Detection of CTC clusters and a dedifferentiated RNA-expression survival signature in prostate cancer

Molly Kozminsky, Shamileh Fouladdel, Jae-Seung Chung, Yugang Wang, David C. Smith, Ajjai Alva, Ebrahim Azizi, Todd Morgan*, Sunitha Nagrath*

Dr. M. Kozminsky, Prof. S. Nagrath

Department of Chemical Engineering, University of Michigan

2300 Hayward Street, Ann Arbor, MI 48109, USA

Dr. M. Kozminsky, S. Fouladdel, Dr. E. Azizi, Prof. S. Nagrath

Translational Oncology Program, University of Michigan

1600 Huron Pkwy, Ann Arbor, MI 48109, USA

Dr. M. Kozminsky, S. Fouladdel, Dr. E. Azizi, Prof. S. Nagrath

Biointerfaces Institute, University of Michigan

2800 Plymouth Road, Ann Arbor, MI 48109, USA

S. Fouladdel, Dr. E. Azizi

Department of Internal Medicine, University of Michigan

1500 E. Medical Center Drive, Ann Arbor, MI 48109, USA

Dr. J. Chung, Dr. Y. Wang, Dr. T. Morgan

Department of Urology, University of Michigan

1500 E. Medical Center Drive, Ann Arbor, MI 48109, USA

Dr. D.C. Smith, Dr. A. Alva

Department of Internal Medicine, Division of Hematology/Oncology, University of Michigan

1500 E. Medical Center Drive, Ann Arbor, MI 48109, USA

E-mail: snagrath@umich.edu, tomorgan@med.umich.edu

Abstract

Rates of progression and treatment response in advanced prostate cancer are highly variable, necessitating non-invasive methods to assess the molecular characteristics of these tumors in real time. The unique potential of circulating tumor cells (CTCs) to serve as a clinically useful liquid biomarker is due to their ability to inform via both enumeration and RNA expression. A microfluidic graphene oxide based device (GO Chip) is used to isolate CTCs and CTC clusters from the whole blood of 41 men with metastatic castration-resistant prostate cancer. Additionally, the expression of 96 genes of interest is determined by RT-qPCR. Multivariate analyses are conducted to determine the genes most closely associated with overall survival, PSA progression, and radioclinical progression. A preliminary signature, comprised of high expression of stemness genes and low expression of epithelial and mesenchymal genes, potentially implicates an undifferentiated CTC phenotype as a marker of poor prognosis in this setting.

Main Text

While men with metastatic castration resistant prostate cancer (mCRPC) have a median survival of approximately 18 months, there is substantial heterogeneity and time-to-progression varies widely.[1] Additionally, given the evolving treatment landscape, there is a clear need for better biomarkers of progression and treatment response in order to help guide therapeutic decisions. While soft tissue and bone biopsies can provide molecular information, many men with mCRPC have already undergone multiple prior invasive biopsies, and the tissue-based information is representative of only that single disease site.

Circulating tumor cells (CTCs) shed from tumors can be detected in the blood stream[2] and have the potential to serve as a liquid biopsy. This approach offers the potential for repeated, non-invasive measurements and may more widely sample the overall disease state. In addition to giving insight into the burden of disease, these cells can relate the overall molecular state and risk of progression through the analysis of gene expression and phenotype of the traveling cells.[3] However, key obstacles to capturing CTCs include their rarity among the millions of surrounding white blood cells and red blood cells.[4] To best interrogate CTCs, they must be detected with high yield and sufficient purity.

This problem has been addressed with a host of isolation technologies.[5] Notably, the first FDA approved CTC isolation technology, CellSearch, has been used to establish survival differences based on CTC enumeration.[6] This macroscale technology uses a magnetic ferrofluid conjugated with an antibody against the epithelial cellular adhesion molecule (EpCAM) to capture EpCAM-expressing cells from 7.5 mL whole blood.[7] For increased sensitivity and flexibility of downstream analysis, microfluidics and nanomaterials[8] have been developed to isolate and study CTCs.[9-12]

Prior studies have used cell enrichment technologies coupled with RT-qPCR to study prostate CTCs and their RNA. For example, mCRPC patients with AR-V7-positive CTCs displayed abiraterone and enzalutamide resistance, indicating the potential for liquid biopsy approaches to provide predictive information.[14] The original microfluidic CTC-Chip was used to study prostate CTCs in localized and metastatic patients, with CTCs detected in 23/36 metastatic patients.[9] RT-qPCR was used to detect the TMPRSS2-ERG fusion in 9/20 metastatic patients. Next-generation CTC chips such as the Herringbone (HB) Chip[10] and the geometrically enhanced differential immunocapture (GEDI) chip[15] have also been applied to prostate cancer. The HB Chip was used to investigate androgen receptor (AR) signaling through immunofluorescence staining for the prostate specific antigen (PSA) and the prostate specific membrane antigen (PSMA),[16] while further immunofluorescence characterization by the GEDI chip examined ERG expression.[17] Recently the CTC-iChip was used to reveal the role of noncanonical Wnt signaling through single-cell RNA-Seq of prostate CTCs from 13 patients isolated by negative selection.[11] However, the majority of these studies reported on a limited set of genes.

The nanomaterial-based graphene oxide chip (GO Chip) affords highly sensitive and gentle capture of rare cells with low white blood cell contamination.[18] Optimized with cell line spike-in samples with as few as 3-5 cancer cells per milliliter of whole blood, the device showed promise in the capture of PC-3 cells under physiologically relevant conditions and concentrations.[18] Coupled with the capability for downstream molecular and morphologic analysis, the GO Chip enables CTC enumeration, characterization, or RNA expression from as little as 1 mL whole blood of patient samples.[18] The efficiency and sensitivity of this device facilitated our study of CTC enumeration and

RNA expression in a clinical cohort using an extensive 96 gene panel. Toward the goal of using CTCs to provide clinically relevant molecular information that could eventually be utilized to assist with patient management, we undertook a prospective study of 41 men with mCRPC (**Figure 1**). We sought to utilize captured CTCs and extracted RNA from parallel GO Chips to determine CTC characteristics associated with progression and survival in advanced prostate cancer.

Blood samples were collected with informed consent from 41 patients with mCRPC (**Table S1**) recruited under institutional approved IRB (HUM00052405) between August 2013 and November 2016 using EDTA tubes. Processing occurred on the day of blood draw. Eight healthy male controls were recruited internally and processed in the same manner as patient samples. The median patient age was 73 years (range: 50-84 years), while the median baseline PSA level was 37.9 ng/mL (range: 1.2-6433 ng/mL). The median number of prior treatments other than first-line hormonal therapy was one (range: 0-7), and at the time of CTC collection there were 17 patients receiving abiraterone, four receiving cabazitaxel, two receiving cabozantinib, seven receiving docetaxel, eight receiving enzalutamide, one receiving olaparib, and one receiving pembrolizumab. During the study and follow-up period, 34 patients experienced PSA progression; 37 experienced radioclinical progression as defined by a ≥20% increase in the sum of the soft tissue lesion diameters during computed tomography, ≥2 new bone lesions on bone scan, or symptomatic progression (worsening pain aggravation or new cancer-related symptoms); and 22 patients died. For surviving patients, the median time to last follow-up was 19.1 months (range: 3.3-37.8 months). Median time to death was 17.5 months (range: 2.6-39.6 months).

In addition to overall survival, PSA at the time of blood draw, radiographic, and clinical progression events were recorded. PSA progression was defined using the PCWG3 criteria of an increase of greater than or equal to 25% from the nadir, with a minimum increase of 2 ng/mL.[19] Radioclinical progression was also used as a clinical endpoint using the date of whichever happened earliest. Radiographic progression entailed one of three events: 20% or more increase in the sum of the diameters of soft-tissue target lesions based on RECIST criteria applied to CT scans; an increase of at least 5 mm in the short axis of a previously normal lymph node (this lymph node must be at least 1.0 cm in the short axis); or at least two new bone lesions. Clinical progression was defined as worsening disease-related symptoms or new cancer-related complications. Radioclinical progression was assessed by a single reviewer using standard PCWG3 criteria.[19]

CTCs were detected in all 41 samples with the number of CTCs ranging from 3-166 CTCs/mL (median: 20 CTCs/mL, **Figure 2A-C**). The median number of CTCs detected in healthy controls was 3 CTCs/mL (range: 0-14 CTCs/mL). CTC counts for patients were significantly higher than those for healthy controls (p = 0.0001). Quantification of contaminating WBCs is summarized in **Table S2**.

While processing patient samples, we observed groups of two or more adjacent CTCs (**Figure 2D, E**), termed CTC clusters. These clusters were only present in patient samples (26/41, 63.4%) and not healthy controls. Both interpatient and intrapatient heterogeneity were evident from the captured clusters, as cells within the clusters showed varying size and cytokeratin expression. Clusters consisted of up to eight CTCs per cluster (**Figure 2F**) with the majority of the clusters comprising fewer numbers of cells. The percentage of CTCs

6

captured in the form of clusters also varied greatly among patients from 0 to 54.8% (**Figure 2G**). While CTC clusters have been observed previously in prostate cancer patient samples,[20, 21, 22] the high frequency of CTC clusters reported in the present study suggests that the GO Chip may be less disruptive to cell-cell interactions and have greater sensitivity for identifying these clusters.

For 36 of the patients, we had the opportunity to run a parallel microfluidic device that ultimately yielded RNA following cell lysis and purification, which was used for RT-qPCR (**Table S3**). Results from one patient sample were discarded due to insufficient expression of housekeeping genes suggesting lack of sufficient RNA for analysis. In the remaining 35 patient samples, 77 of the 96 genes were detectable (C_T < 30) in at least one patient, and 58 genes were detectable at in at least three patients (**Table S4, S5**). This three-patient cut-off was set for two main reasons. The first consideration was to exclude possible technical artifact for genes that show expression only in one to two patients. Secondly, as much of our subsequent analysis involved comparing relative expression levels between two subgroups of patients, we wanted to be able to stratify patients into two groups with at least three patients in each subgroup to evaluate.

We next examined the relationship between gene expression and CTC counts as well as gene expression and the clusters metrics to determine potential associations. The associations between gene expression and several relevant CTC metrics from each patient sample was examined through linear modeling of these parameters as continuous variables (**Figure S1A-F**). If the relationship between the enumeration variable and the gene expression had a negative linear coefficient, the gene expression decreased as the

enumeration variable increased (i.e. as the number of CTCs, the number of clusters, or percentage of CTCs in clusters increased). FOXC2 and CTNND1 were positively associated with the number of CTCs/mL while NFKB1 was negatively associated with CTC counts (p < 0.05). CTNND1 and ZEB2 were negatively associated with the number of clusters present per sample, and CTNND1 and FOXC1 were negatively associated with the percentage of CTCs found in clusters (p < 0.05). As overexpression of δ-catenin has been associated with increased proliferation,[23] the lower expression and therefore potentially lower proliferation would be consistent with an observed stemlike phenotype within the clusters. This is consistent with the lower ZEB2 expression. FOXC1 expression is associated with poor prognosis, androgen independence, and angiogenesis,[24] potentially indicating aggressive properties in CTCs present as single cells in contrast to those in clusters.

Multivariate analysis was then conducted using a best subset selection method. As there were 31 PSA progression events, we opted for subsets of two to three genes, applying the "one in ten" rule.[25] Overall survival, PSA progression, and radioclinical progression were modeled as a function of these two or three subsets using Cox proportional hazards. Models with a global Wald p value of less than 0.05 were selected and sorted by their Akaike information criterion (AIC[26]), with lower AICs denoting better quality models. In total, 32509 combinations were assessed. For overall survival, 311 nominally significant gene combinations (10 two-gene combinations, 301 three-gene combinations); for PSA progression, there were 566 nominally significant gene combinations (16 two-gene combinations, 550 three-gene combinations); and for radioclinical survival, there were 160 nominally significant combinations (six two-gene combinations, 154 three-gene

combinations). Some genes appeared in significant models at higher frequencies than other (**Table S6**).

For overall survival, the three-gene combination of CDH1, CD11B, and STAT3 showed the best performance based on AIC (HR: 0.78, 95% CI: 0.62-0.99; HR: 2.04, 95% CI: 1.32-3.14; and HR: 0.46, 95% CI: 0.30-0.72 respectively). For PSA progression, the three-gene combination of CD44, CASP3, and FOXC2 performed best (HR: 1.51, 95% CI: 1.23-1.87; HR: 0.79, 95% CI: 0.69-0.90; HR: 0.90, 95% CI: 0.82-0.98, respectively). For radioclinical progression, the best model consisted of the three genes CDH1, AR, and CD45 (HR: 0.84, 95% CI: 0.71-0.98; HR: 1.10, 95% CI: 1.02-1.18; HR: 0.78, 95% CI: 0.66-0.92).

To assess the clinical relevance of the experimental data obtained, we compared CTC metrics and gene expression with overall survival, PSA progression, and radioclinical progression. None of the enumeration variables, including CTC and cluster counts and related metrics, were statistically significant in the univariable Cox proportional hazards models. However, patients with a high number of clusters relative to the median had a shorter time to radioclinical progression than those with a low number of clusters in the Kaplan-Meier survival analysis (log rank p < 0.05, **Figure S1G**). A previous study associated CTC clusters detected at any of multiple time points with decreased overall survival;[24] this suggests a future direction for our technology, in which serial sampling may provide more prognostic information.

To construct a bimodal point-based metric relating gene expression (**Figure 3**) to clinical outcomes, we used cut-points to classify patients in to high and low survival groups as determined by subsequent Kaplan-Meier analysis. Cut-points were generated using

regression tree analysis from the rpart (recursive partitioning) package in the R software environment. The rpart package uses regression models based on the input data set to find the variable and location that best splits the data into two groups, where best is defined as minimizing the risk of misclassification. In this case, the input data consisted of the gene expression levels and clinical outcomes. The distribution of expression above or below the generated cut-point was visualized using a beeswarm plot, with expression above the cut-point denoted as "high" and expression below the cut-point denoted as "low." Genes associated with overall survival included CD44, CDH1, EPCAM, ERCC1, IL8, PIK3CA, STAT3, TGFβ, TIMP2, and ZEB2 (**Figure 3B**). The genes CDH1, CD146, FOXC2, and ZEB2 were associated with PSA progression, while the genes associated with radioclinical progression included ACTB, CDH1, CDH2, CD3D, CD45, CASP3, CD146, CXCR1, KLF4, KRAS, MKI67, MMP9, RB1, SPARC, XBP1, and ZEB2. The number of patients in each group along with the relevant median survival metrics are included in **Table S7**.

Using these pre-screened genes, we developed a series of prognostic scores to predict survival/progression. Based on gene expression relative to the cut-point, we assigned genes a point value of 0 or 1 based on whether their gene expression value fell above or below the cut-point based on its relationship to shorter survival or progression time. For example, in the case of overall survival, patients expressing greater levels of CD44 than the cut-point were assigned 1 point, while patients expressing lower levels of CDH1 than the cut-point were assigned 1 point. Conversely, patients expressing lower levels of CD44 than the cut-point were assigned 0 points, while patients expressing higher levels of CDH1 than the cut-point were assigned 0 points. The values for each of the genes were then added in different

combinations with all subsets considered and used to generate receiver operating characteristic (ROC) curves, with the scores ranging from 0 to the number of genes in the score. These curves were then evaluated based on the area under the curve (AUC), with AUCs approaching one having greater discriminatory ability. The 95% confidence interval (CI) was generated by resampling, with 10^4 bootstrap samples. Additionally, the Cox proportional hazard ratio[27] was calculated for each score.

Of the significant genes established for overall survival, a set of eight genes ultimately had the best prognostic power: CD44, CDH1, EPCAM, ERCC1, PIK3CA, STAT3, TGFB1, and ZEB2 (**Figure 4A**). The relative survival of the patients based on the differential expression of these genes is represented in Figure 5A. The group with genes expressed at higher levels than the cut-point is shown in red while the other group below the cut-point is shown in blue, clearly demonstrating the differences in prognosis between the two groups. The resulting ROC had an AUC of 0.88 (95% CI: 0.69-0.98, **Figure 4B**). Assessing the score using the Cox proportional hazard model yielded a hazard ratio of 1.83 (95% CI: 1.33-2.51) for overall survival, representing the increased risk between the patient groups one unit apart on the numerical score. Three of the genes in the score, CDH1, EPCAM, and ZEB2, are reflective of epithelial or mesenchymal phenotypes. The association of the low expression of both the epithelial and mesenchymal genes with lower overall survival suggests the importance of a transitory phenotype in leading to disease progression. This potentially undifferentiated phenotype is in concordance with the high expression of prostate cancer stemness marker CD44.[28] Low ZEB2 was also implicated in the scores for PSA progression and radioclinical progression, while low CDH1 was implicated in the score for PSA progression, suggesting an emerging theme. Additionally, STAT3 is activated by AR signaling loss and is associated with cancer

11

stem cells.[29] These observations are in line with previous descriptions of stemlike tumor-propagating subpopulations in prostate cancer.[30]

Using the same methods to relate the subsets of significant genes to radioclinical progression produced a five-gene score consisting of CD3D, MMP9, RB1, XBP1, and ZEB2 (AUC: 0.82, 95% CI: 0.67-0.96, **Figure S2**). This score carried a hazard of 1.51 (95% CI: 1.19-1.92) for radioclinical progression. Finally, investigation of combinations of genes related to PSA progression resulted in a three-gene score: CDH1, CD146, and ZEB2 (AUC: 0.69, 95% CI: 0.47-0.87, **Figure S3**); and a proportional hazard of 1.75 (95% CI: 1.17-2.63).

Overall, our study has important limitations. Without question, this work will continue to benefit from longer follow-up and corroboration in larger cohorts. Further work to validate this approach and to explore the implications of CTC stemness is ongoing. Our results also highlight the importance of examining protein and RNA expression in CTCs, as we were able to use anti-EpCAM for capture (indicative of protein expression), but saw low expression of EpCAM RNA. While the previous work establishing the device gives us confidence in our ability to capture low EpCAM-expressing cells,[18] future work may incorporate additional capture antibodies to account for CTC heterogeneity. The recruitment of additional controls will be important to examine background levels of cytokeratin positive cells. This false positive rate, a marker of the balance between assay sensitivity and background noise, will need to continue to be refined. Furthermore, our current study makes use of bulk RNA extraction and RT-qPCR. These techniques were selected to efficiently study the maximum number of genes from a cohort of this size. While we have taken steps to

subtract "background" signal from white blood cells in the form of processed healthy control samples, we plan to integrate single-cell techniques into future work to better assess intrapatient heterogeneity and reduce our signal to noise ratio.

CTCs provide an opportunity to study the underlying biology and disease trajectory of prostate cancer in a readily-available liquid biopsy. The incorporation of microfluidics into CTC research offers substantial promise for overcoming the hurdles of low CTC counts and the immense quantity of surrounding normal blood cells. It was our goal to integrate an ultra-sensitive CTC isolation technology enabling both immunofluorescence characterization and RNA expression analysis in order to investigate the relationship of the CTC-based results with key oncologic outcomes. Our work here is an example of a clinically relevant biomarker discovery approach utilizing CTC-related metrics.

Experimental Section

Device fabrication: Fabrication of the graphene oxide chip (GO Chip) has been described previously.[18] Briefly, a suspension was made by probe tip sonicating graphene oxide (CheapTubes.com) and tetrabutyl ammonium hydroxide (TBA, Fluka) in dimethylformamide (Sigma-Aldrich). A phospholipid-polyethyleneglycol-amine (PEG, NOF America Corporation) was added to the resulting suspension and bath sonicated for 1 hour. Silicon wafers with gold features fabricated using photolithography were dipped in the suspension to

allow self-assembly of the GO-TBA-PEG onto the gold. This pattern was then enclosed in a polydimethylsiloxane (PDMS, Dow Corning) microfluidic chamber. The crosslinker N-γ-maleimidobutyryl-oxysuccinimide ester (GMBS, Pierce) was then added to the devices and incubated at which point tubing (Tygon) was inserted. Following a wash, NeutrAvidin (Invitrogen) was added to the devices via syringe pump (Harvard Apparatus). The devices could be stored at 4º C until use, at which time biotinylated anti-EpCAM (**Table S8**) in 1% bovine serum albumin (BSA, Sigma) was added prior to sample processing.

Patient sample processing: Following blocking with BSA, 1 mL whole blood was introduced into the GO Chip at a flow rate of 1 mL/hr. Devices were then flushed with a total volume of 6 mL phosphate buffered saline (PBS, Gibco) at 100 μl/min immediately following blood flow. Subsequent steps were determined based on the ultimate application of the device in the work-flow. For devices that would be stained for enumeration, the contents of the PDMS chamber were fixed using 4% paraformaldehyde (PFA, ThermoFisher). These devices were then stored at 4°C until they were stained. On a parallel device, RNA extraction was performed by first flowing RNA extraction buffer (a component of the PicoPure® RNA Isolation Kit, Arcturus). The device and syringe were then incubated for 30 minutes at 42°C, after which DEPC water (ThermoFisher) was flowed. Collected RNA extraction buffer and DEPC water from the device outlet were stored at -80°C until purification.

Immunofluorescence staining: Subsequent to processing and fixation of the sample, immunofluorescence staining was performed on-chip using a syringe pump. Cells were

permeabilized using Triton X (Sigma) and then blocked a combination of goat serum (ThermoFisher) and BSA. See **Table S8** for all antibody information. Primary antibodies against CD45 and cytokeratin 7/8 were detected using the appropriate secondary antibodies labeled with Alexa Fluors 488 and 546. Antibodies were suspended in 1% BSA while 2-(4-amidinophenyl)-1H-indole-6-carboxamidine (DAPI, Invitrogen) in PBS was used to label cell nuclei. Imaging of fluorescence staining was conducted on a Nikon Eclipse Ti fluorescence microscope using either a 10x or 20x objective. Images were captured using a QImaging cooled mono 12 bit camera and analyzed using NIS-Elements software. Those nucleated cells expressing CK but not CD45 (DAPI+/CK+/CD45- cells), were counted as CTCs.

Quantitative reverse-transcription polymerase chain reaction (RT-qPCR): Bulk cell lysates extracted during sample processing from 36 patients and four healthy controls were subsequently purified using the remaining components of the PicoPure® RNA Isolation Kit; five patients were processed only for enumeration and not RNA expression analysis (**Table S3**). Purification was conducted according to the manufacturer's protocol. Purified RNA was reverse transcribed to cDNA according to the manufacturer's protocol using an Ambion kit (ThermoFisher). The cDNA was pre-amplified after which it underwent RT-qPCR using an Applied Biosystems[TM] TaqMan® Gene Expression Assay. Using the BioMark HD qPCR platform (Fluidigm), C_T levels were determined for 96 genes of interest (complete list, **Table S4**) in the following categories: apoptosis, blood cell, cell cycle, cell junction, cytoskeleton, developmental, DNA repair, extracellular matrix, epithelial, growth factor, hormone, housekeeping, inflammation/immune system, long noncoding RNA, mesenchymal,

oncogene, proliferation, stemness, transcription factor, and tumor suppressor. To visualize this data, a heatmap was generated using the heatmap function from the "stats" package in the R programming environment. Hierarchical clustering was performed using the complete linkage method as a function of Euclidean distance, which is the default setting for that function.

Statistical analysis: The primary outcome of interest was overall survival. Data variables were related to either enumeration or RNA, and in the case of enumeration included CTCs/mL, the presence of clusters, the number of clusters, the percentage of CTCs in clusters, the average number of CTCs/cluster, and the maximum number of CTCs/cluster. The base ten log or z score of the enumeration variables was taken for the purposes of analysis. Analyses were performed using Excel and R with the following R packages: rpart,[31] survival,[32] and survivalROC.[33] The CTC enumeration variables were compared to the clinical metrics using Cox proportional hazards models and Kaplan-Meier survival analysis. The Wald test was used to determine significance for Cox proportional hazards modeling, while the log-rank test was used in the Kaplan-Meier analysis. In other comparisons, statistical significance was determined using the Mann-Whitney test. A nominal p-value of less than 0.05 was considered statistically significant. RT-qPCR results were first normalized to the mean of three housekeeping genes (GAPDH, ACTB, UBB) to obtain a ΔC_t value, and then background corrected by deducting the mean expression level of each in the four healthy controls to obtain a $\Delta\Delta C_t$ value, and subsequently analyzed as $\log_2(2^{-\Delta\Delta C_t} + 1)$. To select genes for the generation of a point-based score, analysis conducted with the rpart

package in R was used as a screening mechanism, with nominally significant genes being considered. Scores were analyzed using receiver operating characteristic (ROC) curves using the survival ROC function. As employed in our analysis, this function plotted the true positive rate against the false positive rate for the score's ability to predict median survival time using the Nearest Neighbor Estimation method. A higher area under the curve (AUC) from this function denoted better model performance.

Supporting Information

Supporting Information is available from the Wiley Online Library or from the author.

Acknowledgements

References

[1]   D. R. Berthold, G. R. Pond, F. Soban, R. de Wit, M. Eisenberger, I. F. Tannock, Journal of Clinical Oncology 2008, 26, 242; H. I. Scher, K. Fizazi, F. Saad, M.-E. Taplin, C. N. Sternberg, K. Miller, R. de Wit, P. Mulders, K. N. Chi, N. D. Shore, New England Journal of Medicine 2012, 367, 1187; C. J. Ryan, M. R. Smith, J. S. De Bono, A. Molina, C. J. Logothetis, P. De Souza, K. Fizazi, P. Mainwaring, J. M. Piulats, S. Ng, New England Journal of Medicine 2013, 368, 138.

[2]   N. Mehra, Z. Zafeiriou, D. Lorente, L. W. Terstappen, J. S. de Bono, Clinical cancer research 2015, 21, 4992.

[3]   K. Pantel, C. Alix-Panabières, Cancer research 2013, 73, 6384.

[4]   S. Maheswaran, D. A. Haber, Current opinion in genetics & development 2010, 20, 96.

[5]   M. Kozminsky, Y. Wang, S. Nagrath, Current Opinion in Chemical Engineering 2016, 11, 59.

[6]   D. R. Shaffer, M. A. Leversha, D. C. Danila, O. Lin, R. Gonzalez-Espinoza, B. Gu, A. Anand, K. Smith, P. Maslak, G. V. Doyle, Clinical Cancer Research 2007, 13, 2023; D. C. Danila, G. Heller, G. A. Gignac, R. Gonzalez-Espinoza, A. Anand, E. Tanaka, H. Lilja, L. Schwartz, S. Larson, M. Fleisher, Clinical Cancer Research 2007, 13, 7053; J. S. de Bono, H. I. Scher, R. B. Montgomery, C. Parker, M. C. Miller, H. Tissing, G. V. Doyle, L. W. W. M. Terstappen, K. J. Pienta, D. Raghavan, Clinical Cancer Research 2008, 14, 6302.

[7]   W. J. Allard, J. Matera, M. C. Miller, M. Repollet, M. C. Connelly, C. Rao, A. G. J. Tibbe, J. W. Uhr, L. W. M. M. Terstappen, Clinical Cancer Research 2004, 10, 6897.

[8]   H. J. Yoon, M. Kozminsky, S. Nagrath, ACS Nano 2014, 8, 1995.

[9]   S. L. Stott, R. J. Lee, S. Nagrath, M. Yu, D. T. Miyamoto, L. Ulkus, E. J. Inserra, M. Ulman, S. Springer, Z. Nakamura, A. L. Moore, D. I. Tsukrov, M. E. Kempner, D. M. Dahl, C. Wu, A. J. Iagrate, M. R. T. R. G. Smith, L. V. Sequist, M. Toner, D. A. Haber, S. Maheswaran, Sci Transl Med 2010, 2, 25ra23.

[10]   S. L. Stott, C. H. Hsu, D. I. Tsukrov, M. Yu, D. T. Miyamoto, B. A. Waltman, S. M. Rothenberg, A. M. Shah, M. E. Smas, G. K. Korir, F. P. Floyd, A. J. Gilman, J. B. Lord, D. Winokur, S. Springer, D. Irimia, S. Nagrath, L. V. Sequist, R. J. Lee, K. J. Isselbacher, S. Maheswaran, D. A. Haber, M. Toner, Proceedings of the National Academy of Sciences 2010, 107, 18392.

[11]   D. T. Miyamoto, Y. Zheng, B. S. Wittner, R. J. Lee, H. Zhu, K. T. Broderick, R. Desai, D. B. Fox, B. W. Brannigan, J. Trautwein, Science 2015, 349, 1351.

[12]   J. M. Sperger, L. N. Strotman, A. Welsh, B. P. Casavant, Z. Chalmers, S. Horn, E. Heninger, S. M. Thiede, J. Tokar, B. K. Gibbs, Clinical Cancer Research 2016.

[13]   S. Park, R. R. Ang, S. P. Duffy, J. Bazov, K. N. Chi, P. C. Black, H. Ma, PloS one 2014, 9, e85264; J. F. Chen, H. Ho, J. Lichterman, Y. T. Lu, Y. Zhang, M. A. Garcia, S. F. Chen, A. J. Liang, E. Hodara, H. E. Zhau, Cancer 2015, 121, 3240; A. S. McDaniel, R. Ferraldeschi, R. Krupa, M. Landers, R. Graf, J. Louw, A. Jendrisak, N. Bales, D. Marrinucci, Z. Zafeiriou, Bju international 2016.

[14]   E. S. Antonarakis, C. Lu, H. Wang, B. Luber, M. Nakazawa, J. C. Roeser, Y. Chen, T. A. Mohammad, Y. Chen, H. L. Fedor, New England Journal of Medicine 2014, 371, 1028.

[15]   J. P. Gleghorn, E. D. Pratt, D. Denning, H. Liu, N. H. Bander, S. T. Tagawa, D. M. Nanus, P. A. Giannakakou, B. J. Kirby, Lab on a chip 2010, 10, 27.

[16]   D. T. Miyamoto, R. J. Lee, S. L. Stott, D. T. Ting, B. S. Wittner, M. Ulman, M. E. Smas, J. B. Lord, B. W. Brannigan, J. Trautwein, N. H. Bander, C. L. Wu, L. V. Sequist, M. R. Smith, S. Ramaswamy, M. Toner, S. Maheswaran, D. A. Haber, Cancer Discovery 2012, 2, 995.

[17]   B. J. Kirby, M. Jodari, M. S. Loftus, G. Gakhar, E. D. Pratt, C. Chanel-Vos, J. P. Gleghorn, S. M. Santana, H. Liu, J. P. Smith, V. N. Navarro, S. T. Tagawa, N. H. Bander, D. M. Nanus, P. Giannakakou, PloS one 2012, 7, e35976.

[18]   H. J. Yoon, T. H. Kim, Z. Zhang, E. Azizi, T. M. Pham, C. Paoletti, J. Lin, N. Ramnath, M. S. Wicha, D. F. Hayes, D. M. Simeone, S. Nagrath, Nature nanotechnology 2013, 8, 735.

[19]   H. I. Scher, M. J. Morris, W. M. Stadler, C. Higano, E. Basch, K. Fizazi, E. S. Antonarakis, T. M. Beer, M. A. Carducci, K. N. Chi, Journal of Clinical Oncology 2016, 34, 1402.

[20]   M. A. Leversha, J. Han, Z. Asgari, D. C. Danila, O. Lin, R. Gonzalez-Espinoza, A. Anand, H. Lilja, G. Heller, M. Fleisher, H. I. Scher, Clinical cancer research : an official journal of the American Association for Cancer Research 2009, 15, 2091; C. P. Meyer, K. Pantel, P. Tennstedt, P. Stroelin, T. Schlomm, H. Heinzer, S. Riethdorf, T. Steuber, "Limited prognostic value of preoperative circulating tumor cells for early biochemical recurrence in patients with localized prostate cancer", presented at *Urologic Oncology: Seminars and Original Investigations*, 2016.

19

[21]   H. I. Scher, D. Lu, N. A. Schreiber, J. Louw, R. P. Graf, H. A. Vargas, A. Johnson, A. Jendrisak, R. Bambury, D. Danila, JAMA oncology 2016, 2, 1441.

[22]   N. Aceto, A. Bardia, D. T. Miyamoto, M. C. Donaldson, B. S. Wittner, J. A. Spencer, M. Yu, A. Pely, A. Engstrom, H. Zhu, Cell 2014, 158, 1110.

[23]   Y. Zeng, A. Abdallah, J.-P. Lu, T. Wang, Y.-H. Chen, D. M. Terrian, K. Kim, Q. Lu, Molecular cancer 2009, 8, 19.

[24]   B. Han, N. Bhowmick, Y. Qu, S. Chung, A. Giuliano, X. Cui, Oncogene 2017, 36, 3957.

[25]   F. E. Harrell, K. L. Lee, D. B. Mark, Statistics in medicine 1996, 15, 361.

[26]   H. Bozdogan, Psychometrika 1987, 52, 345.

[27]   T. M. Therneau, P. Grambsch, Edited by P. Bickel, P. Diggle, S. Fienberg, K. Krickeberg 2000, 51.

[28]   A. T. Collins, P. A. Berry, C. Hyde, M. J. Stower, N. J. Maitland, Cancer research 2005, 65, 10946.

[29]   A. Schroeder, A. Herrmann, G. Cherryholmes, C. Kowolik, R. Buettner, S. Pal, H. Yu, G. Müller-Newen, R. Jove, Cancer research 2014, 74, 1227.

[30]   J. Qin, X. Liu, B. Laffin, X. Chen, G. Choy, C. R. Jeter, T. Calhoun-Davis, H. Li, G. S. Palapattu, S. Pang, Cell stem cell 2012, 10, 556.

[31]   T. Therneau, B. Atkinson, B. Ripley, 2015.

[32]   T. M. Therneau, 2015.

[33]   P. J. Heagerty, P. Saha-Chaudhuri, 2013.
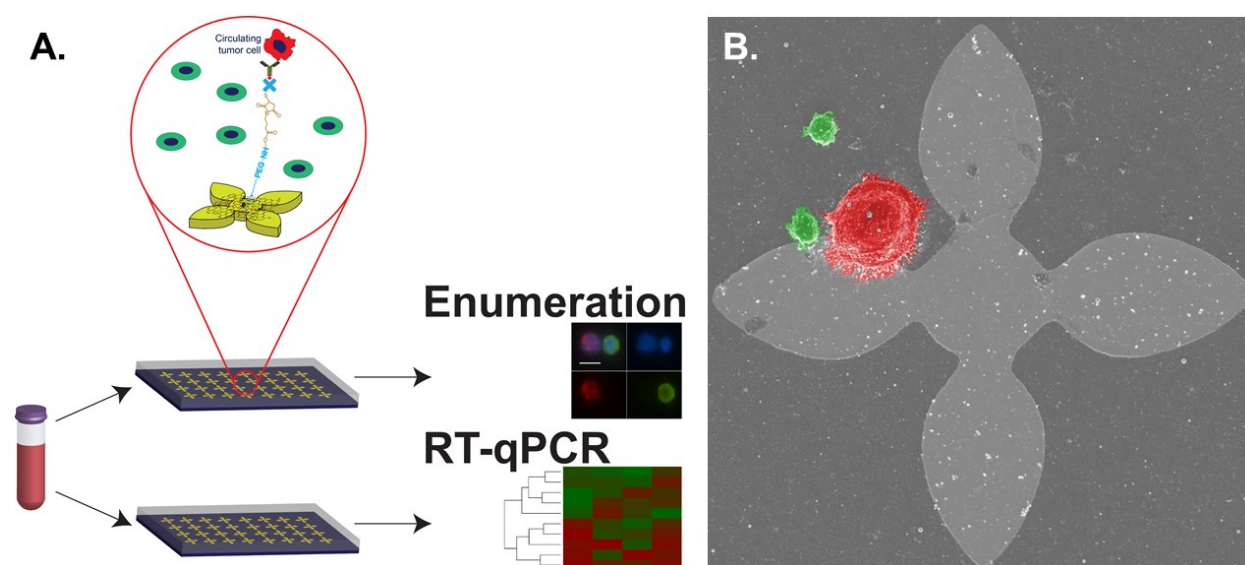
Figure 1. Graphene oxide chip enables isolation of prostate circulating tumor cells. A) Sample workflow. Two parallel devices were processed, one each for circulating tumor cell enumeration and RNA extraction. B) Scanning electron micrograph of PC-3 cell (red pseudocolor) and white blood cells (green pseudocolor) on-chip. Flower patter is 100 µm in height and width.
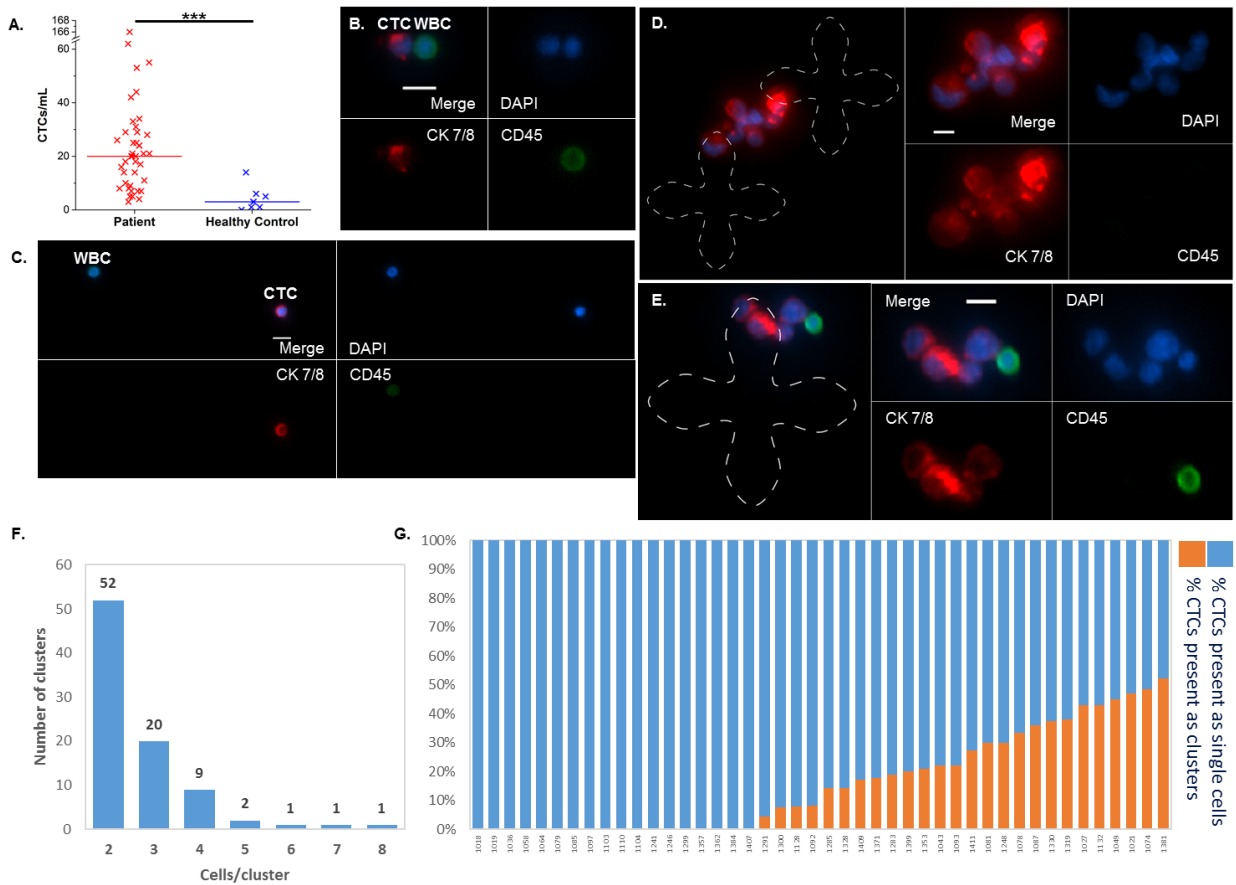
Figure 2. Circulating tumor cells and circulating tumor cell clusters isolated by the graphene oxide chip. A) Circulating tumor cell (CTC) enumeration results for 41 metastatic castrate resistant prostate cancer patient samples (range: 3-166 CTCs/mL, median: 20) and (range: 3-166 CTCs/mL, median: 20) and epithelial cells detected in eight healthy controls (range: 0-14 epithelial cells/mL, median: 3). ***denotes p < 0.001. B, C) Examples of CTCs captured on-chip as well as non-specifically bound white blood cells (WBCs). Nuclear staining is shown in blue, cytokeratin 7/8 in red, and CD45 in green. D, E) Examples of captured circulating tumor cell (CTC) clusters. CTCs captured within clusters had heterogeneous size and cytokeratin expression. The capture pattern is outlined with a dashed line for visualization purposes. Scale bar is 10 μm. F) Captured CTC clusters ranged in size from two to eight cells per cluster. G) The percentage of captured CTCs present in clusters ranged from 0 to 54.8%.

A.
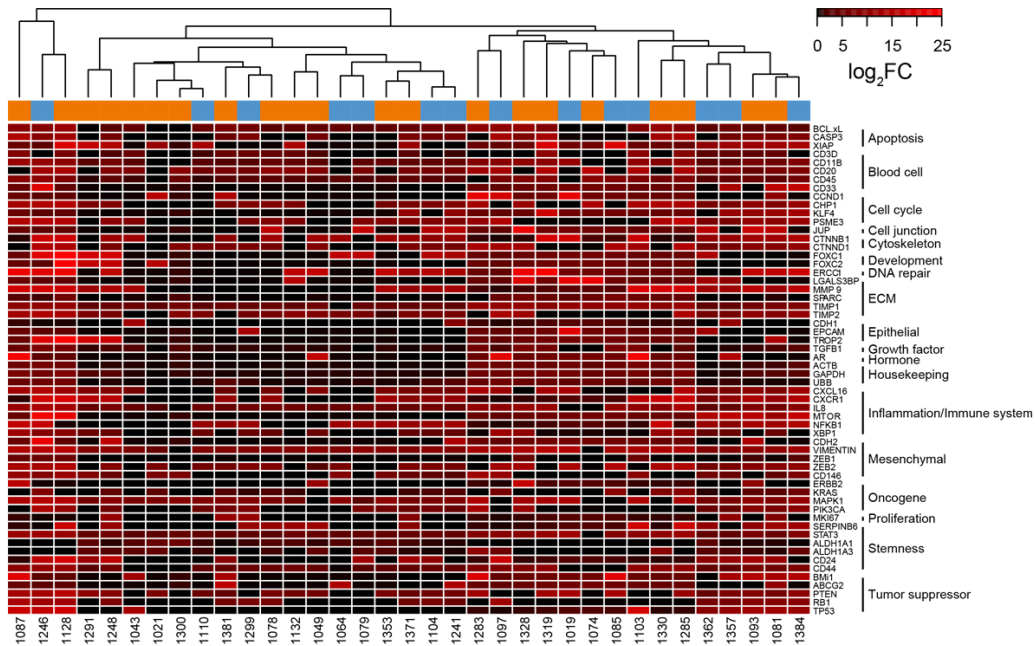
B.
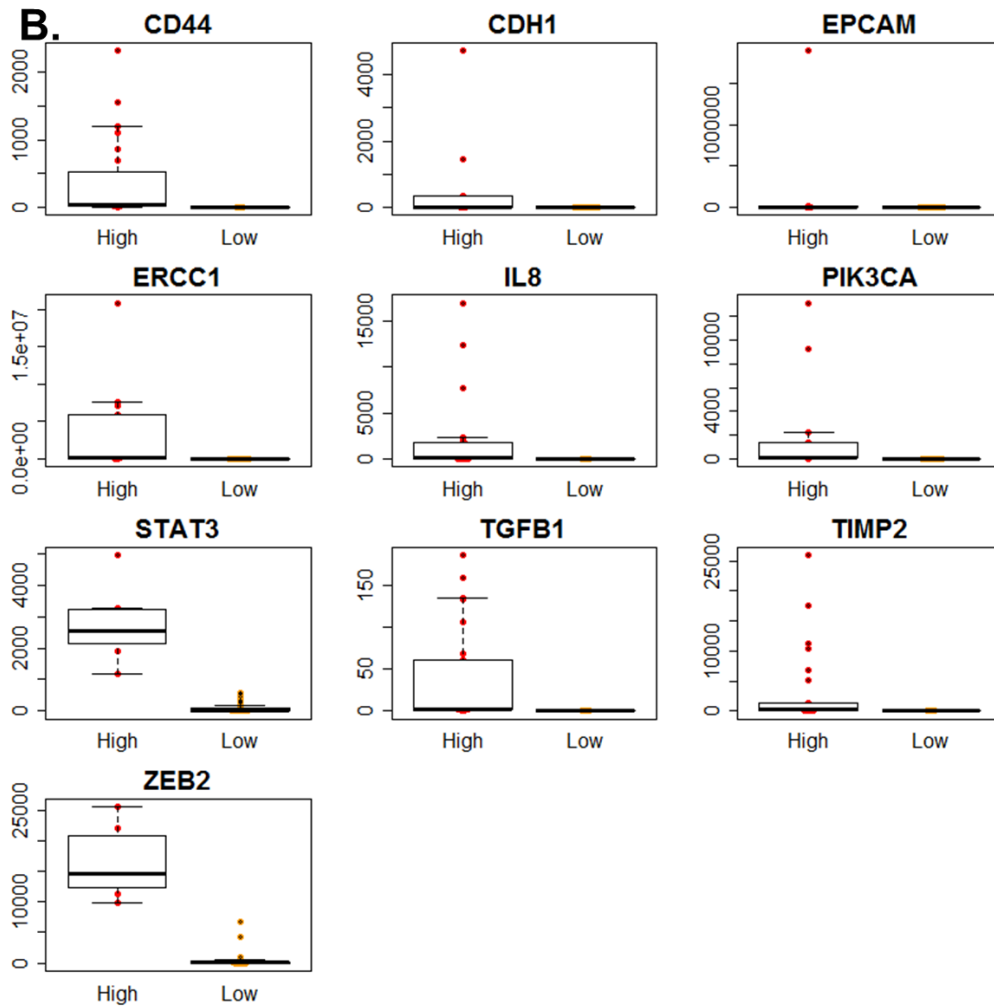
Figure 3. Relative gene expression for use in a bimodal score. A) Heatmap of log$_2$ fold changes (FC) relative to healthy control background for 58 genes detected in patient samples. Associated gene categories are shown to the right of the gene list. Color coding above the heatmap indicates the presence (orange) or absence (blue) of clusters in a sample. B) Beeswarm plot of expression ($2^{-\Delta\Delta C_t}$) for genes enabling stratification based on overall survival.
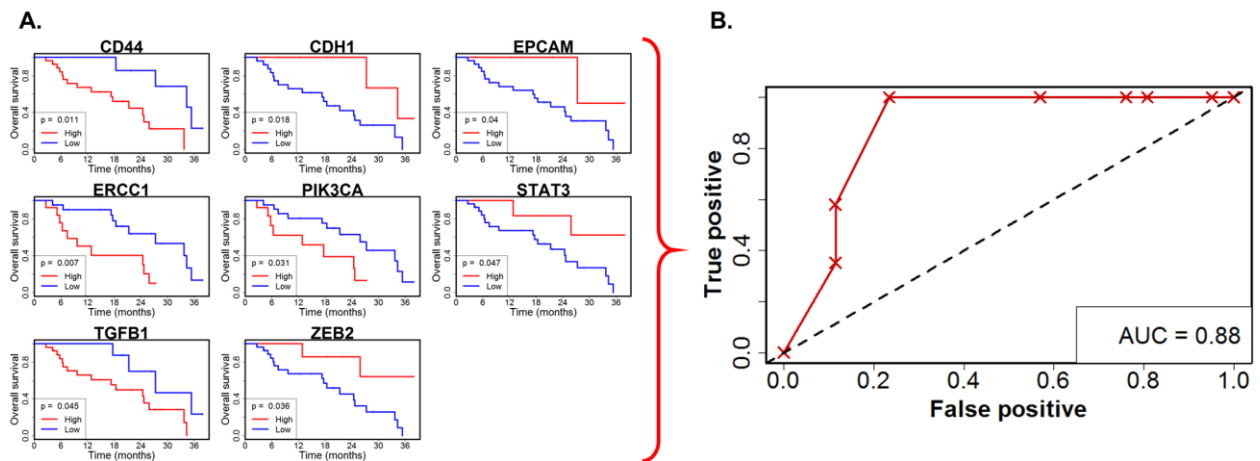
Figure 4. Relationship between RNA expression and overall survival. A) Kaplan-Meier curves for genes with statistically significant relationships with overall survival used to construct the optimized point-based score. B) Scores were optimized by maximizing the area under the curve (AUC) of the associated receiver operating curve.

A microfluidic graphene oxide based device (GO Chip) is used to isolate CTCs and CTC clusters from the whole blood of 41 men with metastatic castration-resistant prostate cancer. A preliminary RNA signature, comprised of high expression of stemness genes and low expression of epithelial and mesenchymal genes, potentially implicates an undifferentiated CTC phenotype as a marker of poor prognosis in this setting.

Keywords: circulating tumor cells, microfluidics, GO Chip, prostate cancer, RNA expression

M. Kozminsky, S. Fouladdel, J. Chung, Y. Wang, D.C. Smith, A. Alva, E. Azizi, T. Morgan*, S. Nagrath*

Detection of CTC clusters and a dedifferentiated RNA-expression survival signature in prostate cancer

ToC figure ((Please choose one size: 55 mm broad × 50 mm high **or** 110 mm broad × 20 mm high. Please do not use any other dimensions))

25