**Supplementary Figures and Tables**

**The Initial Detection and Partial Characterization of Circulating Tumor Cells in**

**Neuroendocrine Prostate Cancer**

Himisha Beltran1,2, Adam Jendrisak3, Mark Landers3, Juan Miguel Mosquera2,4, Myriam Kossai2, Jessica Louw3, Rachel Krupa3, Ryon P Graf3, Nicole Schreiber5, David M Nanus1, Scott T Tagawa1, Dena Marrinucci3, Ryan Dittamore3, Howard I Scher5

1Department of Medicine, Division of Hematology and Medical Oncology, Weill Cornell Medical College, New York, NY.

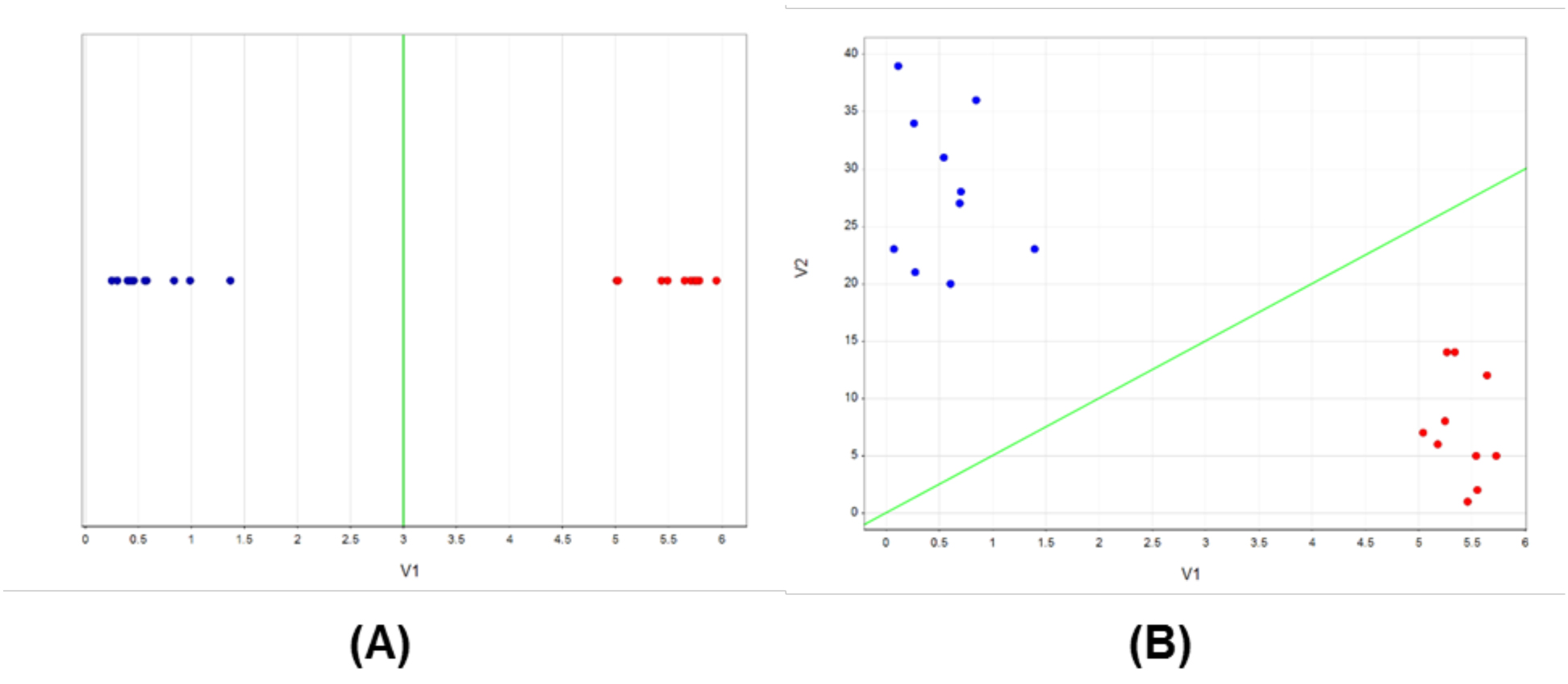
2Institute for Precision Medicine, New York Presbyterian-Weill Cornell Medical College

3Epic Sciences, San Diego, California

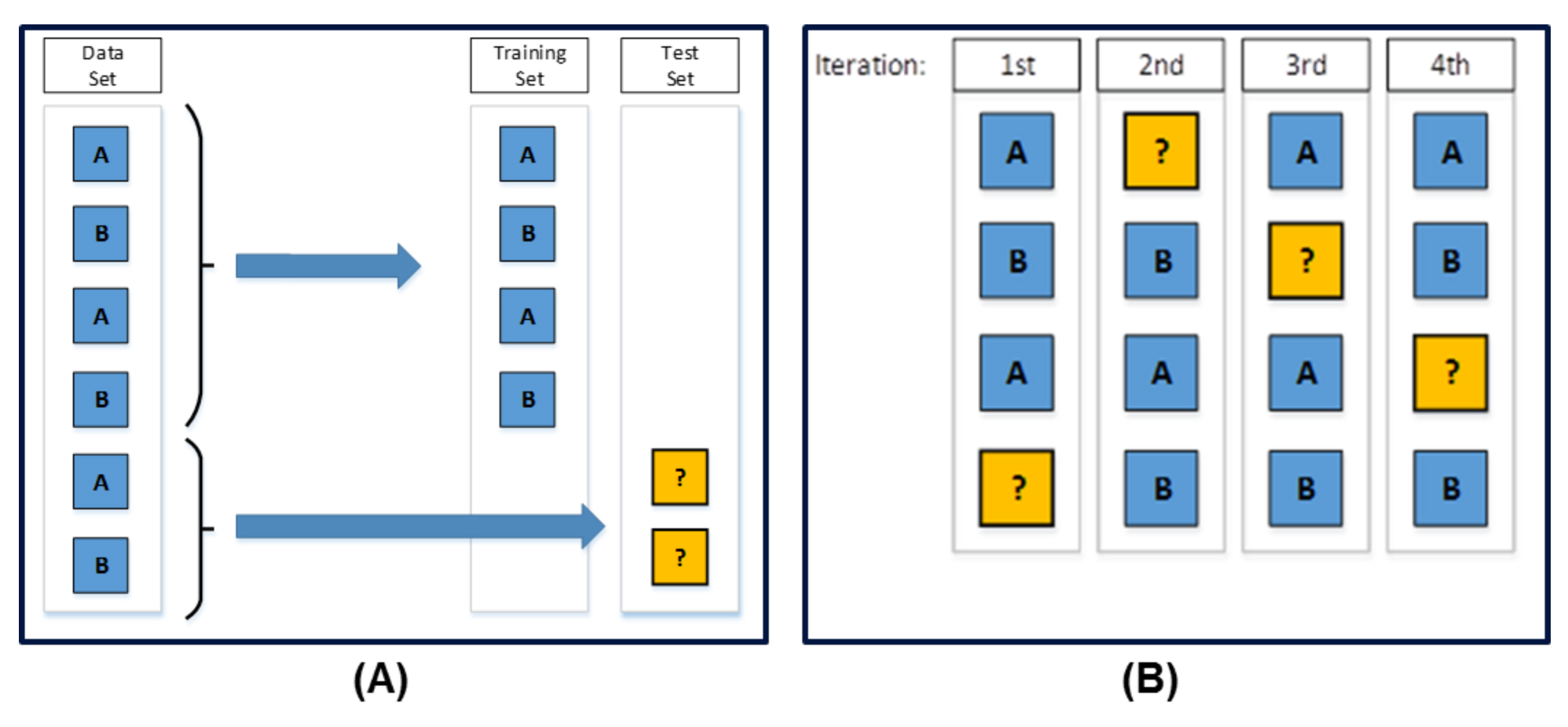
4Department of Pathology and Laboratory Medicine, Weill Cornell Medical College, New York, NY.

5 Genitourinary Oncology Service, Department of Medicine, Memorial Sloan Kettering Cancer Center, Department of Medicine, Weill Cornell Medical College, New York, NY.

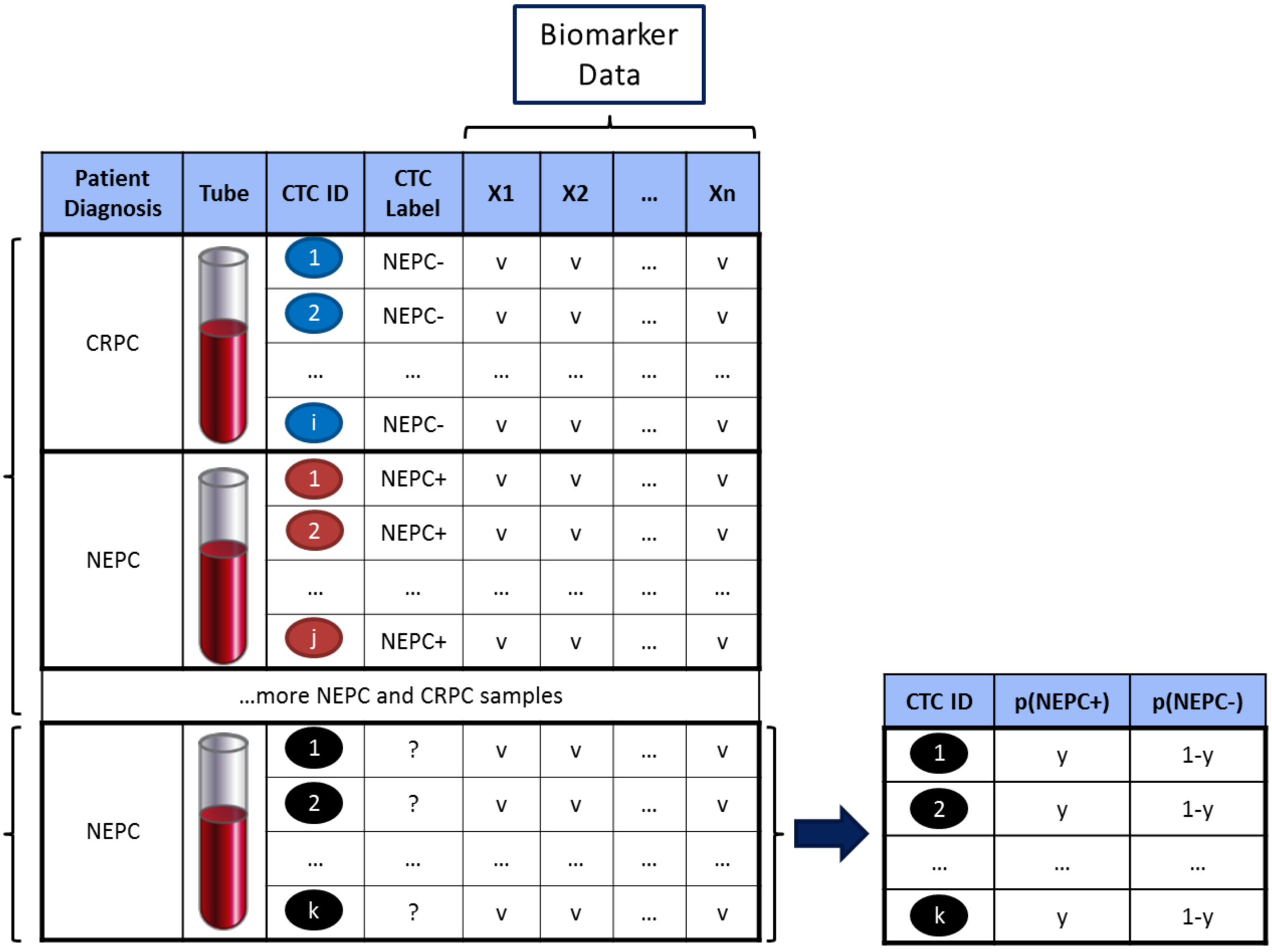
**Figure S1:** Examples of a decision boundary that separate class A (red) from class B (blue) in 1 dimension (figure 1A) and 2 dimensions (figure 1B). In three dimensions, the decision boundary would be a plane. In high dimensional space, this becomes difficult to visualize.



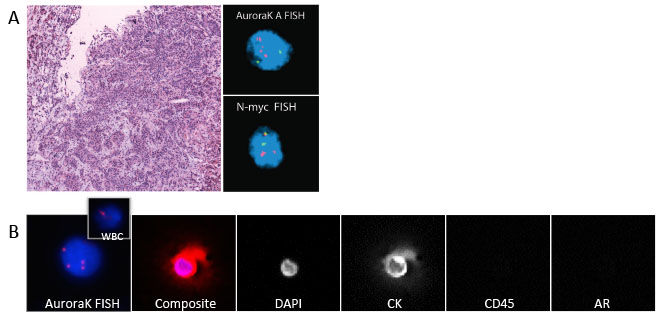
**Figure S2:** Schematic of the supervised learning process (A) and leave one out cross validation (B). During each iteration, a dataset of examples labeled with their class membership (class A, class B) is partitioned into a training set (blue) and a test set (orange).



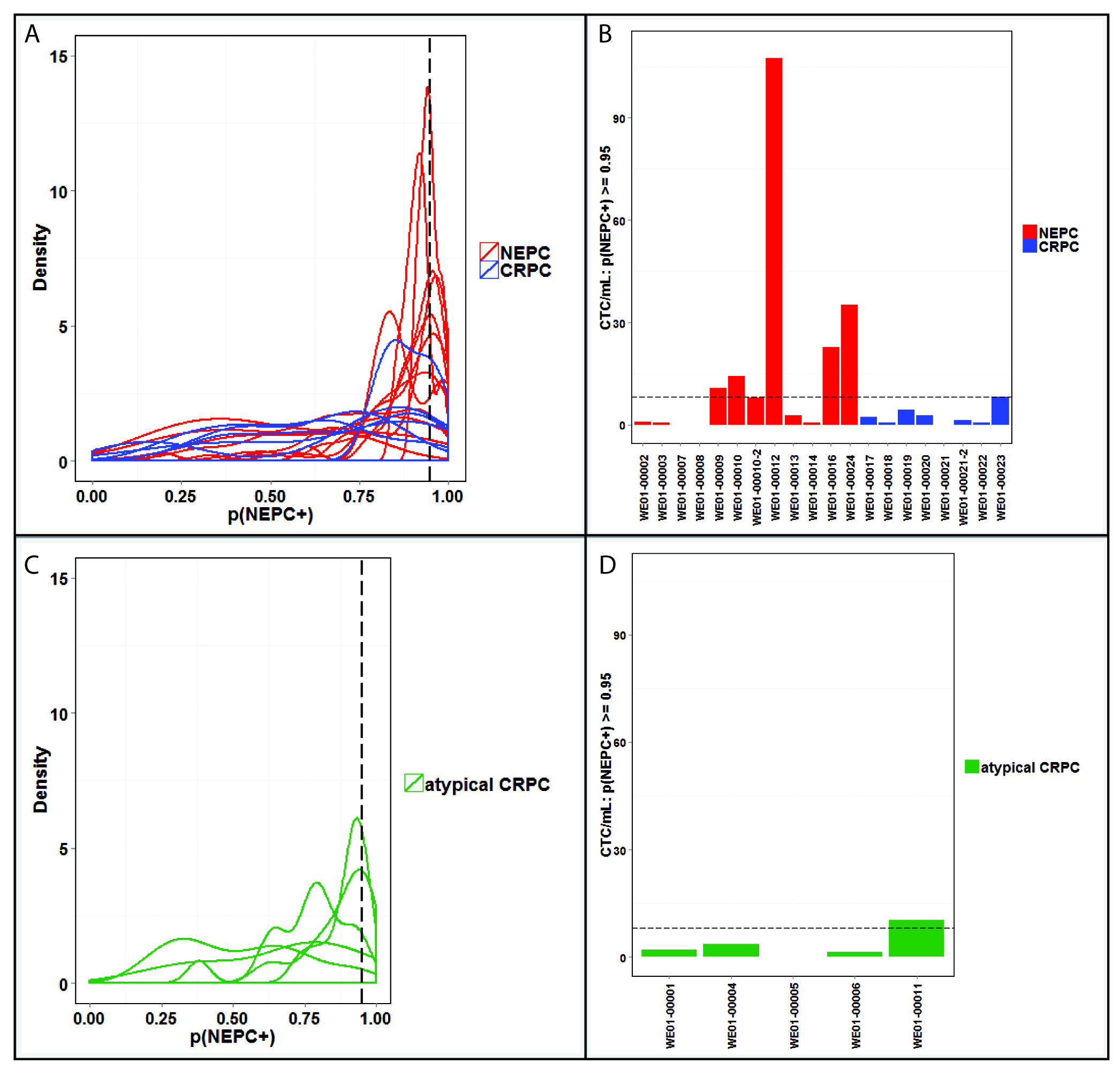
**Figure S3:** Single iteration of Leave-One-Out Cross-Validation performed at the blood sample level. CTCs from every other sample are partitioned as the Training Set, and presented to the classifier as labeled examples. CTCs from the sample held-out as the test set are then analyzed by the trained classifier, which provides an estimated probability of class membership to NEPC+ or NEPC- for each CTC (right).



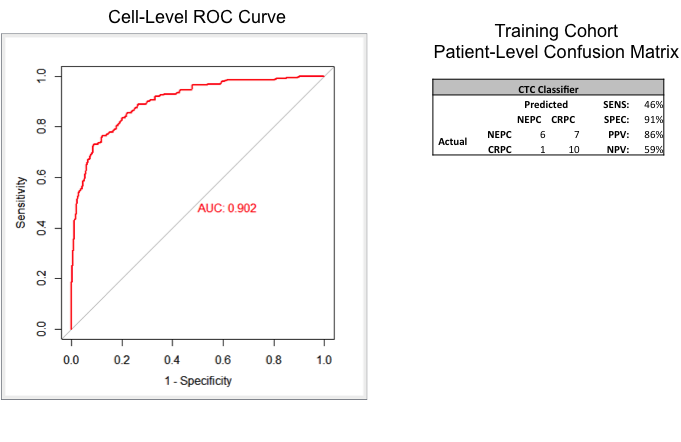
**Figure S4: An example of concordance of AURKA amplification in tumor and CTCs. A)** Palliative transuretheral resection of prostate (TURP) shows AuroraK (AURKA) and N-myc genomic amplification by FISH; **B)** Patient matched circulating tumor cells (CK+, CD45-, AR-), AURKA also amplified by FISH.



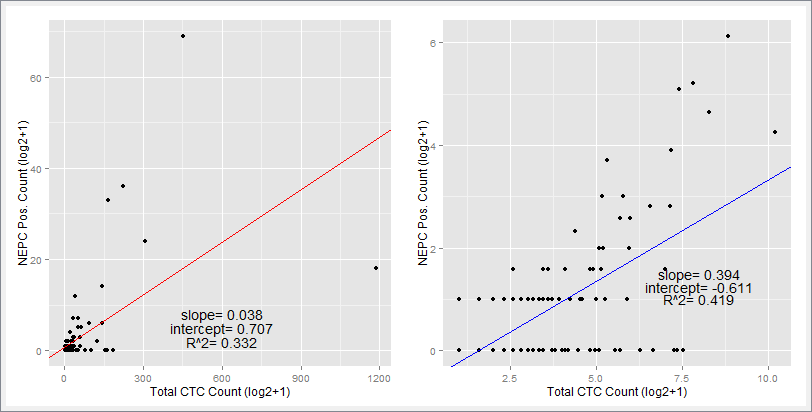
**Figure S5:** (A) Kernel Density Estimate (KDE) curves of the classifier output are plotted for each patient sample colored by their diagnosis: NEPC (red) and CRPC (blue). Note the peak in densities near the far right of the curve corresponding to high probability of NEPC class membership for patient samples diagnosed with NEPC. (B) Bar chart of the number of CTCs/mL that have an estimated probability of class membership to NEPC+ greater than or equal to 0.95. (C) NEPC and CRPC patient samples were used to train a classifier, for which the atypical CRPC patient samples were analyzed as the test set. Kernel Density Estimate (KDE) curves of the classifier output are plotted for each atypical CRPC patient. Note the peak in densities near the far right of the curve corresponding to high probability of NEPC class membership. (D) Bar chart of the number of CTCs/mL that have an estimated probability of class membership to NEPC+ greater than or equal to 0.95.



**Figure S6:** Cell-level NEPC classifier **(A)** Receiver-Operating-Characteristic (ROC) curve generated on the classifier’s single-cell output after LOOCV. (B) Confusion Matrix for the cell-level classifier’s ability to discriminate training cohort patients diagnosed with NEPC vs. CRPC.



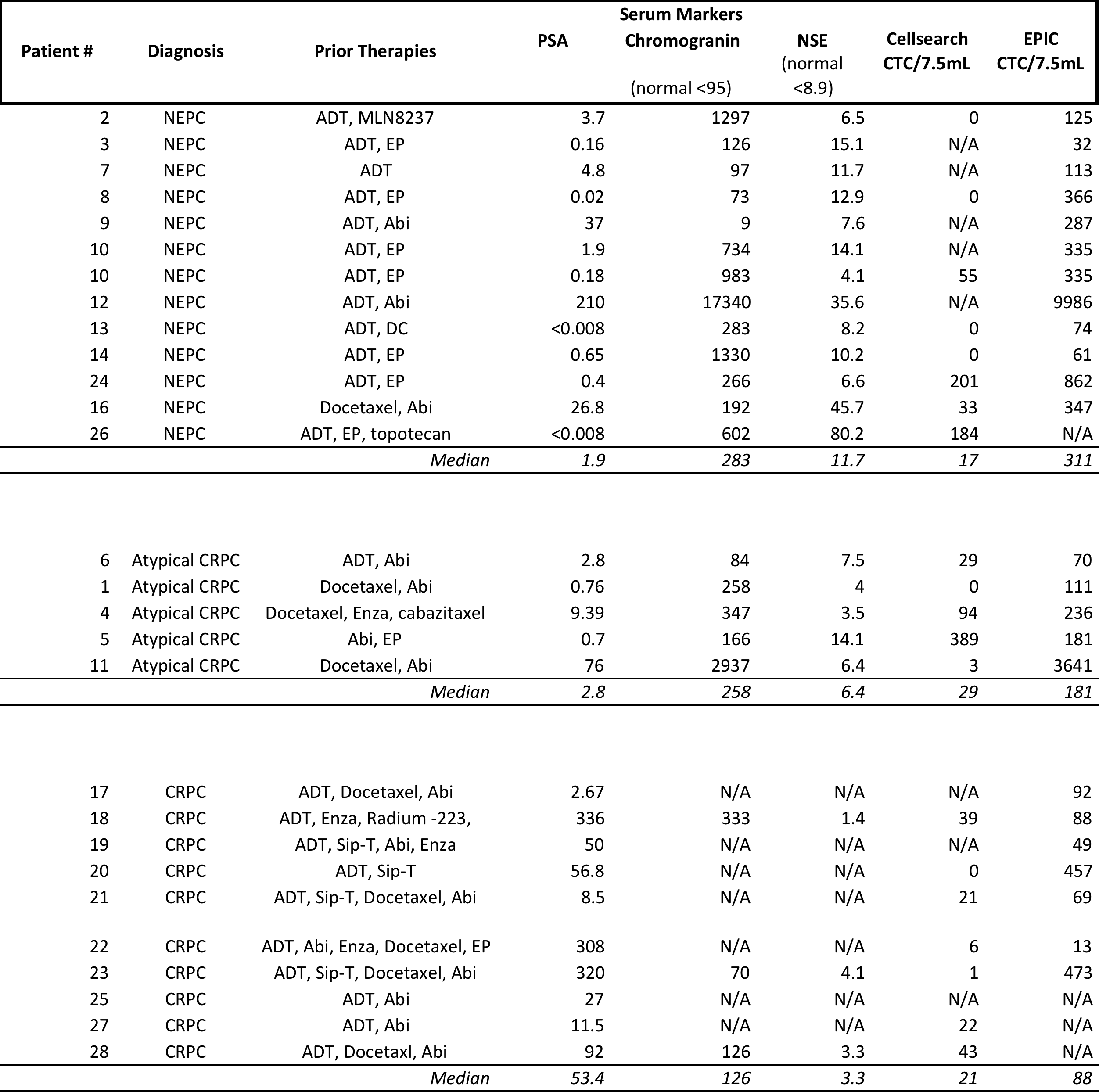
**Figure S7:** To address whether the CTC classifier is simply stochastic, a reflection of an overall higher CTC count, linearity was assessed with a Pearson’s coefficient showing a weak relationship between frequency of NEPC CTCs and total cell count.

**

**Table S1:** A summary of cell-level features utilized to train Random Forest cell-level classifiers in both the LOOCV and for the classification of CTCs in the test cohort.

|  |  |
| --- | --- |
| **Feature** | **Type** |
| **AR Expression** | **Protein Expression** |
| **CK Expression** | **Protein Expression** |
| **CTC Cluster** | **CTC Cluster (Y/N)** |
| **Nuc. Area** | **Nuclear Size** |
| **Nuc. Convex Area** | **Nuclear Size** |
| **Nuc. Major Axis** | **Nuclear Size** |
| **Nuc. Minor Axis** | **Nuclear Size** |
| **Nuc. Circularity** | **Nuclear Shape** |
| **Nuc. Solidity** | **Nuclear Shape** |
| **Nuc. Entropy** | **Nuclear Texture** |
| **Nuc. Speckles** | **Nuclear Texture** |
| **Nucleoli** | **Nuclear Texture** |
| **Cyto. Area** | **Cell Size** |
| **Cyto. Convex Area** | **Cell Size** |
| **Cyto. Major Axis** | **Cell Size** |
| **Cyto. Minor Axis** | **Cell Size** |
| **Cyto. Circularity** | **Cell Shape** |
| **Cyto. Solidity** | **Cell Shape** |
| **CK Speckles** | **CK Texture** |
| **N/C Ratio** | **Cell Size** |

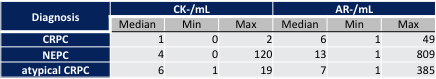
**Table S2:** Patient characteristics (discovery cohort) including prior systemic therapies, serum markers including PSA (ng/ml), Chromogranin (ng/ml), NSE (ng/ml) and CTC counts. ADT= Androgen deprivation therapy with GnRH agonist, Abi= abiraterone, Enza= enzalutamide, DC= Doctaxel- carboplatin, EP= etoposide- platinum, MLN8237= investigational agent, Sip-T= sipuleucel-T.



**Table S3:** Liver metastases in NEPC vs. CRPC

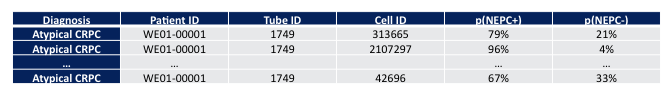
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Pathologically Defined NEPC | Clinically Defined CRPC | | |
|  | Overall N=12 | Overall N=15 | Atypical Clinical features  N=5 | No Atypical Clinical Features  N=10 |
| Liver Mets | 8/12 (66.7%) | 5/15 (33.3%) | 4/5 (80%) | 1/10 (10%) |
| P-value (NEPC to All CRPC) | 0.1283 | |  |  |
| P-value (NEPC, Atypia, No Atypia) | 0.0065 | | | |

**Table S4:** The median concentration of CK-negative and AR-negative CTCs in CRPC, atypical CRPC, and NEPC patients.



**Table S5:** Confusion matrix for the ability of CD56 staining to discriminate patients diagnosed with Small Cell Carcinoma NEPC vs CRPC

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **All Patients** | | | | | | **Cisplatin-Naïve Patients Only** | | | | | |
|  | **CD56+** | **CD56-** |  | **SENS:** | **58%** |  | **CD56+** | **CD56-** |  | **SENS:** | **67%** |
| **NEPC** | **7** | **5** |  | **SPEC:** | **100%** | **NEPC** | **4** | **2** |  | **SPEC:** | **100%** |
| **CRPC** | **0** | **8** |  | **PPV:** | **100%** | **CRPC** | **0** | **8** |  | **PPV:** | **100%** |
|  |  |  |  | **NPV:** | **62%** |  |  |  |  | **NPV:** | **80%** |
|  |  |  |  |  |  |  |  |  |  |  |  |
|  | **CD56+** | **CD56-** |  | **SENS:** | **71%** |  | **CD56+** | **CD56-** |  | **SENS:** | **100%** |
| **Small Cell NEPC** | **5** | **2** |  | **SPEC:** | **100%** | **Small Cell NEPC** | **3** | **0** |  | **SPEC:** | **100%** |
| **CRPC** | **0** | **8** |  | **PPV:** | **100%** | **CRPC** | **0** | **8** |  | **PPV:** | **100%** |
|  |  |  |  | **NPV:** | **80%** |  |  |  |  | **NPV:** | **100%** |

**Table S6:** Example of results from a single tube of blood. The column under “Actual” is the class label used for training, note that the classifier sees this during training, but not during testing. The two columns on the right are the output from classification for each CTC: p(NEPC+) being the probability of the CTC belonging to NEPC, and p(NEPC-) being the probability of the CTC belonging to CRPC.