**Supplementary Material for *Circulating Tumor Cell Phenotypic Heterogeneity Informs Clinical Decisions between Androgen Receptor Signaling Inhibitors and Taxanes in Metastatic Prostate Cancer***

**Elements Included:**

**Supplementary Methods**

**Supplementary Figures and Figure Legends:**

**Supplemental Figure 1:** Hypothesis: Pre-therapy CTC Phenotypic Heterogeneity Associates with Differential Outcomes on Pathway-Specific Hormonal Therapeutics and Non-Pathway-Specific Chemotherapy.

 **Supplemental Figure 2:** Principal Component Analysis and k-selection for Unsupervised Clustering of CTCs for Phenotypic Subtype Classification.

 **Supplemental Figure 3:** Correlations between Shannon index and Other Pre-therapy Features.

 **Supplemental Figure 4:** Multivariate Analyses of Sub-Cohort with Shannon index Greater Than 0

 **Supplemental Figure 5:** CTC and Clinical Association Analysis Overview.

 **Supplemental Figure 6:** The Degree of Inter-Sample Pleomorphism index is Related to Overall Survival of ARSI, but not Taxanes.

**Supplementary Methods**

*Pleomorphism Index: An Alternative Analytical Pipeline for Measuring CTC Phenotypic Heterogeneity*

In pathology, the term pleomorphism is used to describe variability in the size, shape, and staining of cells within a sample. We performed a variance of variance analysis (Supplemental Figure 3) to identify key features that contributed the most variation in size, shape, and staining of CTCs within patient blood samples across the CTC contributing cohort. The same raw digital pathology features (Supplemental Figure 3A) utilized in the Shannon index were employed. First, the coefficient of variation (variance / mean) was calculated per digital pathology feature across the CTCs within each patient sample. Then, the variance of the aforementioned coefficients of variation was calculated across all samples in the clinical association cohort. Four features captured upwards of 85% of the variation in intra-sample CTC phenotypes across the cohort. The coefficients of variation of these four features was summed together to create the Pleomorphism index (Supplemental Figure 3B). Because calculating variance requires the presence of more than one CTC, a value of zero was assigned to a sample if zero or one CTCs were present. All values were multiplied by a constant to normalize differences in the absolute numerical range of the Pleomorphism index relative to the Shannon index and facilitate easier biomarker comparisons.

Similar to analyzing entropy with Shannon index, we evaluated if the degree of CTC phenotypic variance measured with the Pleomorphism index within a sample had differential associations with overall survival (OS) for patients going onto targeted ARSI therapy or taxane chemotherapy (Supplemental Figure 4A-C). We also evaluated if the relationship between CTC feature variance (Pleomorphism index), OS, and therapy class remained significant in the context of multivariate Cox PH models correcting for potential imbalances between the pre-ARSI and pre-taxane patient samples (Supplemental Figure 4D-E).



**Supplemental Figure 1: Hypothesis: Pre-therapy CTC Phenotypic Heterogeneity Associates with Differential Outcomes on Pathway-Specific Hormonal Therapeutics and Non-Pathway-Specific Chemotherapy*.*** Shown is a representation of the central hypothesis. It has been proposed that **(A)** patients with disseminated but relatively non-heterogeneous disease would have strong responses to an appropriately targeted agent, **(B)** while those with heterogeneous diseasewould not. Conversely, a non-pathway-specific agent would have anti-tumor effects in patients with relatively **(C)** low or **(D)** high intra-patient heterogeneity.

**(B)**

**(A)**

**** 

**Supplemental Figure 2: Principal Component Analysis and k-selection for Unsupervised Clustering of CTCs for Phenotypic Subtype Classification.** Principle components of **(A)** z-score transformed raw data from all 9225 CTCs in the CTC Contributing Cohort. Arrow indicates 6th principle component, associated with 85% total variance of data. **(B)** Plotted reduction in within-group sum of squares by increasing k clusters. Arrow indicates chosen “elbow” of flat-lined rate of decrease in within-group sum of squares, at k = 15.

 

**Supplemental Figure 3: Correlations between Shannon index and Other Pre-therapy Features.** Matrices comparing Shannon index to continuous and categorical pre-clinical features with scatter plots and box and whisker plots, respectively. Boxes represent the inter-quartile range.



**Supplemental Figure 4: Multivariate Analyses of Sub-cohort with Shannon index Greater Than 0.** Individual covariates were tested for additive power to predict overall survival using a Cox proportional hazards (PH) model for a subset of the cohort with a Shannon index greater than 0 (n = 112). **(A)** The resulting p-values, hazard ratios, and 95% confidence intervals. **(B)** The interaction of therapy and heterogeneity integrated into the multivariate Cox PH model. The forest plot shows hazard ratios and 95% confidence intervals.



**Supplemental Figure 5: CTC and Clinical Association Analysis Overview.** Shown are schematics for **(A)** CTC detection and digital pathology analysis on single cells, as well as generation of patient-level quantification of phenotypic heterogeneity by **(B)** Pleomorphism index.



**Supplemental Figure 6: The Degree of Inter-Sample Pleomorphism index is Related to Overall Survival of ARSI, but not Taxanes. (A)** The relationship between degree of heterogeneity (Pleomorphism index, x-axis) and overall survival (y-axis) is shown, along with nonparametric kernel estimates of median survival. Colors represent treatment received after pre-therapy draw. “O” = patient alive at last observation, “X” = patient died at time indicated. Overall Survival is alternately visualized with Kaplan-Meier plots from patients starting **(B)** ARSI and **(C)** Taxanes with survival curves dichotomized with the survival crossover point from **(A)**, indicated with an arrow. Individual covariates were tested for additive power to predict overall survival using a Cox proportional hazards (PH) model. **(D)** The resulting p-values, hazard ratios, and 95% confidence intervals. **(E)** The interaction of therapy and heterogeneity integrated into the multivariate Cox PH model. The forest plot shows hazard ratios and 95% confidence intervals.