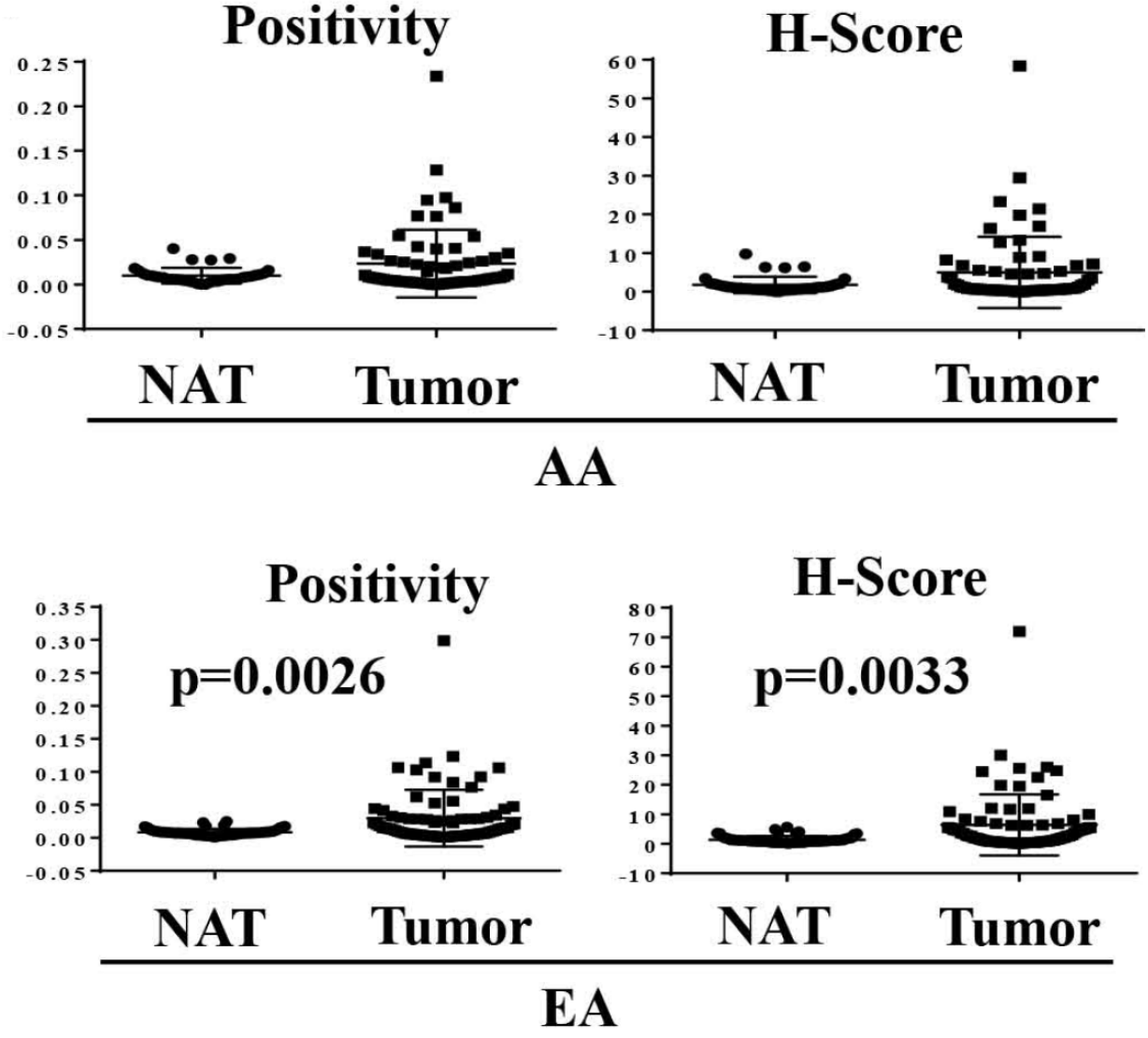

Positivity
H-Score
0.25
0.20
0.15
0.10
0.05
0.00
-0.05
60
50
40
30
20
10
0
-10
NAT
Tumor
NAT
Tumor
AA
Positivity
H-Score
p=0.0026
p=0.0033
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
-0.05
80
70
60
50
40
30
20
10
0
-10
NAT
Tumor
NAT
Tumor
EA

FIG. 8C

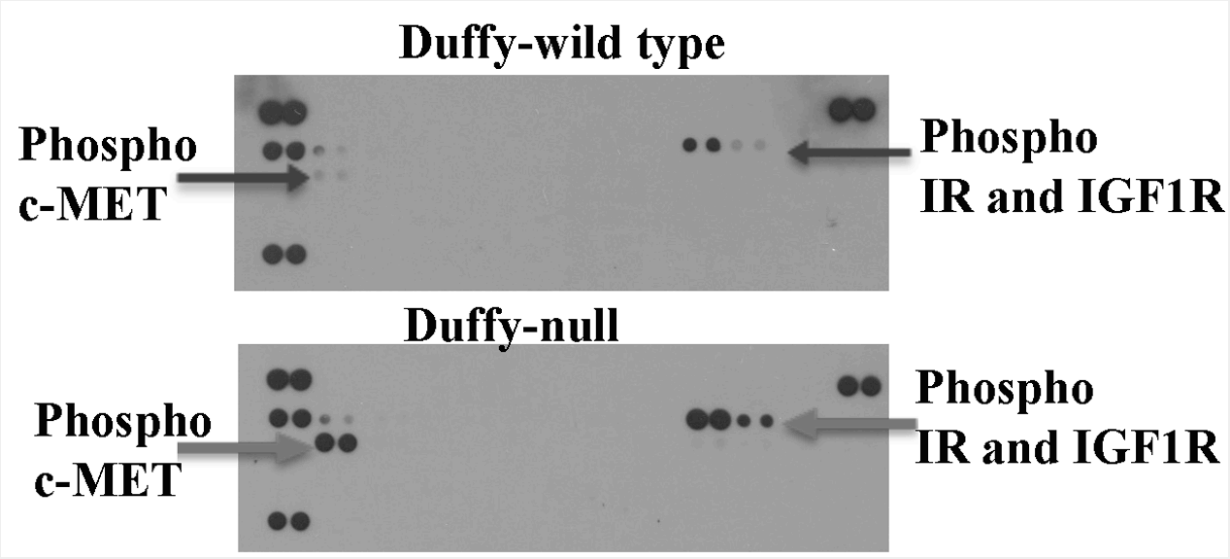

Duffy-wild type
Phospho
c-MET
Phospho
IR and IGF1R
Duffy-null
Phospho
c-MET
Phospho
IR and IGF1R

FIG. 9

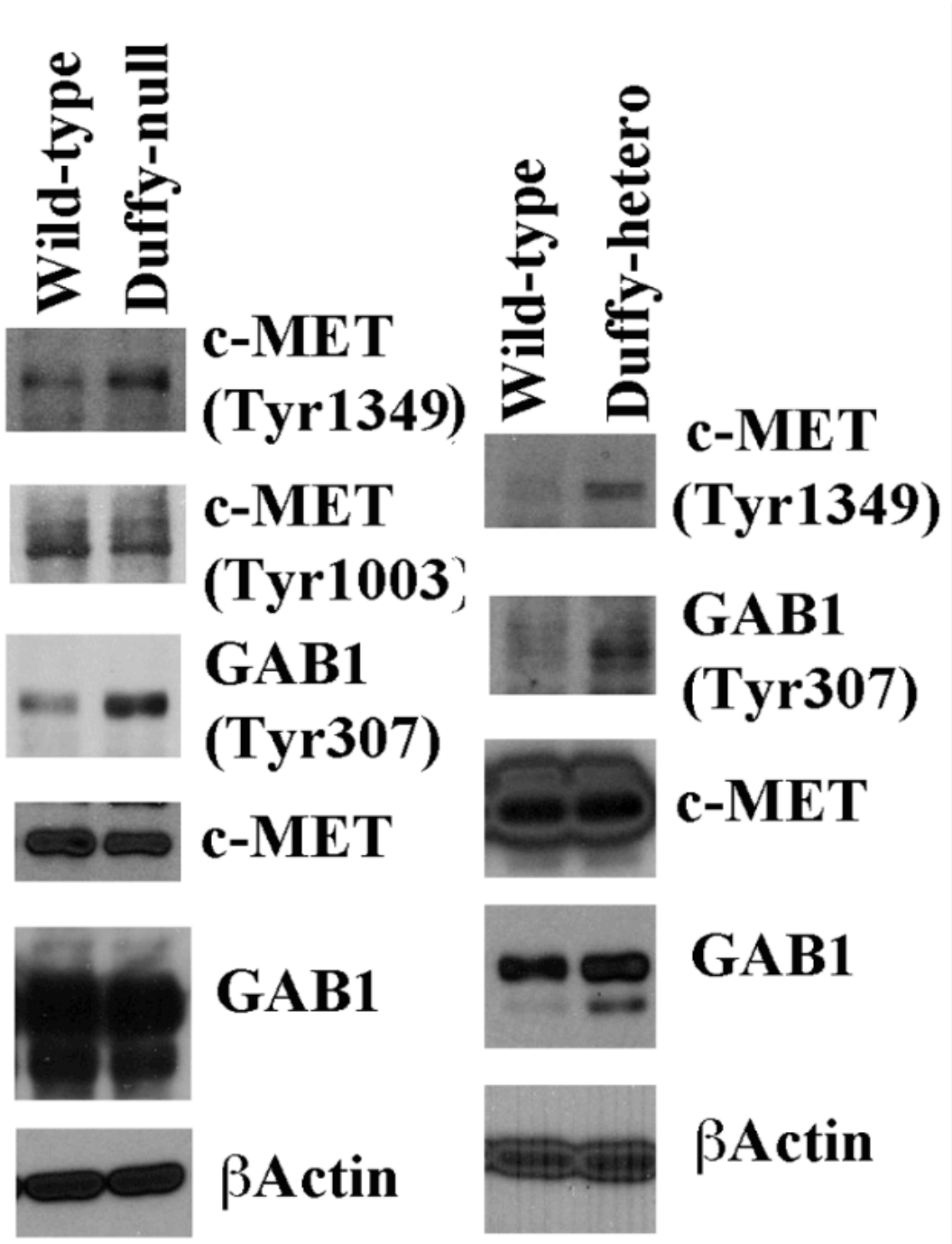

Wild-type
Duffy-null
c-MET (Tyr1349)
c-MET (Tyr1003)
GAB1 (Tyr307)
c-MET
GAB1
βActin
Wild-type
Duffy-hetero
c-MET (Tyr1349)
GAB1 (Tyr307)
c-MET
GAB1
βActin

FIG. 10