# Mutational Changes and Enhanced Cell Migration during Transformation of Breast Non-tumorigenic Cells to Tumorigenic Cells

*Seung Hyuk T. Lee[1], Joon Yup Kim[1], Peter Kim[2], Chia-Yi Su[3], Eun Hyun Ahn[3]**

[1]Department of Pathology, University of Washington, Seattle, WA 98195, USA

[2]Department of Bioengineering, University of Washington, Seattle, WA 98195, USA

[3]Department of Biomedical Engineering, Johns Hopkins University, Baltimore, MD 21205, USA

*Correspondence to:
Eun Hyun Ahn, PhD
Department of Biomedical Engineering, The Johns Hopkins University School of Medicine
Rm. 715B, Ross Research Building, 720 Rutland Avenue, Baltimore, MD 21205
Tel: 410-614-2907, Email: ahn@jhu.edu

Keywords: cancer stem cells, cell migration, extracellular matrix topography, whole exome sequencing, mutation, breast cancer, single-cell

**Abstract**

Although cancer stem cells (CSCs) play a major role in tumor initiation and progression, their mechanisms in invasiveness and metastasis are unclear. Tumor microenvironment signals, such as extracellular matrix (ECM) composition, significantly influence cell and tissue behaviors. Unfortunately, these signals are often lost in *in vitro* cell culture. The current study determines putative CSC populations and contrasts genomic changes in human breast epithelial tumorigenic versus non-tumorigenic cells (both derived from the same normal parental stem cells) and investigates the single-cell migration properties on ECM-mimetic platforms. Whole exome sequencing data indicate that tumorigenic cells have a higher mutation burden than non-tumorigenic cells, and mutations exclusive to tumorigenic cells exhibit higher predictive deleterious scores. ECM-mimetic topography selectively enhances migration speed of tumorigenic cells but not of non-tumorigenic cells. The enhanced migratory ability of tumorigenic cells is lost in conventional two-dimension culture conditions. A wide distribution of tumorigenic single-cell migration speeds suggests heterogeneity in cellular sensing of contact guidance cues. This study identifies mutations acquired during breast tumorigenesis, which can be associated with enhanced invasiveness of breast tumorigenic cells. Results demonstrate that the nanotopographically defined platform can be applied to recapitulate an ECM structure for investigating cellular migration in the simulated tumor microenvironment.

## 1. Introduction

Cancer stem cells (CSCs) account for rapid recurrence and therapeutic resistance after primary cancer treatment, posing difficulties in treatment strategies[1,2]. CSCs contribute to tumor heterogeneity and metastatic dissemination[1]. Breast CSCs have been demonstrated to mediate cancer cell invasiveness and metastasis[1,3]. A subpopulation of breast CSCs characterized with $CD44^{+}/CD25^{-/low}$ are more resistant to therapy and are more invasive than breast non-CSCs[4–6]. Contact guidance of extracellular matrix (ECM) anisotropy that recapitulates a tumor microenvironment is an important factor in modeling tumor progression and metastasis [7–9]. Over the past several years, advances in the fabrication of ECM-mimetic nanoscale topography have facilitated the study of how cells sense contact guidance cues and of cell-matrix interactions [8,10–13]. However, little is known about how the contact guidance of cell migration is affected at a single-cell level by tumorigenicity.

To study the effect of anisotropic contact guidance on cell migration properties, we investigated a human breast epithelial cell (HBEC) model. Our model cells represent multi-stages of breast carcinogenesis, which were sequentially derived from the same parental normal breast epithelial cells with stem cell features. Development, characterization, and culture of these transformed HBECs has been described[14–25] and is highlighted in the *2.1 section* of *Materials and Methods*. In this study, immortalized non-tumorigenic cells (referred to as 'non-tumorigenic cells' hereinafter) and highly tumorigenic-xenograft cells ('tumorigenic cells') were examined.

Our flow cytometry results indicate that the tumorigenic cells have a 10-fold higher putative CSC population than did non-tumorigenic cells. We examined single base substitutions (point mutations) in tumorigenic and non-tumorigenic cells using whole-exome sequencing (WES). WES revealed a higher somatic mutation burden in tumorigenic cells than non-tumorigenic cells and identified somatic mutations in genes associated with breast cancer progression. Using a nanotopography-defined ECM-mimetic platform, we found that anisotropic contact guidance is required to trigger the highly migratory phenotype of tumorigenic cells. The broad distribution of migration speed in tumorigenic cells suggests heterogeneity in cellular sensing of contact guidance cues. Our findings highlight the critical roles of tumor ECM topography in the migration of tumorigenic stem cell-like cells.

## 2. Materials and Methods

### 2.1. Human breast epithelial cells (HBECs) and *in vitro* transformed HBECs

The development and culture of normal HBECs have been described previously[16,20,21]. Immortalized non-tumorigenic, weakly tumorigenic, highly tumorigenic, and highly tumorigenic xenograft cells were transformed sequentially from the same parental normal stem cells with SV40 large T-antigen, x-rays, and the *ERBB2* oncogene[18–20]. Highly tumorigenic cells were injected into nude mice and then the tumors formed in nude mice were collected and grown in culture to develop highly tumorigenic xenograft (M13SV1R2N1-Xeno) cells.

Development, characterization, and culture of these transformed HBECs have been described previously[14–25]. Immortalized non-tumorigenic (M13SV1) cells and highly tumorigenic-xenograft (M13SV1R2N1-xeno) cells were examined in this study. Mutations in normal HBECs with stem cell features (HME13) (previously referred as Type I HBECs)[16,20], that we identified using WES, were used to filter out germline mutations in the non-tumorigenic and tumorigenic cells. The cells (M13SV1R2N1-Xeno, M13SV1, and HME13) were provided by Dr. Chia-Cheng Chang at Michigan State University (MSU) in East Lansing, MI, USA. A material transfer agreement (MTA) was approved by both MSU and the University of Washington (UW) and experiments were performed in UW. The cells were authenticated by short tandem repeat (STR) DNA profiling (Genetica DNA Laboratories, Labcorp brand, Burlington, NC, USA).

### 2.2. Identification of breast cancer stem-like cell population using flow cytometry

Cells were collected two days after initial seeding and were incubated with antibodies labeled with fluorochromes anti-CD24-PE and anti-CD44-APC (BD Biosciences). Cell sorting and immunofluorescence analysis were performed using BD FACS Aria or BD FACSCanto II (BD Immunocytometry Systems). Cell debris and doublets were excluded by using forward and side scatter functions of the FACS instrument. The viable breast cancer stem cell (CSC) population (CD44+/CD24-/low)[14,22,26] was calculated using the FlowJo program version 9.5 (Tree Star, Inc).

### 2.3. DNA library preparation and whole-exome sequencing (WES) data processing

WES procedures for DNA extraction, adapter synthesis, library preparation, and exome capture were carried out as previously described[27,28]. Genomic DNA of normal stem, non-tumorigenic (M13SV1), and tumorigenic-xeno (M13SV1R2N1-xeno) human breast epithelial cells were each extracted using Invitrogen Trizol (ThermoFisher Scientific, Waltham, MA). The extracted DNA was fragmented to approximately 200 bp to make paired-end sequencing libraries. The libraries were captured using SeqCap

EZ Exome v2 (Nimblegen) and hybridized to biotinylated capture probes for paired-end sequencing on an Illumina HiSeq 2500 platform (Illumina Inc.).

Variants were called using the human reference genome hg19 from the University of California Santa Cruz, and only chromosomal positions cited in our exome target BED file were kept. In addition, base positions with unreadable counts >10% of total reads or read coverage below 20 were excluded. Sequence reads were aligned to the targeted exome using BWA v0.6.1-r104 and then were realigned using GATK v1.5-21. The aligned reads were controlled for accuracy by filtering out mapped reads with a MAPping Quality (MAPQ) value below 20 (genome.sph.umich.edu/wiki/Mapping_Quality_Scores). Bases with low calling accuracy (Phred quality score below 13, meaning an error probability of 5%) (www.illumina.com/science/technology/next-generation-sequencing/plan-experiments/quality-scores.html) were not considered for variant calling.

To exclude batch specific sequence artifacts, only positions with a variant read of 5 or greater in at least one sample out of all samples sequenced and prepared from the same WES DNA library preparation experiment were included (i.e., only positions with maximum variant read of 5 or greater among all samples sequenced in the same batch were included). This procedure is analogous to retaining variants representing at least 5% of total reads given average coverage of 100 across all samples sequenced in the same batch. In addition, germline mutations were removed in non-tumorigenic and tumorigenic cells by removing mutations found in the matching normal breast epithelial stem cells.

Mutations in non-tumorigenic and tumorigenic cells were annotated with predicted deleterious scores by CHASMplus, CHASMplus-BRCA[29], and CADD Exome[30] using OpenCRAVAT[29]. Pathogenicity scores by CHASMplus and CHASMplus-BRCA range between 0 and 1, with higher scores indicating a greater likelihood of being a driver mutation. CHASMplus scores were derived using the 32 cancer types from The Cancer Genome Atlas (TCGA), and CHASMplus_BRCA scores were trained using Breast Invasive Carcinoma samples from TCGA.

### 2.4. Fabrication of cell adhesion substrate mimicking nanometer-scale features of the extracellular matrix (ECM)

Cells were cultured on surfaces made from the UV-curable, biocompatible polymer polyurethane-acrylate-301 (PUA) (Minutatek, South Korea). Nanopatterns were generated on a polyethylene terephthalate (PET) film using a silicon master template and UV-assisted capillary force lithography (CFLA)[31]. The nanopatterned PET was used as a template to create nano-grooves on a primer-treated

cover-glass using PUA and UV-assisted CFLA. Each nano-groove (nanopatterened) measured 800nm wide and 800nm high was separated from its neighbor by 800nm to mimic the ~850nm mean spacing between ECM collagen fibril bundles[8,11]. The unpatterned surface was coated with a layer of PUA and cured with UV prior to plasma and collagen treatment to produce a non-textured surface coat. Both unpatterned (UP) and nanopatterned (NP) surfaces were made hydrophilic via treatment with plasma oxygen for 5 minutes at 100W power and with 100 sccm of $O_2$ at 0.5 torr using a Femto Science SUTE-MPR Plasma treatment machine. Then, to further enhance cellular attachment[32], the plasma-treated surfaces were coated with 35 µg/mL of Collagen I Rat Protein, Tail (cat# A1048301, Gibco Thermo Fisher Scientific, Waltham, MA, USA) and incubated at 37°C for 3 hours.

### 2.5. Live cell imaging

Non-tumorigenic and tumorigenic human breast epithelial cells were seeded on unpatterned and nanopatterned PUA301 surfaces for six hours (17 hours for the supplementary data experiment) and were then imaged for 12 hours using a Nikon Eclipse Ti microscope. Images of the cells were taken for 12 hours at 24 spots for each group, for a total of 96 spots, at the intervals of 20 minutes for the main experiment (or 10 minutes for the supplementary data experiment). The cells were incubated at 37°C and 5% $CO_2$ during the live-cell imaging. The images taken at different time points were sequentially compiled to generate a movie file for each spot imaged.

### 2.6. Cell migration quantification

Individual cells were tracked to quantify their speed, persistence, and trajectory using the ImageJ program, version 1.49 (https://imagej.nih.gov/ij/). For each movie of migrating cells, the Manual Tracking plugin was used to establish coordinate points, distance, velocity, and pixel values of individual cells at each time point. These established data were then used with in-house MATLAB scripts to calculate speed and persistence and to generate trajectory graphs of each manually tracked single-cell. Migration speeds of the same cell type seeded on the same surface (unpatterned *vs* nanopatterned) were averaged. Only single, viable cells that did not divide and that survived during the entire time of live cell imaging were included for the single-cell migration quantification.

### 2.7. Statistical analysis

Differences in the average CHASMplus scores for mutations exclusive to non-tumorigenic or tumorigenic cells were analyzed using the Mann-Whitney U-test. Differences in the average migration speed between the two groups were analyzed by the ANOVA on Ranks test using SigmaPlot (version 12.0, Systat Software, Inc., San Jose, CA, USA). Differences between the groups were considered significant at $p < 0.05$. Differences in the distributions of individual cell migration speed were analyzed by the Kolmogorov-Smirnov test using R-3.6.3.

## 3. Results

### 3.1. Higher percentages of putative breast CSCs are present in tumorigenic cells than in non-tumorigenic cells

Putative breast CSC populations were determined using the well-known breast CSC marker $CD44^{+}/CD24^{-/low}$[14,24] via flow cytometry (**Figure 1**). These data indicate that tumorigenic cells contain a higher percentage of CSCs (~27.3%) than in non-tumorigenic cells (~2.78) (**Figure 1**) by about 10-fold.

### 3.2. Whole exome sequencing (WES) identifies somatic mutations in genes associated with breast cancer.

WES was performed on the DNA of normal stem, non-tumorigenic and tumorigenic cells. Somatic mutations in non-tumorigenic and tumorigenic cells were identified after filtering out matching variants present in normal stem cells, which are presumably germline mutations. Only mutations with variant allele frequency (VAF) of at least 5% were included. A total of 175 point mutations across 133 genes were present in the non-tumorigenic cells but not in the tumorigenic cells (**Figure 2A, Table S1A**). Nonsynonymous mutations were most common in the non-tumorigenic cells followed by synonymous, untranslated region (UTR), and intronic mutations.

A higher somatic mutation burden was observed in the tumorigenic cells with a total of 230 point mutations across 81 genes (**Figure 2A, Table S1B-**). In contrast to the non-tumorigenic cells, synonymous mutations were most common in the tumorigenic cells, followed by nonsynonymous, UTR, and intronic mutations. Most of the mutations present exclusively to the tumorigenic cells were found in the gene *ERBB2*, which is related to a transformation step of the *ERBB2* oncogene overexpression (**Figure S1A, B, Table S1B**) introduced during development of this cell model. All other mutations were identified in

genes that have been previously reported to be associated with breast cancer, such as *FANCD2*[33,34], *MMP13*[35], *MSH2*[36], and *RET* [37].

From the total of 175 and 230 mutations exclusive to non-tumorigenic or tumorigenic cells, respectively, we annotated 95 non-tumorigenic exclusive mutations and 65 tumorigenic exclusive mutations with CHASMplus scores[29] to predict pathogenicity. The scores were generated based on all 32 cancer types and on breast cancer alone (**Figure 2B**). The number of genes that have mutations annotated with the CHASMplus scores are identical in both settings (32 cancer types; breast cancer alone). The genes with high CHASMplus scores mutations such as *ITCH*[38], *CHD8*[39], and *SREBP-2*[40] have also been linked to breast cancer progression.

We identified mutations that became increasingly mutated (clonally expanded) during tumorigenesis of HBECs. A total of 74 mutations were identified on exons of 54 genes with higher VAFs in tumorigenic cells than in non-tumorigenic cells by at least 1.5-fold (**Figure 3, Table S1C**). Most of the identified mutations are C>T/G>A transitions, followed by C>G/G>C transversions and T>C/A>G transitions (**Figure 3A, Table S1C**). Among these 74 mutations, the mutations with the three highest scores of CHASMplus-all-cancers (0.164, 0.129, and 0.124) and of CHASMplus-breast-cancer (0.068, 0.073, and 0.034) are: C>T in CHD1, C>A in GNAS, and A>G in CYBA, respectively (**Figure 3A, Table S1C**). These top three mutations are missense mutations (E986K, A436D, and V174A, respectively).

VAFs of 30 exonic mutations were expanded by two-fold or greater including G>T and C>G mutations in *MUC2*[41,42], a C>T mutation in *LAMA5*[43] (**Figure 3B, Table S1C**). Both genes are known to be associated with metastasis of breast cancer. Among the clonally expanded mutations by approximately two-fold (>1.9x), three heterozygous mutations in *APCDD1L*, *GATA5*, and *GNAS* with VAFs near 50% in non-tumorigenic cells became homozygous mutations with VAFs near 100% in tumorigenic cells (**Figure 3B, Table S1C**). *GATA5*[44] and *GNAS* [45,46][ have been previously implicated in breast cancer. *APCDD1L* mRNA abundance has been associated with survival rates in other cancer types[47] (**Table S1C**).

### 3.3. Cells align and migrate along ECM mimetic topography

We evaluated morphologies of non-tumorigenic and tumorigenic cells growing on nanopatterned (NP) and unpatterned (UP) surfaces. The results showed both cell types on NP surfaces aligned along the nano-grooves and displayed elongated morphologies, while these cells displayed no changes in cell morphologies on UP surfaces (**Figure 4**).

Single-cell migration trajectories were plotted in the *X* and *Y* directions with the cell's start position set at coordinate (0,0). Trajectories of the non-tumorigenic or tumorigenic cells cultured on UP and NP surfaces were overlaid to produce a single-cell migration trajectory graph using an in-house MATLAB script. The trajectories of non-tumorigenic and tumorigenic cells on UP surfaces are observed to be nonlinear and multi-directional (**Figure 5A,C**). In contrast, the trajectories of both cell types on NP surfaces are mostly linear along the direction of the nano-groove (**Figure 5B,D**). Migration distances are also longer in cells on NP surfaces than cells on UP surfaces. Overall, the total distances migrated by the cells on NP surfaces are greater than those on UP surfaces for both cell types (**Figure 5**).

### 3.4. ECM mimetic topography enhances the migration speed of tumorigenic cells but not non-tumorigenic cells

We analyzed the migration ability of non-tumorigenic and tumorigenic cells on different ECM topographical surfaces. The average single-cell migration speed of tumorigenic cells on NP surfaces was significantly higher ($p<0.05$) than those of tumorigenic cells on UP surfaces and non-tumorigenic cells on NP and UP surfaces (**Figure 6**). In contrast, no significant difference was observed between the average migration speeds of tumorigenic (9.88 μm/hr) and non-tumorigenic (9.44 μm/hr) cells on UP surfaces (**Figure 6**). This indicates that ECM-mimetic topography enhances migration speed of tumorigenic cells but not non-tumorigenic cells. The enhanced migratory ability of tumorigenic cells is clearly lost in conventional two-dimension culture conditions, demonstrating they lack the capacity to recapitulate behaviors of tumors cell observed *in vivo*.

A greater range of individual single-cell migration speeds was observed in tumorigenic cells (NP 8-38 μm/hr; UP 5-20 μm/hr) than in non-tumorigenic cells (NP 5-25 μm/hr; UP 6-16 μm/hr) (**Figure 6**). This wider range of single-cell migration speeds reflects heterogeneity of tumorigenic cells which contain a significantly higher subpopulation of CSCs than non-tumorigenic cells (**Figure 1**).

An independent cell culture experiment was conducted for the single-cell migration analysis with tumorigenic cells. All experimental conditions were maintained, and cell migration was quantified in the same manner as for the results presented in Figures 4-6. Consistent results were obtained. For example, the average migration speed was significantly higher for tumorigenic cells on NP surface than on UP surface ($p < 0.001$) (**Figure S3**). Furthermore, a similar significant difference in the distribution of individual cell migration speeds and migration trajectories (**Figure S4**) of tumorigenic cells was observed between UP and NP surfaces.

## 4. Discussion

In this study, we evaluated the invasiveness of human breast CSCs by quantifying and comparing migratory properties of non-tumorigenic and tumorigenic cells. Both cell types were sequentially derived from the same parental normal human breast epithelial stem cells (HBECs) with stem cell features[16,20,21]. To recapitulate tumor microenvironment, we cultured these cells (non-tumorigenic vs tumorigenic) on our ECM-mimetic platforms. Our results demonstrate that tumorigenic cells lose their invasiveness and heterogenous properties of cellular sensing on surfaces lacking ECM-mimetic structures. This loss of invasiveness and cellular heterogeneity of tumorigenic cells is recovered on an ECM-mimetic surface.

WES results reveal an increased somatic mutation burden in tumorigenic cells than in non-tumorigenic cells. A greater mutational burden has been associated with metastatic breast cancer[48]. With respect to the distribution of synonymous and nonsynonymous somatic mutations, the proportion of synonymous mutations was greater than that of nonsynonymous mutations in tumorigenic cells. The first evidence that not all synonymous mutations are silent mutations was manifested by Kimchi et. al. (2007), which found that a synonymous mutation in the multidrug resistance 1 (MDR1) gene altered the effectiveness of inhibitors of P-glycoprotein[49]. Since then, synonymous mutations have been receiving more attention insofar as their potential impacts in cancer development. For example, some synonymous mutations in oncogenes can act as driver mutations by altering exonic motifs[50]. Synonymous mutations can alter splicing sites, mRNA stability, miRNA binding sites or translation and can lead to exon skipping and thus a change in protein structure of some of tumor suppressor genes/proteins[51]. In a pan-cancer analysis of 659,194 synonymous mutations identified in TCGA, ICGC, and COSMIC databases, 176,594 synonymous mutations were found recurrently across all tumors.

We identified genes that were mutated exclusively in tumorigenic cells or clonally expanded during breast tumorigenesis. These genes can be candidate genes for breast tumorigenesis and metastasis. Among the 81 genes that contain mutations exclusive to tumorigenic cells, *ERBB2*[52], *FANCD2*[33,34], *MMP13*[35], *MSH2*[36], *RET*[37], *ITCH*[38], *CHD8*[39], and *SREBP-2*[40] have been previously reported to be associated with increased risk of breast cancer. It was demonstrated that the overall survival of patients of breast tumors containing *ERBB2* mutations was lower than that of patients with wild type *ERBB2*[52]. Other genes in the list include *CLASP1*, *RAP2B*, *RET*, and *SEMA3E,* all of which have a role in regulating cell migration or cell-matrix adhesion based on the Gene Ontology[53]. *RAP2B,* in particular, has been

previously reported to be overexpressed in multiple breast cancer cell lines and to increase calcium levels[54]. The elevated calcium level promotes breast cancer cell migration and invasion via the calcium related ERK1/2 signaling pathway[54]. Some of clonally expanded mutations in tumorigenic cells were found in genes that have been associated with breast cancer, such as *AKT3*[55], *GNAS*[45,46], *GATA5*[44], *LAMA5*[43], *MUC2*[41,42], *NRXN2*[32] , *RELN*[56], and *RPA1*[57]. Notably, mutations in *LAMA5* and *MUC2* have high fold-changes of 2.9 and 2.7, respectively. A previous study reported that exosomes derived from non-metastatic breast cancer cells compared to those derived from metastatic breast cancer cells were differentially enriched for *LAMA5*, a gene implicated for cell adhesion[43]. *MUC2* mediated proliferation and metastasis of breast cancer cells[41,42]. Expression of *MUC2* was weak or absent in healthy breast tissue, and patients with high *MUC2* expression were observed to have a significantly shorter survival[41]. In addition, mutations in *GNAS* and *GATA5* expanded from heterozygous mutations (VAFs near 50%) to homozygous mutations (VAFs near 100%), with the mutation in *GNAS* having the highest CHASMplus scores among clonally expanded mutations by greater than 1.5-fold. A previous study reported that the high expression of *GNAS* is associated with distal metastasis and poor survival rates based on 150 breast cancer samples[46]. siRNA-mediated *GNAS* knockdown experiments showed reduced viability and proliferation, impaired migration ability, and induced G1 cell cycle arrest in breast cancer cells[45,46]. Potential mechanisms for these changes induced by the high expression of *GNAS* involve the activation of the PI3K/AKT signaling pathway[46] and overexpression of the extra-long Gαs (XLαs) isoform, which induces cAMP levels[45]. *GATA5* has been reported to activate the progesterone receptor gene promoter under the presence of the progesterone receptor +331G/A gene variant, a variant that is associated with breast cancer risk[44]. Furthermore, *GATA5* has been previously investigated as a potential tumor suppressor in colorectal cancer[58], hepatocellular carcinoma[59], and lung adenocarcinoma[60]. These mutations, acquired or expanded during breast tumorigenesis, that we identified, can account for, at least in part, enhanced migratory ability of tumorigenic cells.

To recapitulate tumor microenvironment, we have applied an ECM-mimetic, nanotopographically-defined cell culture platform, which provides nanoscale contact guidance cues. Previously, we applied this platform to study cell migration in human brain cancer[10], melanoma[61], and breast immortalized[11] cells. Features of our ECM-mimetic culture platform include treatment of the surface with collagen type I to enhance cellular attachment[62] and 800 nm spacing of nanogrooves to resemble fibril spacings in mammary collagen[11,63]. We have demonstrated that this platform provides nanoscale contact guidance cues evidenced by cell alignment along edges of nanogrooves and

enhancement in cell migration. Contact guidance is a strong regulator of directed migration[8] and plays a major role in cancer cell invasion[64]. The presence of nanoscale contact guidance in our study is manifested by the aligned and elongated morphologies of both non-tumorigenic and tumorigenic cells along the edges of nanogrooves.

We observed that the average migration speed of tumorigenic cells is significantly higher on nanopatterned surfaces compared to those on unpatterned surfaces, suggesting that nanoscale contact guidance cues enhance the invasiveness of breast tumorigenic cells. In contrast, we observed no change in migration speed for non-tumorigenic breast cells on nanopatterned surfaces vs. unpatterened surfaces. This indicates that contact guidance cues alone cannot enhance the migratory ability of non-tumorigenic cells. It is possible that non-tumorigenic cells may lack mutations and epigenetic changes that are present in tumorigenic cells that induce more responsiveness to contact guidance cues. Selective enhancement of migratory ability for tumorigenic cells on our ECM-memetic topography supports the idea that this platform can recapitulate cellular migration observed in a tumor microenvironment fostered by the ECM structure.

The linear, as opposed to random, nature of cell trajectories in the presence of nanoscale guidance cues has been suggested as a possible mechanism for enhanced cell migration speed[11,65,66]. Our results indicate that trajectories for both non-tumorigenic and tumorigenic cells are more linear on nanopatterned surfaces than those on the unpatterned surfaces; however, increased average migration speed was only observed for tumorigenic cells. This suggests that contact guidance cues likely trigger distinct alterations in cells that enhance their physical movement rather than through more directed and linear migration.

In addition to higher average migration speeds on NP versus UP surfaces, tumorigenic cells displayed significant differences in the distributions of individual cell speeds on NP surfaces than on UP surfaces, as verified by the Kolmogorov-Smirnov test. This suggests heterogeneity in tumorigenic cells. No such significant differences in individual cell migration speeds were observed for both cell types on unpatterned surfaces and for non-tumorigenic cells on nanopatterned surfaces. To our knowledge, this migration aspect has not yet been investigated at single-cell level, as other studies focused on collective cell migration. Individual breast tumorigenic cells possibly acquire different somatic mutations and epigenetic alterations that change how each cell senses and responds to contact guidance cues in the tumor microenvironment. Future studies should characterize various cell migration speed subpopulations of tumorigenic cells on nanopattern surfaces to ascertain, for example, if higher speeds are found in breast

CSC-like populations. This would test if CSCs are more susceptible to the contact guidance cues associated with enhanced invasiveness.

In the present study, we have investigated mutational changes of the whole exome and performed a single-cell migration assessment on the ECM-mimetic nanoscale cell culture platform in tumorigenic *vs*. non-tumorigenic cells (both derived from the same normal parental stem cells). We have identified groups of mutations as candidates associated with breast tumorigenesis and the varying degrees of breast tumor invasiveness. Our data indicate that ECM-mimetic topography selectively enhances migratory ability of breast tumorigenic cells, but not non-tumorigenic cells. This demonstrates that our nanotopographically defined platform can recapitulate an ECM structure of the tumor microenvironment to study cell migration.

**Author Contributions**

Conceptualization: EHA; Experiments: EHA, JYK, SHTL, PK; Data curation: EHA, SHTL, JYK, PK; Data analysis: SHTL, JYK, EHA, PK; Funding acquisition: EHA; Methodology: SHTL, JYK, EHA, PK; Project administration, EHA; Resources: EHA; Supervision: EHA; Visualization: SHTL, EHA, JYK, CYS; Writing: SHTL, EHA; Editing and review: SHTL, EHA, CYS.

**Funding**

The research was supported by grants from the National Institute of Environmental Health Sciences (NIEHS) P30 ES007033 sponsored-University of Washington (UW) Center for Exposures, Diseases, Genomics and Environment (EDGE) pilot grant (to EHA), UW Office of Research Royalty Research Fund (to EHA), and the National Cancer Institute (NCI) P30 CA015704-39 Fred Hutchinson Cancer Research Center-UW Cancer Consortium Support Grant (to EHA), NCI R21 CA220111 (to EHA) of the National Institutes of Health (NIH). The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

**Acknowledgments**

We thank Drs. Chia-Cheng Chang, Brad L. Upham, and James E. Trosko for obtaining non-tumorigenic immortalized and tumorigenic/xenograft cells; Dr. Deok-Ho Kim for providing access to Nikon Eclipse Ti microscope and Femto Science SUTE-MPR Plasma system; Zhipeng Dong for mutation annotations using CHASMplus; Howard Nebeck for proofreading and editing the manuscript; and Jiwon Min for managing references of the manuscript.

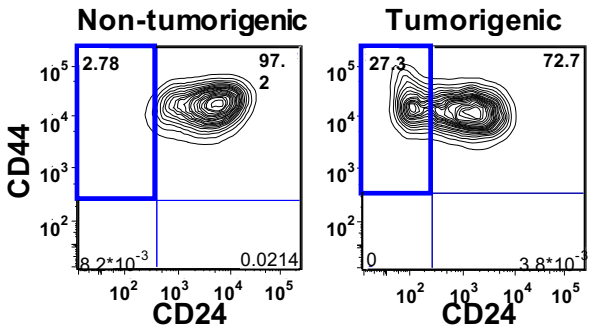


**Figure 1. The percentages of cells in each of four subpopulations, defined by the expressions of CD44 and CD24, were determined using flow cytometry in human breast non-tumorigenic and tumorigenic cells.** Bold and blue lines represent the percentage of putative breast CSC populations ($CD44^{+}/CD24^{-/low}$).

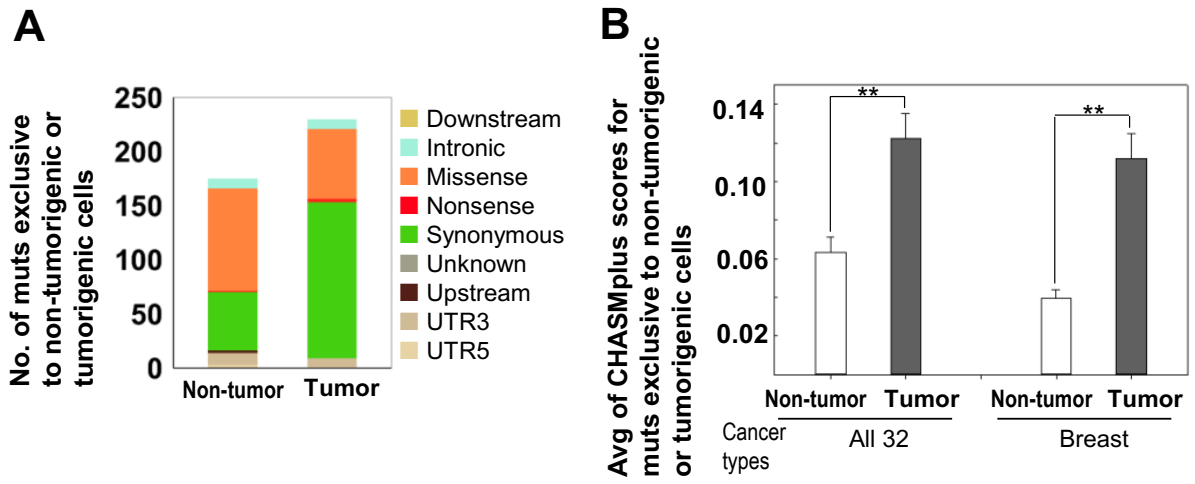


**Figure 2. Mutations exclusive to human breast non-tumorigenic or tumorigenic cells were identified using whole exome sequencing.** Mutations (muts) with variant allele frequencies less than 5% and mutations found in the matched normal cells (HME13) were filtered out. **(A)** Numbers of mutations exclusive to non-tumorigenic or tumorigenic cells. Each color represents proportion of each category of point mutations annotated using ANNOVAR. Abbreviations used are: UTR3, three prime untranslated region; UTR5, five prime untranslated region. **(B)** Averages (Avg) of CHASMplus scores based on all 32 cancer types and on breast cancer alone for mutations exclusive to non-tumorigenic or tumorigenic cells. $^{**}p < 5\text{x}10^{-4}$ by the Mann–Whitney U-test.

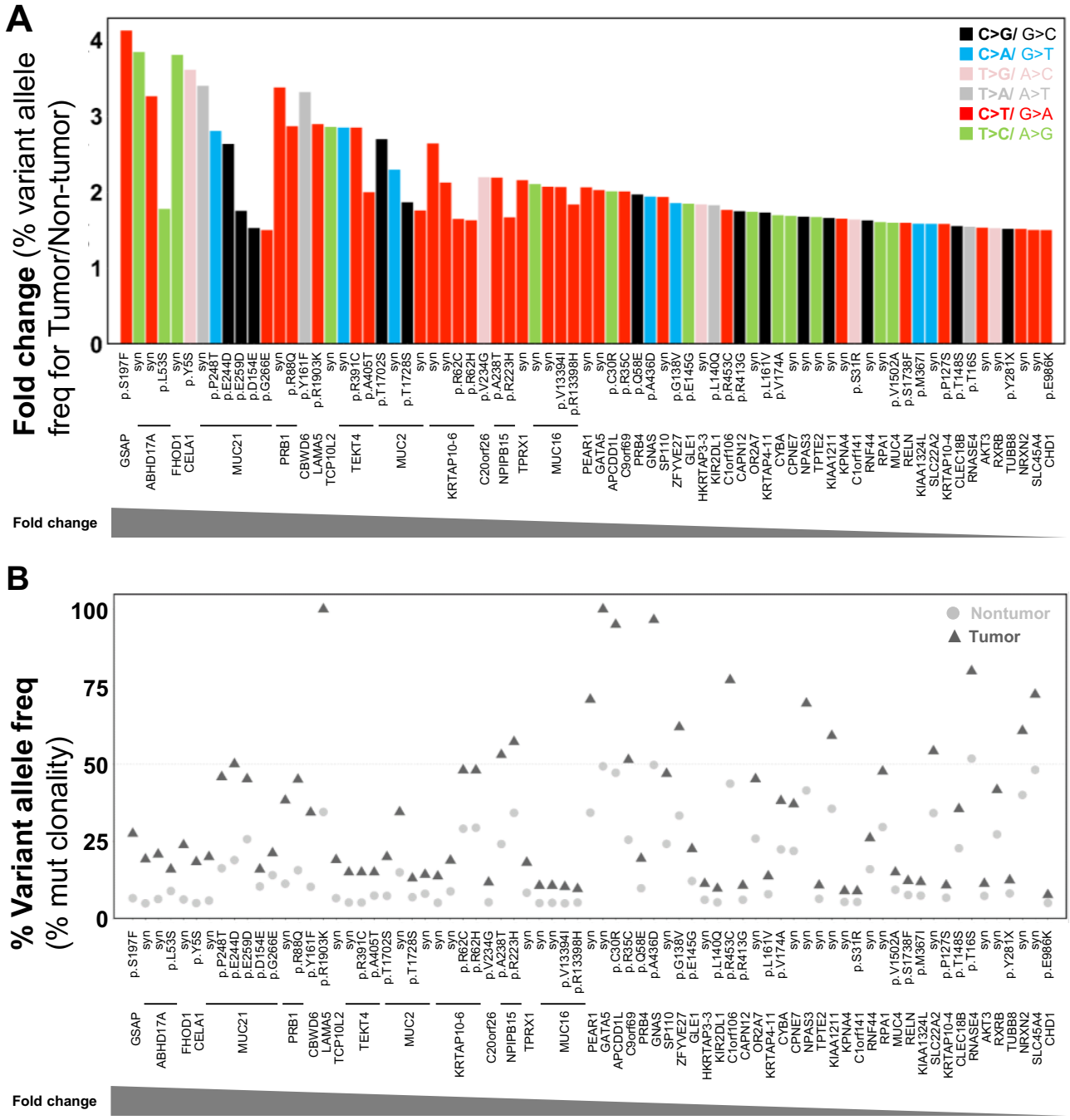


**Figure 3. Somatic mutations that are clonally expanded during tumorigenesis were identified by whole exome sequencing.** Mutations with variant allele frequencies (freq) less than 5% and mutations found in the matched normal cells (HME13) were removed. **(A)** Fold changes of clonally expanded somatic mutation allele frequencies between non-tumorigenic and tumorigenic cells colored by mutation types. **(B)** Variant allele frequencies (%) of each clonally expanded somatic mutation are displayed.

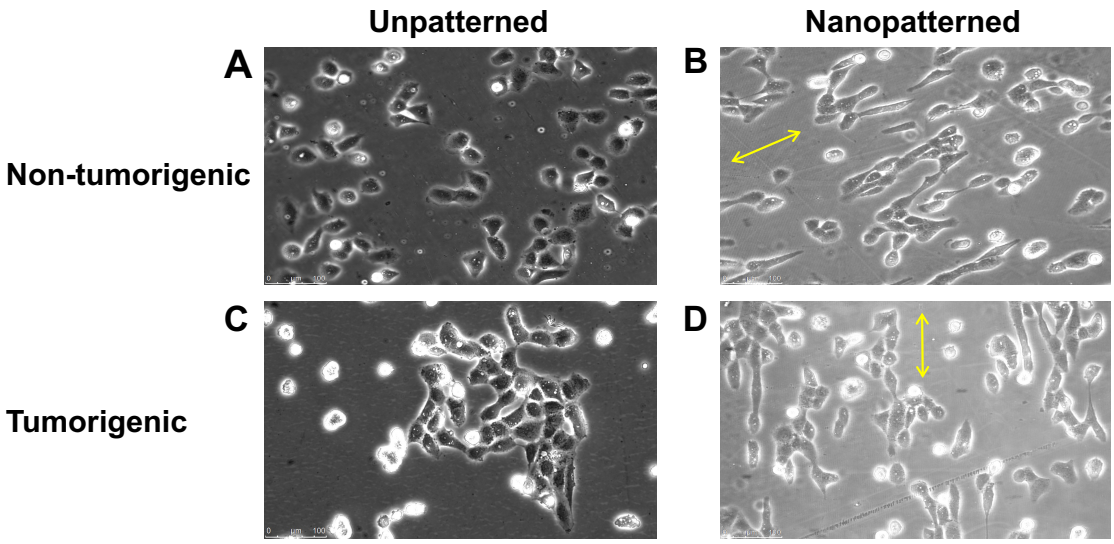


**Figure 4. Both non-tumorigenic and tumorigenic human breast epithelial cells align along ECM mimetic topography.** Human breast non-tumorigenic and tumorigenic cells were seeded on unpatterned and nanopatterned biocompatible polymer polyurethane-acrylate-301 (PUA301) substrates. Unlike the cells on unpatterned substrates **(A,C)**, both cell types align along the directions of nanogrooves and have elongated morphologies on nanopattered substrates **(B,D)**. Images were taken 3.5 hours after cell seeding. Yellow arrows represent directions of nanogrooves. Scanning electron microscope image of a nanopatterned substrate is shown in **Fig S2**.

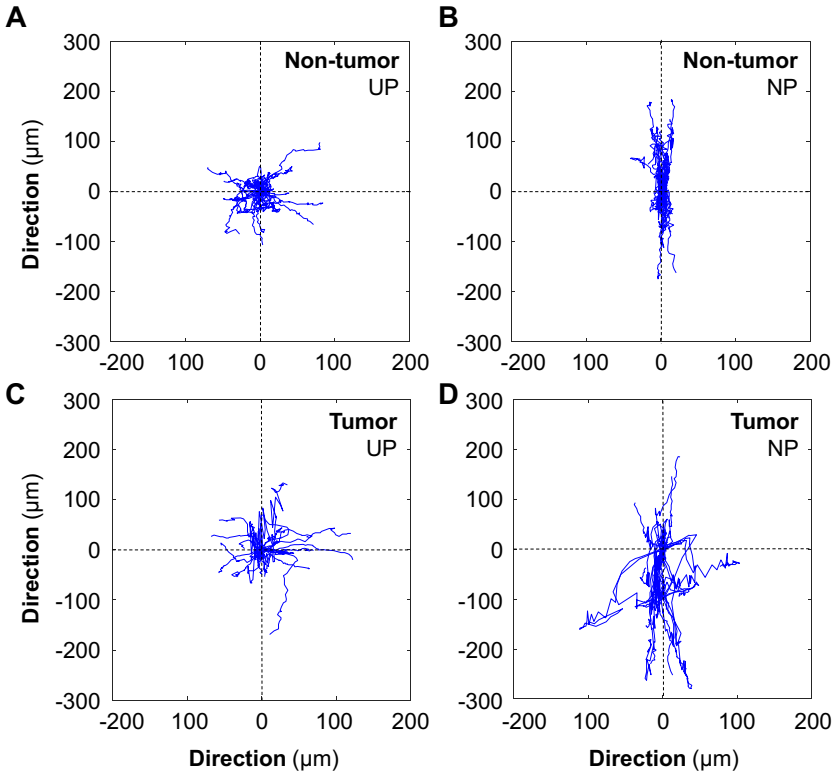


**Figure 5. ECM-mimetic topography regulates the persistence of cell migration.** The trajectories of a 12-hour single-cell migration are presented for human breast non-tumorigenic and tumorigenic cells cultured on unpatterned (UP) **(A,C)** and nanopatterned (NP) **(B,D)** PUA301 surfaces. Migration trajectories of each single-cell were overlaid for each group and all starting points were set to the coordinate (0,0).

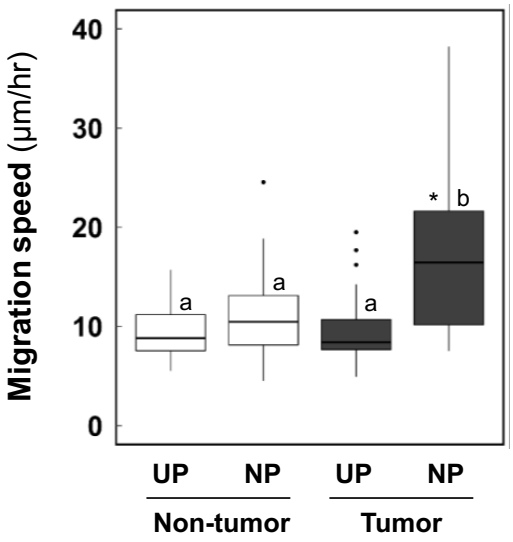


**Figure 6. ECM-mimetic topography enhances the migration speed of tumorigenic cells but not non-tumorigenic cells.** Non-tumorigenic and tumorigenic cells were seeded on unpatterned (UP) and nanopatterned (NP) PUA301 substrates. Single-cells were individually tracked to calculate migration speed (non-tumorigenic UP n= 40, non-tumorigenic NP n=50, tumorigenic UP n=26, tumorigenic NP n=28). Box and whisker plots show an increase in the migration speed of tumorigenic cells on NP (ECM mimetic topography) (a,b: $p < 0.05$ by the ANOVA on ranks). Significant differences in distribution of individual single-cell migration speed are indicated ($^{*}$: $p < 0.05$) by the Kolmogorov-Smirnov test.

**References**

[1] F. Yang, J. Xu, L. Tang, X. Guan, *Cell Mol Life Sci* **2017**, *74*, 951.
[2] A. Wang, L. Chen, C. Li, Y. Zhu, *Cancer Lett* **2015**, *357*, 63.
[3] M. A. Velasco-Velazquez, N. Homsi, M. De La Fuente, R. G. Pestell, *Int J Biochem Cell Biol* **2012**, *44*, 573.
[4] K. K. Tan, C. S. Giam, M. Y. Leow, C. W. Chan, E. K. Yim, *J Funct Biomater* **2015**, *6*, 241.
[5] W. Yan, Y. Chen, Y. Yao, H. Zhang, T. Wang, *Cancer Cell Int* **2013**, *13*, 62.
[6] C. Sheridan, H. Kishimoto, R. K. Fuchs, S. Mehrotra, P. Bhat-Nakshatri, C. H. Turner, R. Goulet, S. Badve, H. Nakshatri, *Breast Cancer Res* **2006**, *8*, R59.
[7] P. P. Provenzano, D. R. Inman, K. W. Eliceiri, J. G. Knittel, L. Yan, C. T. Rueden, J. G. White, P. J. Keely, *BMC Med* **2008**, *6*, 11.
[8] A. Ray, O. Lee, Z. Win, R. M. Edwards, P. W. Alford, D. H. Kim, P. P. Provenzano, *Nat Commun* **2017**, *8*, 14923.
[9] M. W. Conklin, J. C. Eickhoff, K. M. Riching, C. A. Pehlke, K. W. Eliceiri, P. P. Provenzano, A. Friedl, P. J. Keely, *Am J Pathol* **2011**, *178*, 1221.
[10] C. L. Smith, O. Kilic, P. Schiapparelli, H. Guerrero-Cazares, D. H. Kim, N. I. Sedora-Roman, S. Gupta, T. O'Donnell, K. L. Chaichana, F. J. Rodriguez, S. Abbadi, J. Park, A. Quinones-Hinojosa, A. Levchenko, *Cell Rep* **2016**, *15*, 2616.
[11] K. H. Nam, P. Kim, D. K. Wood, S. Kwon, P. P. Provenzano, D. H. Kim, *Sci Rep* **2016**, *6*, 29749.
[12] T. Q. Huang, X. Qu, J. Liu, S. Chen, *Biomed Microdevices* **2014**, *16*, 127.
[13] A. S. Meyer, S. K. Hughes-Alford, J. E. Kay, A. Castillo, A. Wells, F. B. Gertler, D. A. Lauffenburger, *J Cell Biol* **2012**, *197*, 721.
[14] E. H. Ahn, S. H. Lee, J. Y. Kim, C. C. Chang, L. A. Loeb, *Cancer Res* **2016**, *76*, 4569.
[15] J. S. Park, D. Y. Noh, S. H. Kim, S. H. Kim, G. Kong, C. C. Chang, Y. S. Lee, J. E. Trosko, K. S. Kang, *Int J Oncol* **2004**, *24*, 1545.
[16] E. H. Ahn, C. C. Chang, J. J. Schroeder, *Exp Biol Med (Maywood)* **2006**, *231*, 1664.
[17] M. H. Tai, C. C. Chang, M. Kiupel, J. D. Webster, L. K. Olson, J. E. Trosko, *Carcinogenesis* **2005**, *26*, 495.
[18] C. C. Chang, W. Sun, A. Cruz, M. Saitoh, M. H. Tai, J. E. Trosko, *Radiat Res* **2001**, *155*, 201.
[19] K. S. Kang, I. Morita, A. Cruz, Y. J. Jeon, J. E. Trosko, C. C. Chang, *Carcinogenesis* **1997**, *18*, 251.
[20] C. Y. Kao, K. Nomata, C. S. Oakley, C. W. Welsch, C. C. Chang, *Carcinogenesis* **1995**, *16*, 531.
[21] K. S. Kang, W. Sun, K. Nomata, I. Morita, A. Cruz, C. J. Liu, J. E. Trosko, C. C. Chang, *Mol Carcinog* **1998**, *21*, 225.
[22] K. H. Wang, A. P. Kao, C. C. Chang, J. N. Lee, M. F. Hou, C. Y. Long, H. S. Chen, E. M. Tsai, *Mol Cancer* **2010**, *9*, 288.
[23] E. H. Ahn, C. C. Chang, D. A. Talmage, *Anticancer Res* **2009**, *29*, 2899.
[24] E. H. Ahn, K. Hirohata, B. F. Kohrn, E. J. Fox, C. C. Chang, L. A. Loeb, *PLoS One* **2015**, *10*, e0136216.
[25] C. Y. Kao, C. S. Oakley, C. W. Welsch, C. C. Chang, *In Vitro Cell Dev Biol Anim* **1997**, *33*, 282.
[26] M. Al-Hajj, M. S. Wicha, A. Benito-Hernandez, S. J. Morrison, M. F. Clarke, *Proc Natl Acad Sci U S A* **2003**, *100*, 3983.
[27] E. H. Ahn, S. H. Lee, *Int J Mol Sci* **2019**, *20*, DOI 10.3390/ijms20010199.
[28] S. Gulsuner, T. Walsh, A. C. Watts, M. K. Lee, A. M. Thornton, S. Casadei, C. Rippey, H. Shahin, S. Consortium on the Genetics of, Paartners Study Group, V. L. Nimgaonkar, R. C. Go, R. M. Savage, N. R. Swerdlow, R. E. Gur, D. L. Braff, M. C. King, J. M. McClellan, *Cell* **2013**, *154*, 518.

[29] C. Tokheim, R. Karchin, *Cell Syst* **2019**, *9*, 9.
[30] P. Rentzsch, D. Witten, G. M. Cooper, J. Shendure, M. Kircher, *Nucleic Acids Res* **2019**, *47*, D886.
[31] J. Macadangdang, H. J. Lee, D. Carson, A. Jiao, J. Fugate, L. Pabon, M. Regnier, C. Murry, D. H. Kim, *J Vis Exp* **2014**, DOI 10.3791/50039.
[32] P. Fernandez-Nogueira, P. Bragado, V. Almendro, E. Ametller, J. Rios, S. Choudhury, M. Mancino, P. Gascon, *Oncotarget* **2016**, *7*, 5313.
[33] H. Kobayashi, S. Ohno, Y. Sasaki, M. Matsuura, *Oncol Rep* **2013**, *30*, 1019.
[34] E. Barroso, R. L. Milne, L. P. Fernandez, P. Zamora, J. I. Arias, J. Benitez, G. Ribas, *Carcinogenesis* **2006**, *27*, 1930.
[35] B. Zhang, X. Cao, Y. Liu, W. Cao, F. Zhang, S. Zhang, H. Li, L. Ning, L. Fu, Y. Niu, R. Niu, B. Sun, X. Hao, *BMC Cancer* **2008**, *8*, 83.
[36] F. J. Couch, H. Shimelis, C. Hu, S. N. Hart, E. C. Polley, J. Na, E. Hallberg, R. Moore, A. Thomas, J. Lilyquist, B. Feng, R. McFarland, T. Pesaran, R. Huether, H. LaDuca, E. C. Chao, D. E. Goldgar, J. S. Dolinsky, *JAMA Oncol* **2017**, *3*, 1190.
[37] B. S. Paratala, J. H. Chung, C. B. Williams, B. Yilmazel, W. Petrosky, K. Williams, A. B. Schrock, L. M. Gay, E. Lee, S. C. Dolfi, K. Pham, S. Lin, M. Yao, A. Kulkarni, F. DiClemente, C. Liu, L. Rodriguez-Rodriguez, S. Ganesan, J. S. Ross, S. M. Ali, B. Leyland-Jones, K. M. Hirshfield, *Nat Commun* **2018**, *9*, 4821.
[38] Q. Yin, C. J. Wyatt, T. Han, K. S. M. Smalley, L. Wan, *Semin Cancer Biol* **2020**, *67*, 117.
[39] A. A. Mills, *Cold Spring Harb Perspect Med* **2017**, *7*, DOI 10.1101/cshperspect.a026450.
[40] Z. Jie, Z. Xie, W. Xu, X. Zhao, G. Jin, X. Sun, B. Huang, P. Tang, G. Wang, S. Shen, A. Qin, S. Fan, *Biochim Biophys Acta Mol Basis Dis* **2019**, *1865*, 115.
[41] P. Mukhopadhyay, S. Chakraborty, M. P. Ponnusamy, I. Lakshmanan, M. Jain, S. K. Batra, *Biochim Biophys Acta* **2011**, *1815*, 224.
[42] A. Astashchanka, T. M. Shroka, B. M. Jacobsen, *Breast Cancer Res Treat* **2019**, *173*, 289.
[43] L. Gangoda, M. Liem, C. S. Ang, S. Keerthikumar, C. G. Adda, B. S. Parker, S. Mathivanan, *Proteomics* **2017**, *17*, DOI 10.1002/pmic.201600370.
[44] G. S. Huggins, J. Y. Wong, S. E. Hankinson, I. De Vivo, *Cancer Res* **2006**, *66*, 1384.
[45] I. Garcia-Murillas, R. Sharpe, A. Pearson, J. Campbell, R. Natrajan, A. Ashworth, N. C. Turner, *Oncogene* **2014**, *33*, 2478.
[46] X. Jin, L. Zhu, Z. Cui, J. Tang, M. Xie, G. Ren, *Clin Transl Oncol* **2019**, *21*, 1207.
[47] S. V. Vasaikar, P. Straub, J. Wang, B. Zhang, *Nucleic Acids Res* **2018**, *46*, D956.
[48] F. Bertucci, C. K. Y. Ng, A. Patsouris, N. Droin, S. Piscuoglio, N. Carbuccia, J. C. Soria, A. T. Dien, Y. Adnani, M. Kamal, S. Garnier, G. Meurice, M. Jimenez, S. Dogan, B. Verret, M. Chaffanet, T. Bachelot, M. Campone, C. Lefeuvre, H. Bonnefoi, F. Dalenc, A. Jacquet, M. R. De Filippo, N. Babbar, D. Birnbaum, T. Filleron, C. Le Tourneau, F. Andre, *Nature* **2019**, *569*, 560.
[49] C. Kimchi-Sarfaty, J. M. Oh, I. W. Kim, Z. E. Sauna, A. M. Calcagno, S. V. Ambudkar, M. M. Gottesman, *Science* **2007**, *315*, 525.
[50] F. Supek, B. Minana, J. Valcarcel, T. Gabaldon, B. Lehner, *Cell* **2014**, *156*, 1324.
[51] Y. Sharma, M. Miladi, S. Dukare, K. Boulay, M. Caudron-Herger, M. Gross, R. Backofen, S. Diederichs, *Nat Commun* **2019**, *10*, 2569.
[52] C. Brueffer, S. Gladchuk, C. Winter, J. Vallon-Christersson, C. Hegardt, J. Hakkinen, A. M. George, Y. Chen, A. Ehinger, C. Larsson, N. Loman, M. Malmberg, L. Ryden, A. Borg, L. H. Saal, *EMBO Mol Med* **2020**, *12*, e12118.

[53] M. Ashburner, C. A. Ball, J. A. Blake, D. Botstein, H. Butler, J. M. Cherry, A. P. Davis, K. Dolinski, S. S. Dwight, J. T. Eppig, M. A. Harris, D. P. Hill, L. Issel-Tarver, A. Kasarskis, S. Lewis, J. C. Matese, J. E. Richardson, M. Ringwald, G. M. Rubin, G. Sherlock, *Nat Genet* **2000**, *25*, 25.
[54] J. Di, H. Huang, D. Qu, J. Tang, W. Cao, Z. Lu, Q. Cheng, J. Yang, J. Bai, Y. Zhang, J. Zheng, *Sci Rep* **2015**, *5*, 12363.
[55] A. Grottke, F. Ewald, T. Lange, D. Norz, C. Herzberger, J. Bach, N. Grabinski, L. Graser, F. Hoppner, B. Nashan, U. Schumacher, M. Jucker, *PLoS One* **2016**, *11*, e0146370.
[56] M. H. Yeh, Y. J. Tzeng, T. Y. Fu, J. J. You, H. T. Chang, L. P. Ger, K. W. Tsai, *Anticancer Res* **2018**, *38*, 4593.
[57] Y. Wang, C. D. Putnam, M. F. Kane, W. Zhang, L. Edelmann, R. Russell, D. V. Carrion, L. Chin, R. Kucherlapati, R. D. Kolodner, W. Edelmann, *Nat Genet* **2005**, *37*, 750.
[58] D. M. Hellebrekers, M. H. Lentjes, S. M. van den Bosch, V. Melotte, K. A. Wouters, K. L. Daenen, K. M. Smits, Y. Akiyama, Y. Yuasa, S. Sanduleanu, C. A. Khalid-de Bakker, D. Jonkers, M. P. Weijenberg, J. Louwagie, W. van Criekinge, B. Carvalho, G. A. Meijer, S. B. Baylin, J. G. Herman, A. P. de Bruine, M. van Engeland, *Clin Cancer Res* **2009**, *15*, 3990.
[59] H. Feng, M. Zhu, R. Zhang, Q. Wang, W. Li, X. Dong, Y. Chen, Y. Lu, K. Liu, B. Lin, J. Guo, M. Li, *J Cell Mol Med* **2019**, *23*, 2536.
[60] W. Ji, L. Zhang, H. Zhu, *Bioengineered* **2022**, *13*, 2878.
[61] J. Park, D. H. Kim, H. N. Kim, C. J. Wang, M. K. Kwak, E. Hur, K. Y. Suh, S. S. An, A. Levchenko, *Nat Mater* **2016**, *15*, 792.
[62] C. S. Szot, C. F. Buchanan, J. W. Freeman, M. N. Rylander, *Biomaterials* **2011**, *32*, 7905.
[63] P. P. Provenzano, K. W. Eliceiri, J. M. Campbell, D. R. Inman, J. G. White, P. J. Keely, *BMC Med* **2006**, *4*, 38.
[64] L. M. Sander, *Cancer Res* **2014**, *74*, 4588.
[65] K. M. Riching, B. L. Cox, M. R. Salick, C. Pehlke, A. S. Riching, S. M. Ponik, B. R. Bass, W. C. Crone, Y. Jiang, A. M. Weaver, K. W. Eliceiri, P. J. Keely, *Biophys J* **2014**, *107*, 2546.
[66] A. Ray, Z. M. Slama, R. K. Morford, S. A. Madden, P. P. Provenzano, *Biophys J* **2017**, *112*, 1023.

## Supplementary Information

**Figure S1.** Numbers of somatic mutations exclusively found in tumorigenic cells for each gene including **(A)** and excluding **(B)** ERBB2 gene. Mutations with variant allele frequencies less than 5% and mutations found in the matched normal cells (HME13) were removed. Colors represent proportion of each mutation type. Abbreviations used are: UTR3, three prime untranslated region; UTR5, five prime untranslated region.

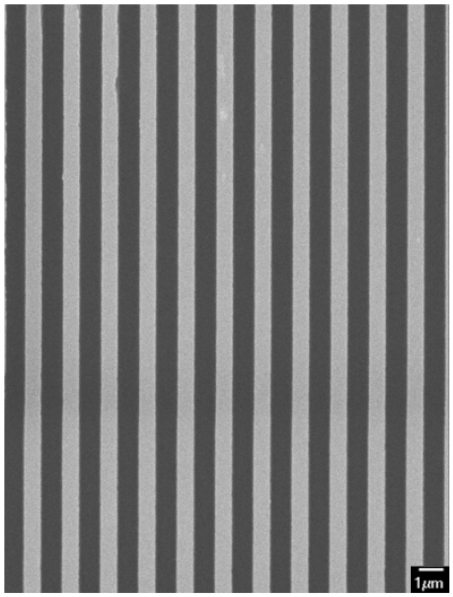


**Figure S2**. A representative scanning electron microscope image of extracellular matrix (ECM)-mimetic nanotopographical model. Each nano-groove is 800nm by 800nm and are separated by 800nm.

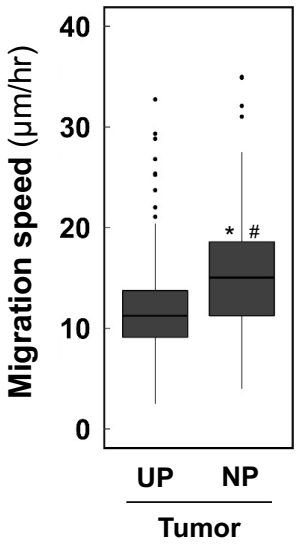


**Figure S3.** Independent cell culture experiments generate reproducible results of migration speeds of human breast tumorigenic cells. Single-cells were individually tracked to calculate the speed of tumorigenic cell migration on unpatterned (UP) and nanopatterned (NP) PUA301 substrates. Box plots indicate average migration speed of each group and distributions of individual cell migration speeds. Significant differences in average migration speeds between the two groups are indicated ($^{*}$: $p < 0.001$) by the Mann-Whitney Rank Sum test. Significant difference in distribution of individual cell migration speeds is indicated ($^{\#}$: $p < 0.05$) by the Kolmogorov-Smirnov test.

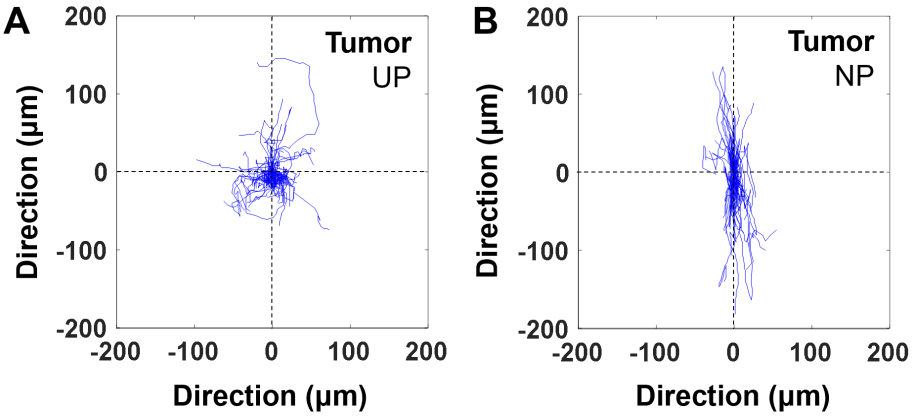


**Figure S4.** Independent cell culture experiments generate reproducible results for trajectories of a 12-hour single-cell migration in human breast tumorigenic cells cultured on unpatterned (UP) **(A)** and nanopatterned (NP) **(B)** PUA301 substrates. Migration trajectories of each cell were overlaid for each group and all starting points were set to the coordinate (0,0).