nature biomedical engineering

SUPPLEMENTARY INFORMATION

https://doi.org/10.1038/s41551-018-0285-z

In the format provided by the authors and unedited.

Live-cell phenotypic-biomarker microfluidic assay for the risk stratification of cancer patients via machine learning

Michael S. Manak[1,6], Jonathan S. Varsanik[1,6], Brad J. Hogan[1], Matt J. Whitfield[1], Wendell R. Su[1], Nikhil Joshi[1], Nicolai Steinke[1], Andrew Min[1], Delaney Berger[1], Robert J. Saphirstein[1], Gauri Dixit[1], Thiagarajan Meyyappan[1], Hui-May Chu[2], Kevin B. Knopf[3], David M. Albala[4], Grannum R. Sant[5] and Ashok C. Chander[1]*

[1]Cellanyx Diagnostics, Beverly, MA, USA. [2]Anoixis Corporation , Natick, MA, USA. [3]Cancer Commons, Los Altos, CA, USA. [4]Associated Medical Professionals of New York, New York, NY, USA. [5]Department of Urology, Tufts University School of Medicine, Boston, MA, USA. [6]These authors contributed equally: Michael S. Manak, Jonathan S. Varsanik. *e-mail: ashok@cellanyx.com

SUPPLEMENTARY INFORMATION

Table of Contents:

Supplementary Tables

A. Technology rationale

In prostate cancer (PCa), the current reference standard used to diagnose men at-risk for PCa, the Gleason score, a formalin-fixed based staining of tumor cells, is challenged with qualitative uncertainty due to observation bias, and has limited prognostic capability. Similarly, in BrCa, static staining of formalin-fixed paraffin-embedded tissue is the current reference standard to predict stage and grade, and suffers from the inability to predict if a nascent cancer lesion (i.e. DCIS) will develop into an aggressive and/or metastatic cancer [1-3]. Recent histochemical, molecular, genetic, epigenetic, genomic, and proteomic techniques have emerged toward better risk stratification of newly diagnosed cancer. However, the clinical utility and actionability of these tests have not been firmly established [4-6]. Traditional immuno-histochemical staining and fixed-tissue molecular/genetic/genomic techniques suffer from a poor "signal-to-noise" ratio due to tumor heterogeneity and only measure biomarkers at a single time-point without integrating the effect of the tumor microenvironment – leading to suboptimal predictive performance characteristics [7].

Existing cancer diagnostic tests have revealed prostate and breast cancers as clusters of heterogeneous cells and are classically difficult to diagnose and treat due to significant inter- and intra-cellular genetic heterogeneity and complex microenvironment interactions [8-10]. This makes it difficult to ascertain the specific genetic mutation(s) and physiological contexts responsible for aberrant cellular behavior and tumor aggressiveness – a necessary step to create a successful biomarker panel [11,12]. Further, individual cancer cells may evolve their own survival mechanisms and biochemical pathways, leading to specific, individual, "rogue," cells that may be highly aggressive in invasive and metastatic potential [13]. Further compounding the difficulty in risk stratifying cancer, the differences amongst individual and unique cells make bulk fixed-tissue-core or fixed-tissue-slice diagnosis of cancer via histochemical, genomic, or other molecular approaches problematic since it is not known, on a biopsy-sampling level basis which, or how many, cells within the biopsy are indeed aggressive and metastatic and, conversely, which cells are benign [14]. As such a live-cell phenotypic biomarker assay, with single-cell resolution, enabled with machine vision and machine learning for the risk stratification by adverse pathology of cancer patients (STRAT-AP) was developed and investigated via prostate and breast cancer primary tissue samples.

B. Biomarkers and biomarker ranking

The set of all measured biomarkers are listed in Supplementary Table 2. After primary biomarkers are measured via machine vision, aggregate biomarkers are derived from primary biomarkers via algebraic formulas detailed in Supplementary Table 3. As STRAT-AP is designed to probe cell signalling and cytoskeletal activity generated from cell-ECM interactions, focal adhesion size and actin retrograde flow velocity are heavily utilized when generating aggregate biomarkers. All listed biomarkers are necessary to train the machine learning algorithm to maximize the sensitivity, specificity for each prediction (groups of adverse pathology features and individual adverse pathology features). Of the ~600 biomarkers, biomarkers were selected by objective machine-based ranking during the training of the machine learning software via decision tree analysis. The top 10 rankings for local adverse pathology predictor (LAPP) and metastatic adverse pathology predictor (MAPP) values for prostate cancer are listed in

Supplementary Table 4. The biomarker rankings for LAPP and MAPP were defined by the machine learning algorithm based on maximizing sensitivity and specificity for a specific prediction.

Further, biomarker rankings are specific to a given prediction and the total set of ~600 biomarkers are necessary to achieve maximum sensitivity and specificity. Given the decision tree – based analysis and biomarker rankings, biomarkers cannot be replaced by other biomarkers. That is to say, it is necessary for each biomarker to be present and ranked, or weighted, in a specific manner to achieve the predictive performance of STRAT-AP. STRAT-AP applies the same set of ~600 biomarkers to both prostate and breast cancer samples. That said, analysis of prostate tissue and breast tissue yield different biomarker rankings after machine learning analysis. A representative comparison of biomarker rankings for prostate and breast cancer is detailed in Supplementary Table 4 & 5 respectively. Given the machine learning-derived biomarker rankings, the following biomarkers emerge as important for predicting LAPP and MAPP: Focal Adhesion intensity, cell height / adhesion / cell spreading as measured by mean square gray value (MSGV), cell perimeter, nucleus perimeter, and cell tortuosity. Further studies are planned to understand the individual importance of these biomarkers for given predictions.

C. Extracellular matrix formulation (ECMf) and media formulations

STRAT-AP was designed to minimize genetic and phenotypic alterations upon culturing primary cells *in vitro*. Toward that aim STRAT-AP was designed to enable rapid and short-term culturing of primary cells, promoting adhesion and analysis of biomarkers in vitro. That said, we recognize that in vitro culturing will inevitably alter certain genetic and phenotypic signatures of the primary cells analysed. As such, STRAT-AP's culturing methodology and technique was designed to provide a reference standard to promote a robust and reproducible response from primary cultured cells derived from primary tissue samples. In other words, the ECMf and media conditions of STRAT-AP's culturing conditions serves three functions: 1) to enable rapid adhesion on a substrate easy to analyse in vitro; 2) to promote short-term survival such that biomarkers can be measured within 72 hours; and 3) to elicit a specific cell-ECM response to probe cell-ECM interactions, cell signalling and cytoskeletal activation. STRAT-AP's success in measuring meaningful biomarkers is contingent on the ECMf's ability to enable adhesion, survival, and stimuli to the cell such that the machine learning algorithm can compare biomarker measurements effectively toward classifying and characterizing the heterogeneous population of cancer and non-cancer cells from a patient's primary tissue sample. Further studies are planned to further elucidate the role of the ECMf and media formulations in culturing and exhibiting specific biomarker characteristics.

D. Culture conditions and characteristics

Toward analyzing dynamic biomarkers over multiple time-points, the ability to establish live-primary cell cultures rapidly and analyze them in a meaningful way to generate informative biomarker measurements is paramount. To accomplish this, STRAT-AP utilizes standard culturing and physical chemistry techniques to engineer a discretely-defined environment with known media and extracellular matrix formulation conditions described in detail in Chander,

A.C., et al., 2017[15]. Such conditions were optimized to provide rapid (< 72 hours) and robust (> 80% adhesion and survival) culturing of primary cells. Standard methods of using classic media and serum formulations along with uncoated plastic culturing techniques were not successful in that they did not meet rapid and robust criteria though they may be sub-optimally permissive for culturing primary cells. Specifically, STRAT-AP's successful culturing techniques were based on media with defined growth and extracellular factors[15,16] and defined extracellular matrix protein formulation[15,16] used to coat glass culturing surfaces. Given that both prostate and breast cancer are thought to arise from epithelial tissue and that epithelial tissue, by definition, is composed of epithelial cells, STRAT-AP's ECMf formulation and media formulation was selected to be most permissive to epithelial cells. The distribution of cell types found in STRAT-AP's culturing conditions and image analysis conditions is presented in Supplementary Fig. 5. To optimize consistency of measurements and work-flow, culture conditions used for image analysis are identical to the culturing conditions.

As detailed in Chander, et al., 2017[8] the ECMf is critical for rapid and robust (>80%) adhesion and survival of primary prostate cells. Furthermore, the ECMf is critical to elicit a strong cell-ECM response from the cell such that biomarkers can be measured and analysed by machine vision and machine learning software, respectively. That said, given that the ECMf is reference standard, it is conceivable that other extracellular proteins may be utilized with varying degrees of accuracy and success if they satisfy the criteria for adhesion, survival and biomarker measurement. While the ECMf may be most critical and important for STRAT-AP's success, the media formulation is less critical in that ECMf may provide enough of an external signal to allow for adhesion, survival and biomarker measurement across numerous types of media. Future studies are planned to better understand the importance of the media and its role in selecting for different cell-types. As detailed in Chander et al., 2017[8,9] the critical components of STRAT-AP's ECMf are collagen type 1 and fibronectin. While future studies are planned to understand the critical component of the media, preliminary data supports the idea that the media is a less critical component of the culturing conditions though it may optimize the speed at which cells can be cultured, cell-type composition, and perhaps biomarker measurements. Alternative media formulations are discussed briefly in the Alternative methodology section of the Supplement. Indeed, the ECMf is the same as that is provided in Chander et al., 2017[8]. Future studies are planned to understand the optimal ratio of collagen & fibronectin. In this study we utilized a 1:1 ratio of Collagen to Fibronectin with a protein concentration great enough to saturate the surface of culturing form factor (i.e. microfluidic device).

E. Alternative methodology

With respect to alternative methodology, alternative reagents and software, albeit with sub-STRAT-AP efficiency, can be utilized to approximate the results of STRAT-AP. Specifically, commercial Dulbecco's Minimal Essential Media (DMEM) is sufficient in conjunction with the ECMf to promote adhesion and survival of primary prostate tissue with varying efficacy. Further studies are designed to understand the role the media plays in adhesion, survival, and biomarker measurements as well as cell-type composition. Data supporting the ability to culture cells (promote adhesion and survival) in a commercially available media are presented in the following. Specifically, data presented in Supplementary Fig. 10, 11, and 12 support the

utilization of DMEM, supplemented with 10 mM glutamine for transport, dissociation, and culturing as measured by cell adhesion in vitro on ECMf coated glass. Data supporting the ability to utilize 10 µg/ml of laminin as an alternative ECM is presented in Chander et al., 2017[8,9] and discussed in the following.

In this manuscript we present data gathered using custom reagents such as transport media, dissociation media, culturing media, and an extracellular matrix formulation (ECMf). Additionally, we present data that is collected by custom automated machine vision software and analyzed by custom machine learning software. In this section we describe how alternative, previously available commercial reagents and software can be utilized to collect similar results. Further, we present data that supports the use of commercially available medias and software to obtain similar data.

Toward the aim of successfully transporting tissue from the operating room to a central lab, a transport media is employed such that core-biopsy samples from surgical tissue can be stored during a transport time of <72 hours. Preliminary data support the ability to transport tissue with DMEM, supplemented with 10 mM glutamine while maintaining ~70% viability as measured by cell adhesion after dissociation. The use of DMEM with 10 mM glutamine exhibits a 10% decrease in viability when compared to Cellanyx's custom transport media[8]. Supplementary Fig. 10, 11, and 12 display data supporting the ability to transport tissue, dissociate tissue into cells, culture cells and measure biomarkers using commercially available DMEM supplemented with 10 mM glutamine.

Similar to tissue transport, DMEM with 10 mM glutamine may be employed to dissociate cells, albeit with less viability efficiency, as measured by cell adhesion, cell survival and biomarker measurement (Supplementary Fig. 10, 11, and 12). DMEM with 10 mM glutamine may also be employed to culture cells albeit with ~10% less efficiency than Cellanyx's custom media (Supplementary Fig. 10, 11, and 12).

As mentioned earlier, the ECMf is thought to contribute most to efficient culturing. In Chander et al., 2017[8], we present data comparing ECMf with laminin as an alternative protein coating to enable adhesion and survival of primary prostate cancer cells. Laminin was achieved ~60% adhesion and survival as compared to ECMf which achieved >80% adhesion and survival.

While data presented from this study was collected using automated MatLab machine vision software utilizing functions found in the 'Image Processing Toolbox,' machine vision software to quantify biomarkers can be also built using ImageJ (NIH) utilizing the following functions: 'find edges', 'optimum threshold', 'fill holes', and 'remove noise'. Supplementary Fig. 12 demonstrated biomarker measurements quantified using manual (open source) ImageJ techniques. Further, manual machine vision software to quantify biomarkers can also be employed using MatLab utilizing the functions in the 'Image Processing Toolbox'.

In addition to primary biomarkers (Supplementary Table 2) aggregate biomarker were utilized to highlight and quantify the interdependence, synergistic and antagonistic relationships, of primary biomarkers. Supplementary Table 3 lists the aggregate biomarkers that were used. Both primary and aggregate biomarkers were initially presented in patent[16].

While less efficient with respect to time and performance, commercial methods and reagents may be utilized to obtain similar results and analysis presented in this manuscript.

Machine learning methodology were initially presented in the patent. The following is a description of how a machine learning pipeline in Matlab can be built:

Capture raw data

Using a microscope's automation software, capture images of the sample cells. There are three collection regimes: cell spreading, cell tracking, and focal adhesion.

Cell spreading is the first collection regime. It is the time immediately after the cells are first introduced to the microfluidic environment. The images are taken in quick succession to capture the fast movement of cells adhering to the ECM and spreading.

Cell tracking is the next regime. As the cells have adhered and interact with the ECM, the cells are tracked over a period of hours, with images taken at many different locations to capture as many cells as possible, but ensuring constant time differences between each subsequent capture at a specific location. The result is images of many cells over the period of hours.

The final regime is after cells are stained to image focal adhesions. They are observed through the proper filter cube and imaged to obtain images of the focal adhesions.

Get data for each cell for each time point

Metrics are calculated for every time point for every cell during the cell spreading and cell tracking regimes. To calculate these metrics, first segment the cells. It is necessary to identify cells that are touching. Some cells move between time points. We want to track these cells across frames of data. To do this, for each cell, calculate the center of mass. In the next time point, look for the center of mass that is closest to this one, and identify those as the same cell. For each cell, also segment the nucleus.

Segmentation can be achieved in Matlab using a combination of commands like 'multithresh,' 'bwdist,' imhmin, 'watershed,' and 'bwmorph,' all in the image processing toolbox.

For each cell and nucleus, calculate the parameters: area, perimeter, tortuosity, mean grey scale value. These can be calculated in Matlab using the 'regionprops' command in the image processing toolbox.

For cells from the cell spreading regime, calculate the change in the area to calculate the spreading velocity.

For cells in the cell tracking regime, calculate the movement of the center of mass of the cell between frames to get the migration velocity.

The output from this is a list of directly measured metrics for each cell for each time point, these are called the primary biomarkers. Aggregate biomarkers are calculated from combinations of the primary biomarkers.

Measure actin retrograde flow

Retrograde flow is the movement of actin towards the center of a cell at the leading edge of its leading edge connection with the ECM. The retrograde flow is measured from the cell tracking regime. In this regime, radial samples are taken from the center of mass of the cell to the perimeter at periodic angles. These line samples are lined up for all time points to generate a kymograph. From this kymograph, the identification and measurement of retrograde flow can be calculated.

Measure focal adhesions

Observing the images taken in the focal adhesion regime, focal adhesions can be segmented, sized, counted, and locations measured using matlab functions like multithresh, bwmorph, and regionprops.

Aggregate data for all time points

To aggregate data from all time points, to produce a single metric per cell, several statistical operations are used. The mean, median, maximum, minimum, mean of top quartile, mean of bottom quartile of all metrics are taken. Also, the inverse of these values is taken. This results in 12 values for each metric for each cell. Reject cells that were not able to be tracked for all time points.

Train classifier

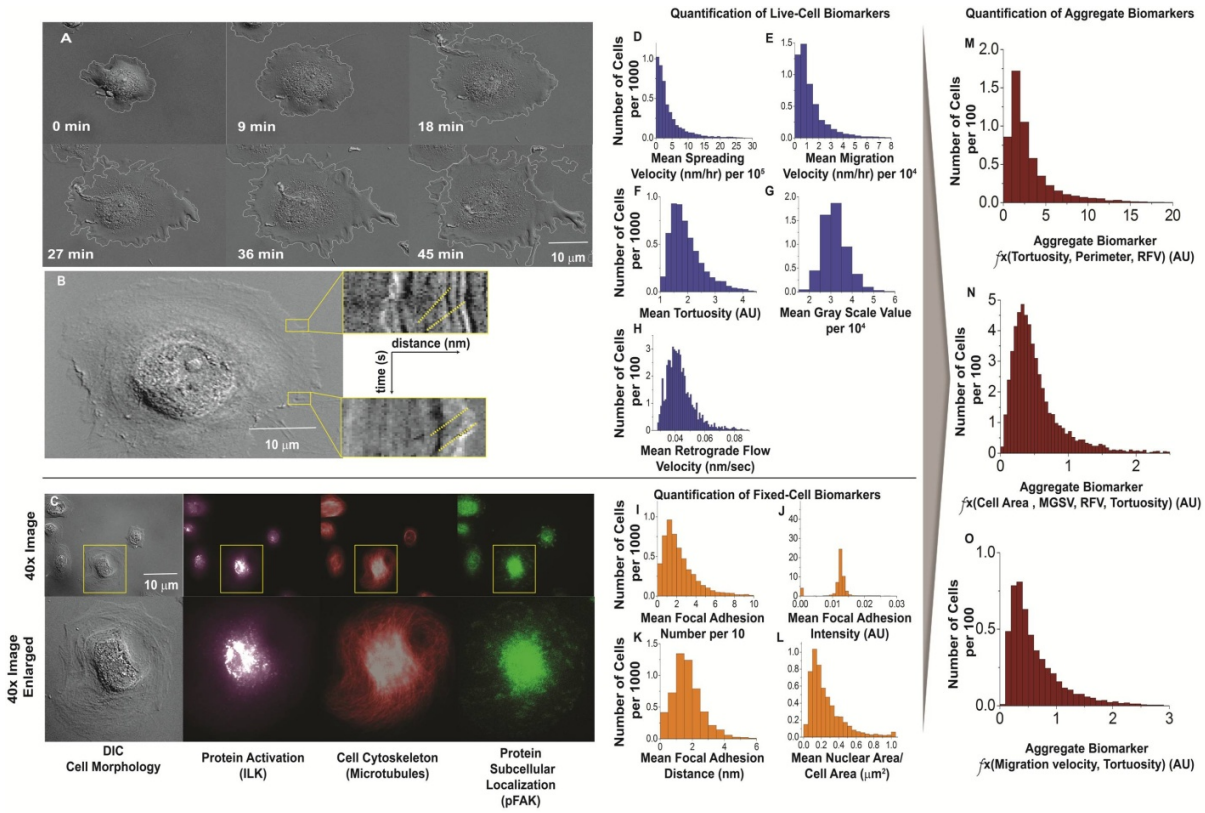
With all the metrics for each cell calculated, they can be used to train and validate a machine learning classifier. Each cell is treated as independent observations. As a result, the pathology output for each patient from which the cell came was used as the ground truth for each of the cells.

The cells are separated into independent training and validation sets using a standard 70%/30% training/validation split. The cell metrics and output pathologies are used to train a random forest model.

In matlab, this can be done using the TreeBagger function.

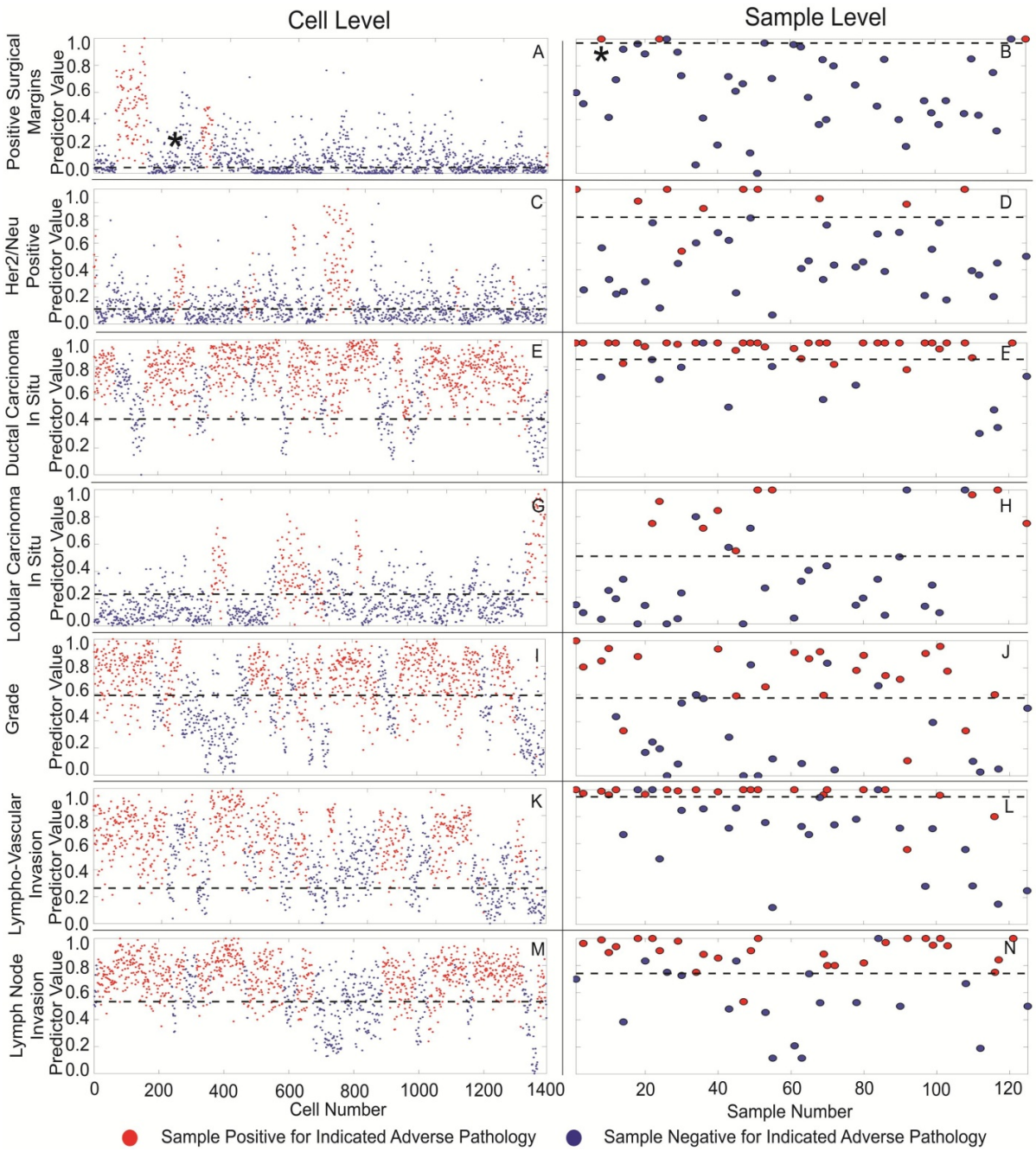
Validate classifier

Take the validation set that was held apart from training the classifier and use that cell data to predict the output. Combine cell-by-cell results into a sample-level response by taking the average classifier output. Compare this predictor output to the ground truth of the actual pathology report for the cell that the patients came from. Calculate performance metrics.



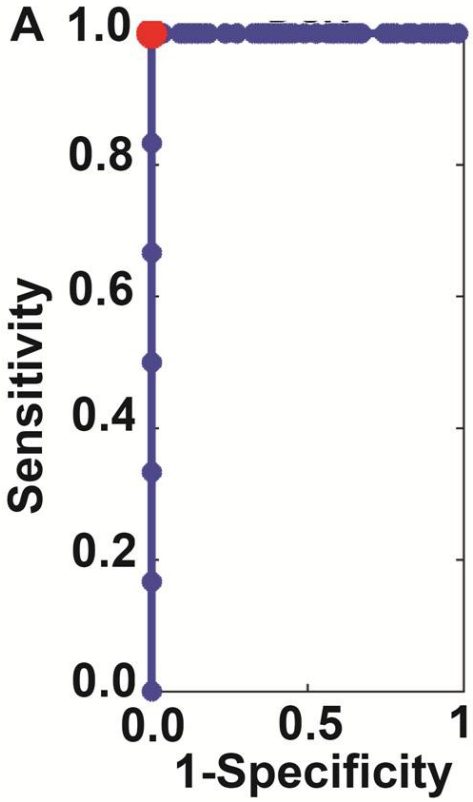
Supplementary Fig. 1: STRAT-AP's breast cancer phenotypic (cellular and molecular) biomarkers are measured via sequential live-cell imaging and fixed-cell imaging in a standardized microfluidic environment and quantified using automated machine vision software to obtain live- and fixed-cell biomarkers with single-cell resolution. Example breast cancer biomarkers measured include A) Cell spreading + tortuosity: cell adhesion rate to device substrate, cell area change during adhesion, tortousity of the cell membrane as a measure of morphology B) Membrane fluctuations: rapid dynamics of the membrane surface are measured as retrograde flow through kymographs C) expression, localization, and phosphorylation state of subcellular protein complexes (phospho-focal adhesion kinase (pFAK)) and individual proteins (integrin-linked kinase (ILK)) as well as microtubules are measured on corresponding fixed cells matched to the live cells 20x DIC and 40x fluorescence images were measured via a standard automated fluorescent microscope. D-H) Live-cell biomarker examples quantified. D) Quantitative measure of mean cell spreading velocity for the total population of cells. E) Quantitative measure of mean cell migration velocity for the total population of cells. F) Quantitative measure of mean cell tortuosity for the total population of cells. G) Quantitative measure of mean cell mean gray scale value for the total population of cells. H) Quantitative measure of cell mean retrograde flow velocity for the total population of cells. I-L) Fixed cell biomarker examples quantified. I) Quantitative measure of mean cell focal adhesion number for the total population of cells. J) Quantitative measure of mean cell focal adhesion intensity for the total population of cells. K) Quantitative measure of mean cell focal adhesion distance from

membrane edge for the total population of cells. L) Quantitative measure of mean cell nuclear area/cell area for the total population of cells. M) Quantitative measure of LAPP2 (Tortuosity, Perimeter, RFV) aggregate biomarker for the total population of cells. N) Quantitative measure of MAPP10 (Area, MGSV, RFV, Tortuosity) aggregate biomarker for the total population of cells. O) Quantitative measure of MAPP17 (Migration velocity, Tortuosity) aggregate biomarker for the total population of cells.



Supplementary Fig. 2: STRAT-AP's breast cancer machine learning algorithms trained to predict specific surgical adverse pathology features classify positive and negative single-cells and predict patient surgical adverse pathology features in blinded, test sample sets based on machine learning-derived thresholds. (A) A 'cell-level' plot stratifying negative (blue circles) and positive (red circles) cells for positive surgical margins (PSM). Black dashed line indicate the machine-derived threshold. (B) A 'patient-level' plot stratifying positive (red circles) and negative (blue circles) for patients positive or negative for PSM. Results are obtained by summarizing cell level results. Black dashed line indicates machine-derived threshold set by

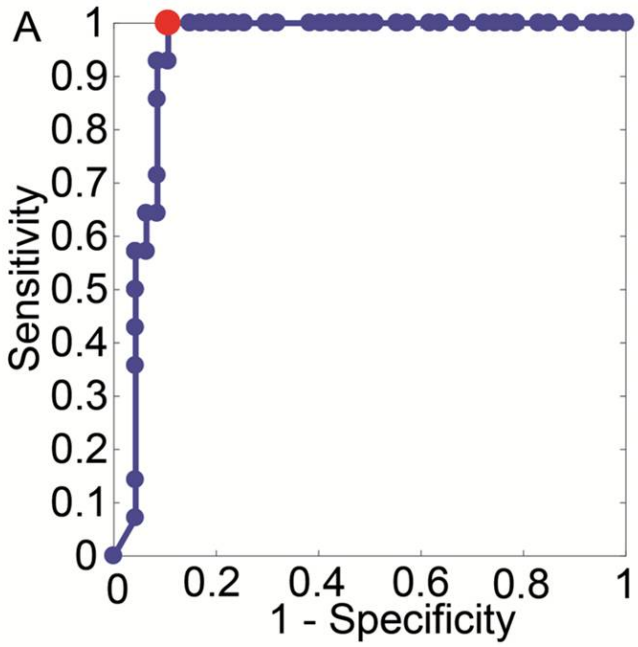
algorithmic analysis. (C) Cell-level plot stratifying negative and positive cells for Her2/Neu positive (H/NP). (D) Patient-level plot stratifying for patients positive or negative for H/NP results. (E) Cell-level plot stratifying negative and positive cells for ductal carcinoma in situ (DCIS). (F) Patient-level plot stratifying for patients positive or negative for DCIS. (G) Cell-level plot stratifying negative and positive cells for lobular carcinoma in situ (LCIS). (H) Patient-level plot stratifying for patients positive or negative for LCIS. (I) Cell-level plot stratifying negative and positive cells for grade (Gr). (J) Patient-level plot stratifying for patients positive or negative for Gr. (K) Cell-level plot stratifying negative and positive cells for lympho-vascular invasion (LVI). (L) Patient-level plot stratifying for patients positive or negative for LVI results. (M) Cell-level plot stratifying negative and positive cells for lymph node invasion (LI). (N) Patient-level plot stratifying for patients positive or negative for LI.



B **Prediction of Biochemical Recurrence Performance Results from Prostate Tissue Sample Study**

Predictive Metrics	Sensitivity	Specificity	AUC	N	Number Positive	Number Negative
Biochemical Recurrence (BCR)	0.99	0.99	0.99	61	6	55

Supplementary Fig. 3: A Receiver Operating Characteristic (ROC) curve was calculated for Biochemical Recurrence (BCR) to quantify how well the algorithmically derived metrics could predict disease recurrence. (A) The optimal sensitivity and specificity was calculated based on ROC curve analysis. (B) The platform is able to predict BCR with high sensitivity and specificity.



B **Prediction of Upgrading Between Biopsy Gleason Score and RP Gleason Score Performance Results from Prostate Tissue Sample Study**

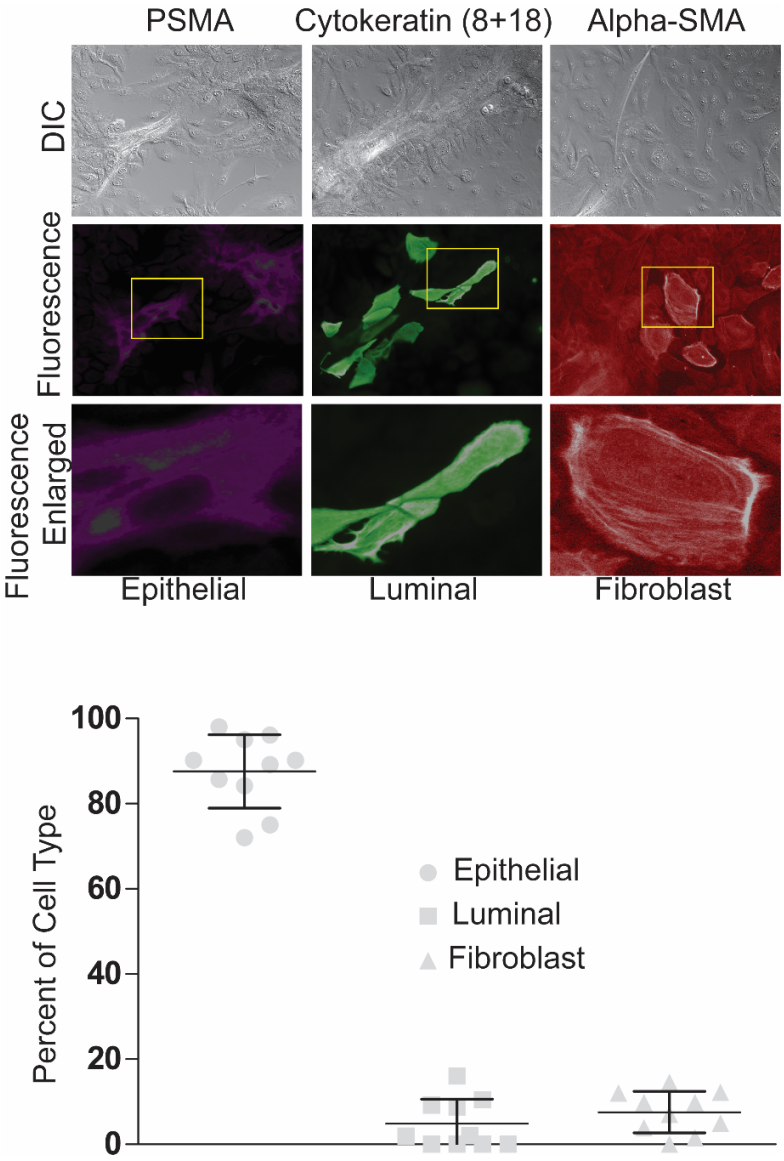
Predictive Metrics	Sensitivity	Specificity	AUC	N	Number Positive	Number Negative
Upgrading	0.99	0.90	0.94	61	14	47

Supplementary Fig. 4: A Receiver Operating Characteristic (ROC) curve was calculated for upgrading between biopsy Gleason score and RP Gleason score to quantify how well machine learning-derived metrics could predict RP Gleason. (A) ROC curve analysis. (B) The platform is able to predict BCR with high sensitivity and specificity.

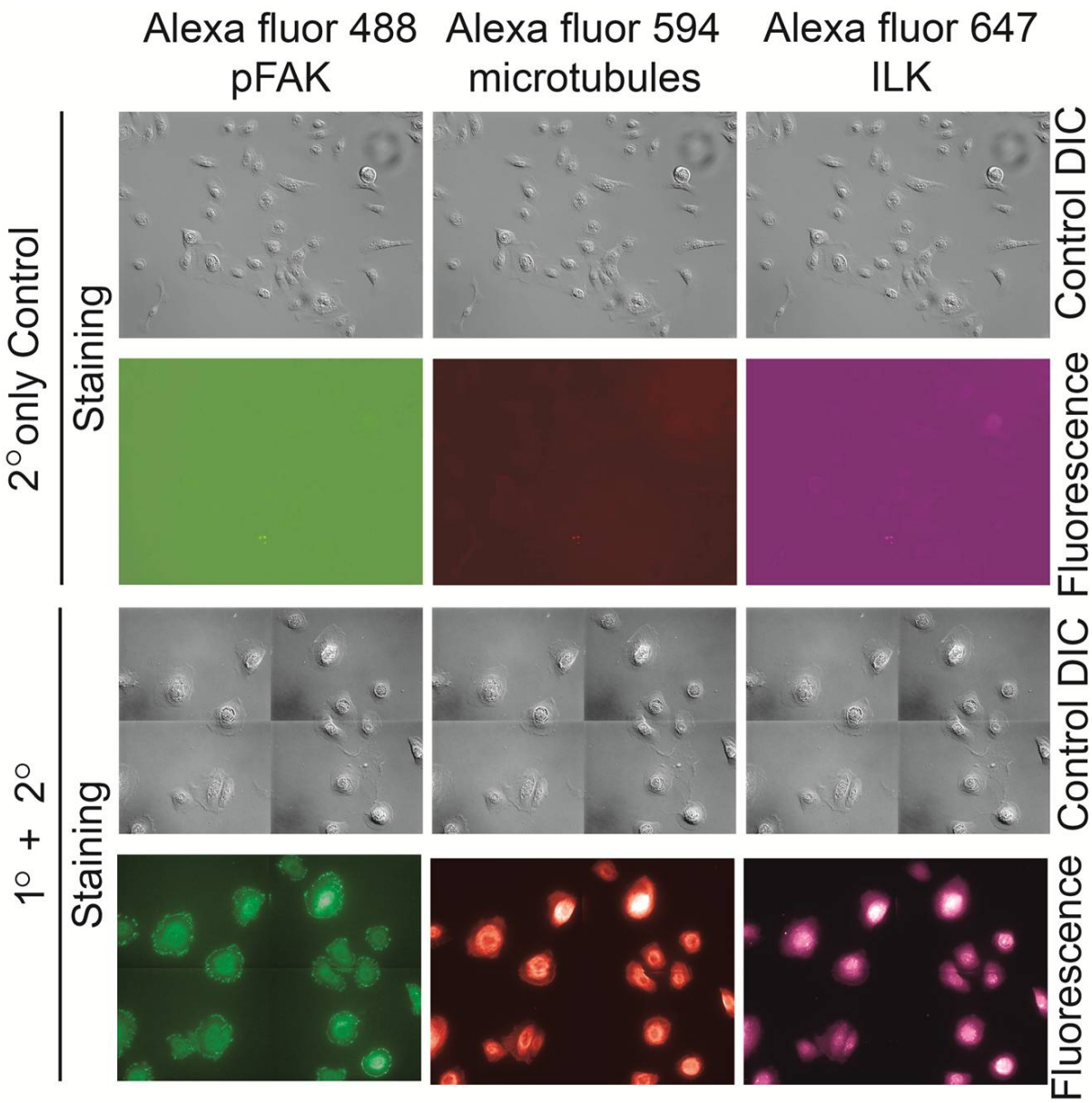
Supplementary Table 1: Summary of adverse pathology data for prostate cancer patients positive for Biochemical Recurrence.

BCR Positive Sample Adverse Pathology Data						
	Sample Number					
Adverse Pathology	1	2	3	4	5	6
Positive surgical margin	-	+	+	+	-	-
Seminal vesicle invasion	-	+	+	ND	-	-
Extra-prostatic extension	+	+	+	+	+	-
Perineural invasion	ND	+	+	+	+	+
Lymph node positive	-	ND	ND	ND	-	+
Lymph vascular invasion	-	ND	+	ND	-	-

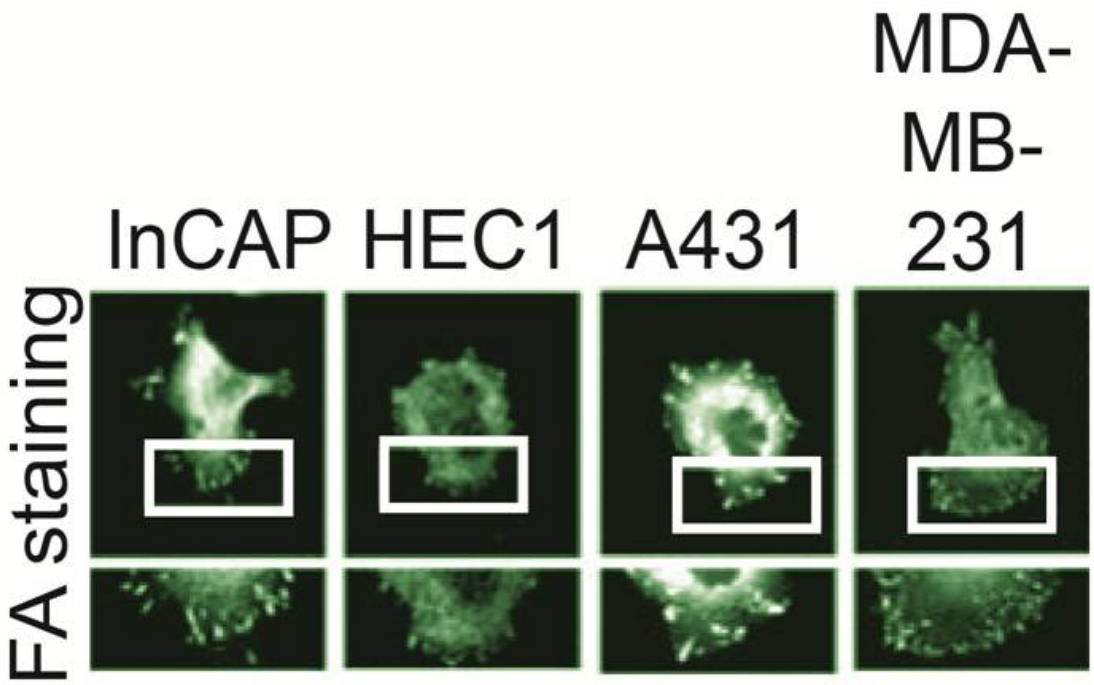
Supplementary Table 1: Samples Positive for BCR and Their Associated Adverse Pathologies: (+) positive for indicated pathology prediction, (-) negative for indicated pathology prediction, ND for that sample adverse pathology was available from the path report.



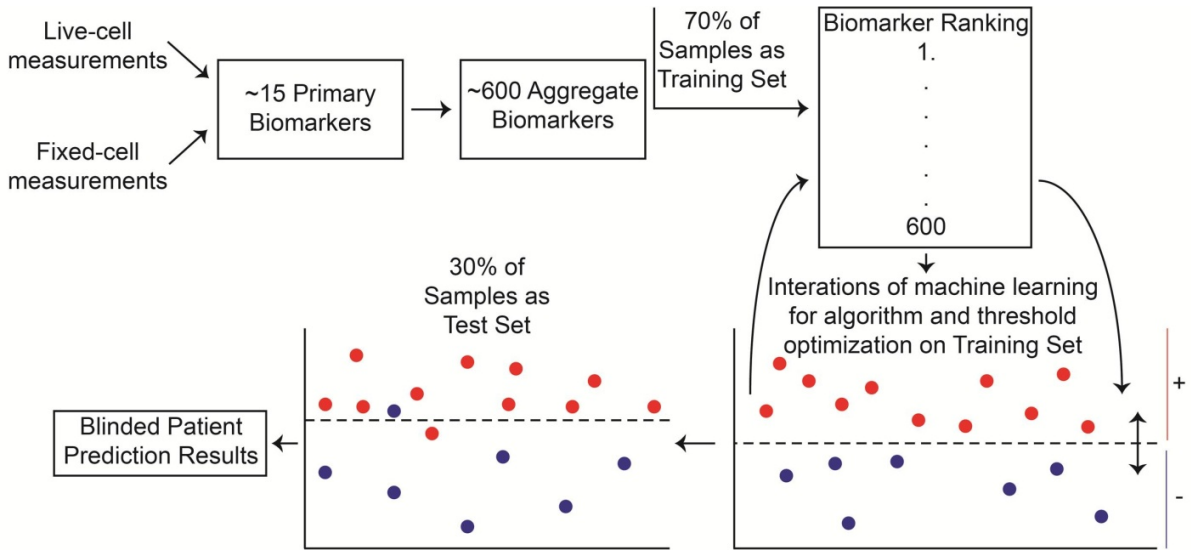
Supplementary Fig. 5: Various cell types are cultured in the STRAT-AP platform. Fresh primary prostate cells were seeded onto ECM coated surfaces incubated overnight at 37°C with 5% CO_2 The cells were stained with prostate specific membrane antigen (PSMA) as an epithelial cell marker, Cytokeratin 8 + 18 as a luminal epithelial cell marker, and alpha-smooth muscle actin as a fibroblast marker. Controls were treated with secondary antibody only and treated with the same conditions as the primary antibodies. All images are at 20x magnification. Graph of percentage of each type of cell found in the culture system. 500 cells were counted as either epithelial, luminal, or fibroblasts based on staining patterns for cultures derived from RP tissue samples (n=3 samples). Then the percentage of each cell type was calculated. Graph shows the percentage of each cell type.



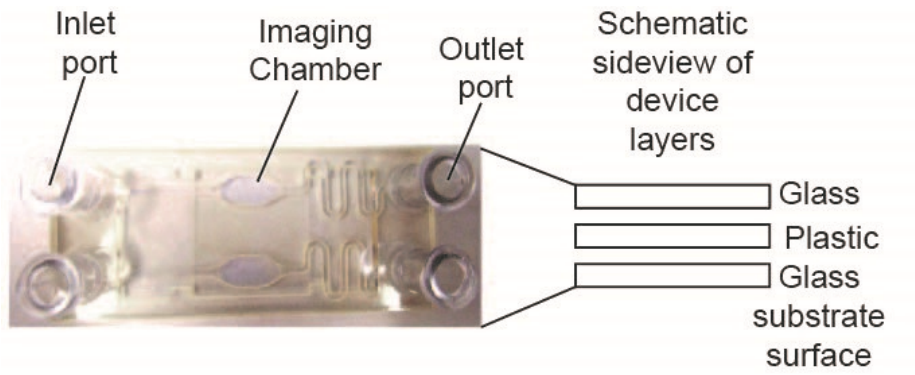
Supplementary Fig. 6: Fluorescence staining of molecular biomarkers and corresponding staining controls. Each sample was treated with either secondary antibodies only as control for background and non-specific staining (upper two image panels) or corresponding primary antibodies and secondary antibodies (lower two image panels).



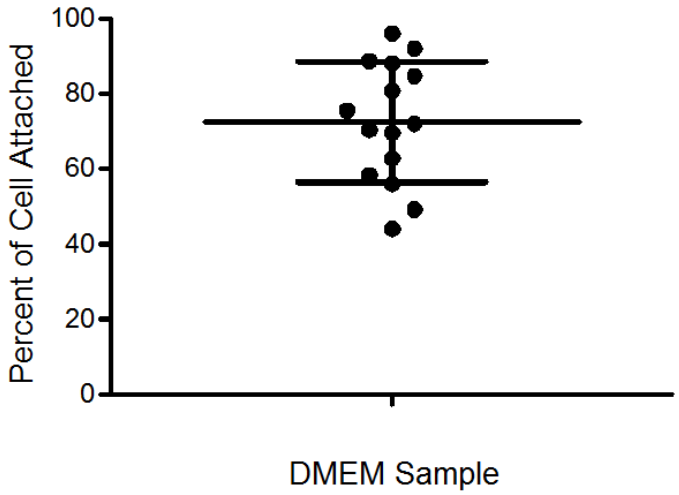
Supplementary Fig. 7: Fluorescence staining of molecular biomarkers in cancer cell lines. Each cell line was treated with focal adhesion staining primary antibodies and corresponding secondary antibodies. lnCAP = human prostate cancer cell line,HEC1 = human endometrial cancer cell line, A431 = epidermoid carcinoma cell line, MDA-MB-231 = breast cancer cell line



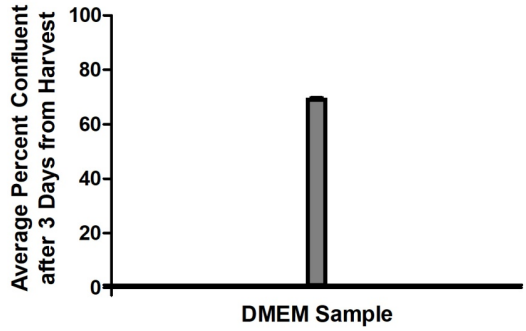
Supplementary Fig. 8: Flow diagram of machine learning sample training, classification and blinded predictions. Live-cell and fixed-cell primary biomarkers and measured by machine vision software at single cell resolution. Machine learning algorithms are trained using 70% of the cells across all samples associating biomarkers with corresponding adverse pathology reports where several iterations are run to optimize cell and sample level classification and train for biomarker ranking based on accuracy of adverse pathology prediction. Once optimal biomarker rankings are established for the highest accuracy of predictions at the cell and sample level the remaining 30% of cells across samples are analyzed in a blinded fashion and classified at the cell level and sample level for their likelihood of having a particular adverse pathology with the pathology reports remaining blinded to the system. After the prediction is made at the cell and sample level based on the machine vision measured biomarkers and machine learning analysis predictions are checked against the unblinded pathology reports to test prediction performance. ROC curve analysis and AUC calculations are made as a read out of prediction performance and accuracy.



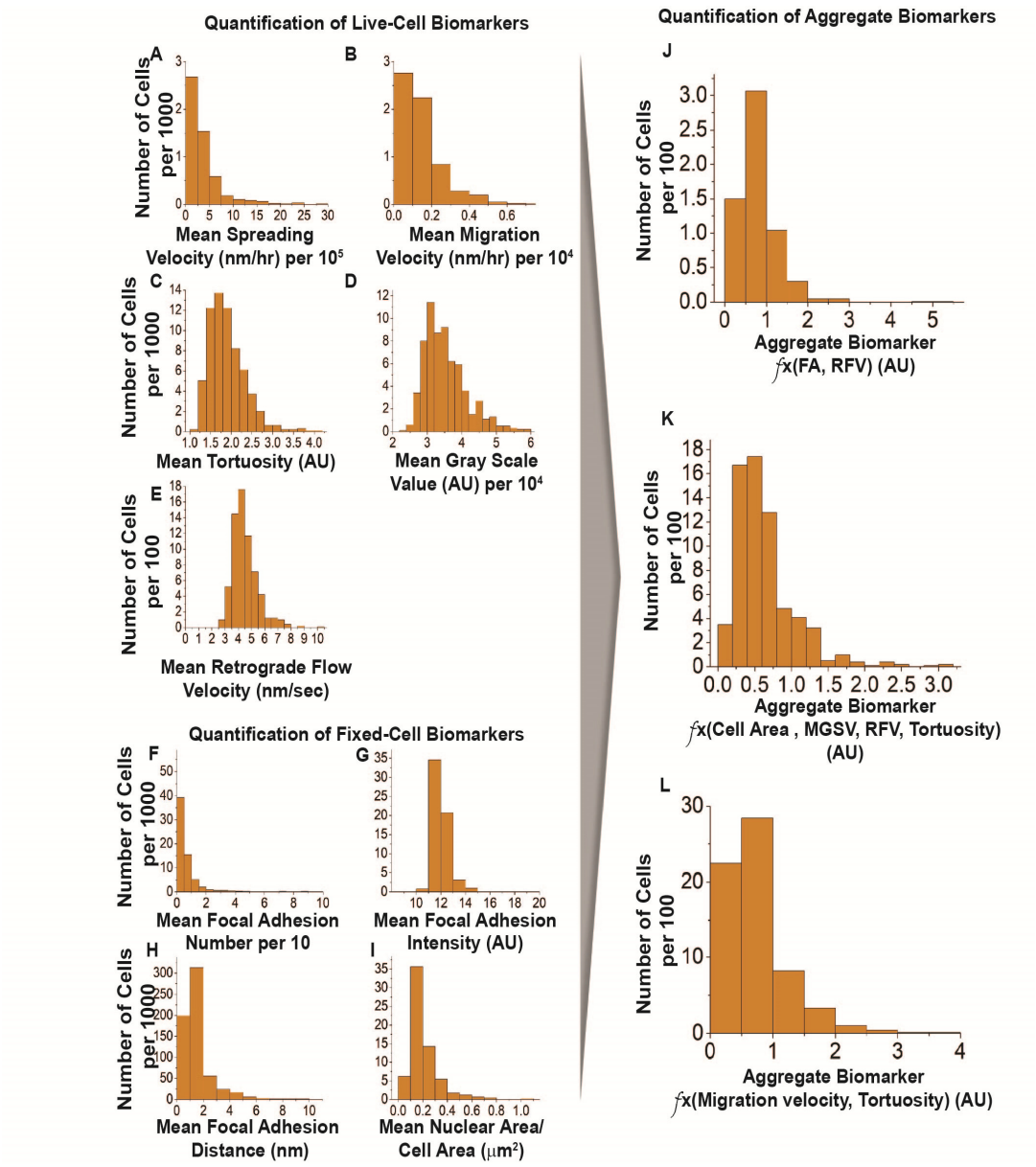
Supplementary Fig. 9: **STRAT-AP Microfluidic Device.** Image of the top of STRAT-AP's microfluidic device and schematic drawing of the glass, plastic, glass layers of STRAT-AP's microfluidic device.



Supplementary Fig. 10: Immediate cell attachment (viability) after transport and dissociation in alternative media: DMEM with 10 mM glutamine. The percentage number of cells were counted that attached to the surface of the microfluidic device upon initial seeding relative to the total possible based on seeding density (n= 9).



Supplementary Fig. 11: Cell survival after 3 days in culture using alternative media: DMEM with 10 mM glutamine. Percent confluence of cells after 3 days in culture (n=3).



Supplementary Fig. 12: Biomarker measurement from cells cultured and imaged in alternative media: DMEM with 10 mM glutamine and via alternative machine vision software: manual ImageJ-based biomarker quantification. Histograms of biomarker measurements collected from cells exposed to DMEM + 10 mM glutamine media demonstrate biomarkers can be measured in alternative media and alternative machine vision techniques.

Primary Biomarkers
Migration velocity
Cell area
Cell perimeter
Cell tortuosity
Cell aspect ratio
Nuclear area
Nuclear perimeter
Nuclear tortuosity
Cell mean gray square value
Actin retrograde flow velocity
Focal adhesion number
Focal adhesion intensity
Focal adhesion distance from cell edge
Cell spreading velocity

Supplementary Table 2: **List of primary biomarkers measured using STRAT-AP.**

Mathematical Functions to Compute Aggregate Biomarkers
OP1 = Tortuosity / RFV
OP2 = Tortuosity * Perimeter / RFV
OP3 = Area * RFV / Tortuosity
OP4 = FA Size / RFV
MP2 = OP3 * Migration Velocity
P4 = Area / RFV
P5 = RFV / Area
P6 = FA Size / Area
P7 = Area / FA Size
P8 = Area / (RFV * Tortuosity)
P9 = RFV * Tortuosity / Area
P10 = Area * MGSV / (RFV * Tortuosity)
P11 = Area / (RFV * Tortuosity * MGSV)
P12 = FA Size * Tortuosity / Area
P13 = Area * FA Size / Tortuosity
P14 = Area / Migration Velocity
P15 = FA Size / Tortuosity
P16 = Migration Velocity * Tortuosity
P17 = Migration Velocity / Tortuosity
P18 = Tortuosity / FA Size
P19 = Area * Migration Velocity
PAC1 = FAS^2 * Spreading Velocity
PAC2 = FAS^2 / Spreading Velocity
PAC3 = (FAS / Migration Velocity) * CA
PAC4 = (CA * FAS * RFS) / Migration Velocity
PAC5 = (Spreading Velocity) / (Migration Velocity)
PAC6 = (CA * FAS (Migration Velocity)) / RFS

Supplementary Table 3: List of aggregate biomarkers and mathematical formulas utilized to derive aggregate biomarkers measured using STRAT-AP. After all primary biomarkers are collected and aggregate biomarkers the following statistical functions are applied to each biomarker measurement:

1. All timepoints
    a. Median
    b. Mean
    c. Standard deviation
2. Top quartile of values
    a. Median
    b. Mean
    c. Standard deviation

3. Bottom quartile of values
    a. Median
    b. Mean
    c. Standard deviation

After statistical functions are performed ~600 different biomarker measurements are generated that are then input into the machine learning training algorithm.

	Prostate Tissue Biomarker Ranking for Local Adverse Pathology (LAPP) Prediction
1	MGSV median
2	MGSV median inverse
3	bottom quartile FA intensity mean inverse
4	MGSV mean inverse
5	MGSV stdev
6	MGSV stdev inverse
7	bottom quartile MGSV median inverse
8	bottom quartile FA intensity mean
9	FA intensity mean inverse
10	FA intensity median

	Prostate Tissue Biomarker Ranking for Metastatic Adverse Pathology (MAPP) Prediction
1	PAC4 bottom quartile mean
2	PAC3 mean inverse
3	P18 bottom quartile mean inverse
4	P15 stdev inverse
5	P13 median
6	OP3 top quartile mean
7	P16 bottom quartile mean inverse
8	bottom Cell Perimeter median
9	Nucleus Perimeter stdev
10	PAC6 top quartile median

Supplementary Table 4: Biomarker rankings determined by objective machine learning algorithms are presented and demonstrate that different biomarkers are utilized for different predictions within prostate tissue analysis.

Breast Tissue Biomarker Ranking for Local Adverse Pathology (LAPP) Prediction	
1	MGSV median
2	MGSV median inverse
3	bottom quartile MGSV median inverse
4	top quartile MGSV std
5	bottom quartile Cell Tortuosity mean
6	Cell Tortuosity median
7	top quartile FA mean
8	top quartile MGSV median inverse
9	top quartile MGSV median
10	PAC2 top quartile mean

Breast Tissue Biomarker Ranking for Metastatic Adverse Pathology (MAPP) Prediction	
1	bottom quartile FA Intensity median
2	bottom quartile FA Intensity mean inverse
3	top quartile FA Intensity median inverse
4	FA Intensity mean inverse
5	FA Intensity median inverse
6	FA Intensity median
7	bottom quartile FA Intensity median inverse
8	top quartile FA Intensity mean inverse
9	bottom quartile FA intensity mean
10	top quartile FA intensity mean

Supplementary Table 5: Biomarker rankings determined by objective machine learning algorithms are presented and demonstrate that different biomarkers are utilized for different predictions within breast tissue analysis.

Supplementary References:

1 Alvarado, M., Ozanne, E. & Esserman, L. Overdiagnosis and overtreatment of breast cancer. *Am Soc Clin Oncol Educ Book*, e40-45, doi:10.14694/EdBook_AM.2012.32.e40 (2012).

2 Evans, A. & Vinnicombe, S. Overdiagnosis in breast imaging. *Breast*, doi:10.1016/j.breast.2016.10.011 (2016).

3 Virnig, B. A., Tuttle, T. M., Shamliyan, T. & Kane, R. L. Ductal carcinoma in situ of the breast: a systematic review of incidence, treatment, and outcomes. *J Natl Cancer Inst* **102**, 170-178, doi:10.1093/jnci/djp482 (2010).

4 Loeb, S. *et al.* Overdiagnosis and overtreatment of prostate cancer. *Eur Urol* **65**, 1046-1055, doi:10.1016/j.eururo.2013.12.062 (2014).

5 Mukhtar, R. A., Wong, J. M. & Esserman, L. J. Preventing Overdiagnosis and Overtreatment: Just the Next Step in the Evolution of Breast Cancer Care. *J Natl Compr Canc Netw* **13**, 737-743 (2015).

6 Shieh, Y. *et al.* Population-based screening for cancer: hope and hype. *Nat Rev Clin Oncol* **13**, 550-565, doi:10.1038/nrclinonc.2016.50 (2016).

7 Marusyk, A. & Polyak, K. Tumor heterogeneity: causes and consequences. *Biochim Biophys Acta* **1805**, 105-117, doi:10.1016/j.bbcan.2009.11.002 (2010).

8 Albini, A. & Sporn, M. B. The tumour microenvironment as a target for chemoprevention. *Nat Rev Cancer* **7**, 139-147, doi:10.1038/nrc2067 (2007).

9 Barber, L. J., Davies, M. N. & Gerlinger, M. Dissecting cancer evolution at the macro-heterogeneity and micro-heterogeneity scale. *Curr Opin Genet Dev* **30**, 1-6, doi:10.1016/j.gde.2014.12.001 (2015).

10 Koren, S. & Bentires-Alj, M. Breast Tumor Heterogeneity: Source of Fitness, Hurdle for Therapy. *Mol Cell* **60**, 537-546, doi:10.1016/j.molcel.2015.10.031 (2015).

11 Burrell, R. A., McGranahan, N., Bartek, J. & Swanton, C. The causes and consequences of genetic heterogeneity in cancer evolution. *Nature* **501**, 338-345, doi:10.1038/nature12625 (2013).

12 Venkatesan, S. & Swanton, C. Tumor Evolutionary Principles: How Intratumor Heterogeneity Influences Cancer Treatment and Outcome. *Am Soc Clin Oncol Educ Book* **35**, e141-149, doi:10.14694/EDBK_158930 (2016).

13 Roth, A. *et al.* Clonal genotype and population structure inference from single-cell tumor sequencing. *Nat Methods* **13**, 573-576, doi:10.1038/nmeth.3867 (2016).

14 Spratt, D. E., Zumsteg, Z. S., Feng, F. Y. & Tomlins, S. A. Translational and clinical implications of the genetic landscape of prostate cancer. *Nat Rev Clin Oncol*, doi:10.1038/nrclinonc.2016.76 (2016).

15 Chander, A., Manak, M., Varsanik, J., Hogan, B., Mouraviev, V., Zappala, S., Sant, G., Albala, D. . Rapid and short-term extra-cellular matrix-mediated in vitro culturing of tumor and non-tumor human primary prostate cells from fresh radical prostatectomy tissue. *Urology* (2017).

16 Chander, A. C. Systems, methods and devices for measuring growth/oncogenic and migration/metastatic potential United States of America patent (2011).