SUPPLEMENTAL METHODS

Cell lines and cell culture

Immortalized non-transformed prostate epithelial cell line (BPH1) [1] and PCa cell lines LNCaP, PC3, NCI-H660 (CRL-5813) [2] were obtained from the American Type Culture Collection (ATCC) and cultured under recommended conditions. LNCaP-AR and LNCaP-AR-Enz resistant cell lines were a kind gift from Dr. Felix Feng at UCSF. LNCaP, PC3, LNCaP-AR and LNCaP-AR-Enz resistant cells were maintained in RPMI 1640 media each supplemented with 10% fetal bovine serum (FBS) (Atlanta biologicals) and 1% penicillin/streptomycin. Enz resistant cells were cultured in presence of 50µM Enzalutamide (Selleck Chemicals). BPH1 and NCI-H660 cell lines were maintained in RPMI 1640 media and HITEs media, respectively, each supplemented with 5% FBS, and 1% penicillin/streptomycin. All cell lines were maintained in an incubator with a humidified atmosphere of 95% air and 5% CO_2 at 37°C.

Cell line transfections

Cells were plated in growth medium without antibiotics ~24hrs before transfection with miRNA mimics/inhibitors (Ambion) or with siRNA/ORF constructs (Origene). Transient transfections of miR-375/miR-301a mimics (catalog no. 4464066) /negative control or mimic (catalog no. 4464058) or anti-miR-363 inhibitor (catalog no. 4464084)/control inhibitor (catalog no. AM17010) was carried out by using Lipofectamine 2000 (Invitrogen) per the manufacturer's protocol. Similarly, siRNA/ ORF clone transfections were carried out using control siRNA/ AURKA siRNA (Catalog no. SR321919) or control ORF/*MYCN* ORF/*AURKA* ORF/ *E2F1* ORF (Origene). All transfections were for 72h followed by functional assays. For miR-106a~363 blockade, miR-106a~363 targeting sponge that contains multiple miR bulge sites cloned into pGreen-lentiviral expression system (System Biosciences) with an RNA polymerase III-driven H1 promoter [3,4] was kindly gifted by Dr. Paul Jedlicka at University of Colorado. C42B cells were infected with miR-106a~363 targeting lentiviral/control lentiviral constructs followed by puromycin selection (1µg/ml).

Clonogenicity assays

72 hours post-transfection, cells were harvested, counted, seeded at low densities (10,000 cells/plate) and allowed to grow until visible colonies appeared. Cells were stained with Giemsa stain, colonies were counted and average colony numbers were plotted.

Migration and invasion assays

Control inserts (for migration) or Matrigel inserts (for invasion) (BD Biosciences) were used for performing *in vitro* transwell migration and invasion assays as per manufacturer's protocol. Briefly, 48 hrs post-transfection, cells were counted and 50,000 cells in a volume of 300µl serum-free medium were placed on Matrigel inserts and control inserts, respectively. Cells were allowed to migrate for 24 hrs at 37°C. Following this, cells were removed from the top of the inserts and cells that migrated/invaded though the polycarbonate/basement membrane were fixed, stained, counted under light microscope.

Immunoblotting analyses

Cells were lysed with RIPA buffer [50 mmol/L Tris (pH 8.0), 150 mmol/L NaCl, 0.5% deoxycholate, 0.1% SDS, and 1.0% NP-40] containing protease inhibitor cocktail (Roche). Protein lysates (60-80 µg) were loaded onto a 4–20% Tris-glycine gradient gel (Biorad) and Western blotting was performed as per standard protocols. The following antibodies were used: AURKA (Cell Signaling Technology, 14475), KLF4 (Cell Signaling Technology, 4038), N-Myc (Cell Signaling Technology, 84406), STAT3 (Cell Signaling Technology, 4904), E2F1 (Cell Signaling Technology, 3742) and GAPDH (Santa Cruz Biotechnology, sc-32233).

Luciferase assays

For *AURKA, KLF4, STAT3, E2F1* and *NMYC*, 3'UTR region containing target sequences complementary to the miR-363 seed sequence were cloned downstream of the luciferase gene in the pmiRGLO luciferase vector (Promega) in accordance with manufacturer's instructions. For *AURKA* and *KLF4,* mutated 3'UTR sequences were cloned in the same vector. The primers used for clonings were synthesized from Invitrogen and sequences are listed in Table S8. 3'-UTR reporter (wt/mutant) / control constructs (0.2 ug each) were transfected into LNCaP/C42B cells cultured in 24-well plates along with 50nM miR-CON/miR-363 mimics (Life Technologies) using Lipofectamine 2000 (Invitrogen). After 48 hrs, cells were harvested and firefly and renilla luciferase activities were measured by using the dual luciferase reporter assay system (Promega) as per the manufacturer's protocol. Firefly luciferase was normalized to corresponding renilla luciferase activity.

Functions, Pathway and interactive network analysis

Ingenuity Pathway Analysis (IPA 8.0, Qiagen) was used to identify the pathways and interaction networks that are significantly affected by significantly differentially expressed transcripts as per the manufacturer's instructions. The knowledge base of IPA consists of functions, pathways and network models derived by systematically exploring the peer reviewed scientific literature. It calculates a P-value for each function, pathway according to the fit of users' data to the IPA database using one-tailed Fisher exact test. The pathways with P-values <0.05 were considered significantly affected. For each network, IPA calculates a score derived from the P-value of one-tailed Fisher exact test [score = -log(P-value)] and indicates the likelihood of focus transcripts appearing together in the network due to random chance. A score of 2 or higher has at least a 99% probability of not being generated by random chance alone. The ability to rank the networks based on their relevance to the queried data sets allows for prioritization of networks with the strongest association with differentially expressed transcripts.

Generation of miRNA-based classifier

To develop an optimal classifier to differentiate between NE samples from those with adenocarcinoma histology, miRNAs detected above threshold were used as a seed set. We used a combination of techniques to mitigate the effects of the n < p data set (n=number of samples and p=number of variables). We chose to pair the random forest machine learning technique with leave-pair-out cross validation (LPOCV) because of its ability to lessen the effects of high bias/variance (30). Also, before each cross-validation iteration, Boruta feature selection (https://pypi.org/project/Boruta/) was applied to the training set of a given iteration followed by training the random forest model on the newly subsetted training set. The performance of classifier was measured using receiver operating characteristic (ROC) analysis with area under the curve (AUC) as the primary evaluation metric.

Classifier Methods

Tools used: Python – Scikit-learn, Numpy, Pandas, BorutaPy

Outline:

I. **Data Intro**
   a. Features:
      i. *Phase 0*: Original Data set – 9511 features
      ii. *Phase 1*: Filtered Data set –553 features (filtering was done based on **(1)** p-value column (0.05), **(2)** status column (Low, OK, Outlier), and **(3)** miRNA vs. ISO-miRNA.
      iii. *Phase 2*: ISO-mIR features filtered out (88 features left).
   b. Observations:
      i. 35 total observations; 25 = Control, 10 = Test
      ii. Labels were generated based on Control, Test categorization (Control = FALSE, Test = TRUE).

II. Data Cleaning

a. Filtering (as detailed in section ii of Data Intro, Dimensions)

b. Transpose dataframe

   i. The data was initially presented with features as rows and observations as columns.

III. Feature Selection

a. BorutaPy library

   i. Feature selection wrapper built around Random Forest algorithm.

   ii. Forest tree depth for BorutaPy feature selection = 6 (Suggested tree depth from documentation: 3 to 7)

IV. Classifier Setup/Cross Validation

a. Run 1: Random Forest with LPOCV

   i. Purpose:

      1. Evaluate Random Forest model performance in conjunction with LPOCV.

   ii. LPOCV setup:

      1. Compute all possible pair combinations of positive/negative classes (25 negative and 10 positive = 25 x 10 = 250 cross validation folds).
      2. Iterate through combinations one by one:
         a. Generate data subset (aka training/validation sets)
         b. Feed training set into BorutaPy feature selection.
         c. Based on features suggested by BorutaPy, subset training/validation sets again.
         d. Train Random Forest model on new training set.
         e. Run predict function on validation pair (output are prediction probabilities from Random Forest predict_proba function).
         f. REPEAT until all 250 leave pair out cross validation folds are complete.

b. Evaluation

   i. Pooled AUC Calculation

      1. Using the probability outcomes from all 250 cross validation iterations, area under the curve was calculated using Scikit-Learn's roc_auc_score metric (based on trapezoidal rule).

   ii. Average AUC Calculation

      1. Calculate area under the curve score for each validation pair.
      2. Average all 250 calculations.

   iii. Feature Importance

1. Ranking of each features' average importance across all 250 models trained.

Logarithmic Loss

Logarithmic loss measures the performance of a supervised learning classification model using ground truth labels and the predicted probabilities from the model. The goal is to minimize the log loss with 0 being a perfect score. The score increases when the predicted probability diverges away from the actual label. So, a probability of 0.25 for a label 1 (CRPC-NE) would increase model log loss score while a probability of 0.85 for a label 1 (CRPC-NE) would decrease the score. To calculate Log loss for a single observation/sample, the following equation is used:

$-(y \log(p)+(1-y)\log(1-p)$; where variable 'y' represents binary class (CRPC-Adeno=0 or CRPC-NE=1) and 'p' represents the model's predicted probability for that y.

The log loss for all observations/sample that a classifier has predicted on is the average of all single log loss calculations.

Validation of 'miRNA classifier' against Taylor *et al* dataset [5]

Since Taylor *et al* dataset [5] dataset did not include expression for nine of the miRNAs included in our classifier (Table S6), we subsetted our 'miRNA classifier' excluding these missing features (Fig. S6A). The performance of this 'subsetted classifier' was re-evaluated by ROC analyses that confirmed high AUC of 0.97. The 'subsetted classifier' was then applied to Taylor *et al* dataset [5] (Fig. S6B) using single model prediction (Fig. 4E).

SUPPLEMENTAL FIGURE LEGENDS

Fig. S1 Histology of CRPC-Adeno and CRPC-NE tissues

Representative H&E staining for sequenced CRPC samples used in the study: adenocarcinoma without NED (left panel) and with NED (right panel).

Fig. S2 Average distribution of small RNA reads in sequenced samples (related to Fig. 1)

Average distribution of small RNA reads from small RNA sequencing analyses of CRPC-Adeno and CRPC-NE clinical samples, PDX and cellular models.

Fig. S3 Principal Component Analyses for CRPC-Adeno and CRPC-NE based on piRNAs

Unsupervised principal component analyses (PCA) based on differential expression of piRNAs, as performed in CRPC-adeno cases (n=25; clinical tissues, n=21 + PDX models AdLuCaP-70, -78, -81) and CRPC-NE models (n=10; clinical tissues, n=6 + PDX models LuCaP-49, -145.1, -145.2 + NCI-H660 cell line).

Fig. S4 Differential expression of microRNA isoforms (iso-miRs) in neuroendocrine differentiation states in CRPC (related to Fig. 3)

A. Range of length of differentially expressed iso-miRs (14-45 nucleotides) and their abundance across sequenced CRPC-Adeno and CRPC-NE samples.

B. Random forest machine learning technique with leave-pair-out cross validation as applied to the NGS dataset of analyzed NE tissues + PDX models+ NCI-H660 cell line (CRPC-NE, n=10) vs those with adenocarcinoma features (CRPC-Adeno, n=25) including miRNA isoforms. miRNAs are listed in the order of feature importance.

Fig. S5 Performance of miRNA classifier in validation cohort of mCRPC patients

A. Heat map showing differentially expressed miRNAs in a validation cohort of human metastatic CRPC clinical samples (n=20).

B. Single model prediction algorithm for 'miRNA classifier data' from cohort 1 as applied to NGS data for validation cohort,

C. ROC curve analyses showing the ability of 'miRNA classifier' to distinguish between class 0 (CRPC-Adeno) and class 1 (CRPC-NE) in validation cohort.

Fig. S6 Application of 'miRNA classifier' to Taylor *et al.* dataset [5] of prostate adenocarcinomas

A. For application of the classifier to Taylor *et al* dataset [5] of primary and metastatic prostate cancer cases, we subsetted our classifier as this dataset did not include expression for nine of the miRNAs. Inset: ROC curve analyses to evaluate the performance of this 'subsetted classifier'.

B. The 'subsetted classifier' was then applied to Taylor *et al* dataset [5] using single model prediction algorithm. miRNAs are listed in the order of feature importance.

SUPPLEMENTAL TABLE LEGENDS

Table S1 Clinicopathologic characteristics of metastatic CRPC patients (Cohort 1+2)

Table summarizing the age, race, Gleason score of primary tumor, final serum PSA, metastatic sites and prior therapies of clinical cohorts 1 and 2 used in the study.

Table S2 MicroRNA expression in CRPC Adeno vs CRPC NE tissues (cohort 1) and PDX models

Table showing significantly dysregulated miRNAs identified by sequencing of CRPC-Adeno vs CRPC-NE clinical samples (cohort 1), PDX models and NCI-H660 cell line.

Table S3 MicroRNA families altered in CRPC-NE vs CRPC-Adeno (cohort 1)

Table showing significantly dysregulated miRNA families identified by sequencing of CRPC-Adeno vs CRPC-NE clinical samples (cohort 1), PDX models and NCI-H660 cell line.

Table S4 List of miRNAs in 'miRNA classifier including iso-miRs' to distinguish between CRPC-Adeno and CRPC-NE cases

Random forest machine learning technique with leave-pair-out cross validation was applied to the NGS dataset of analyzed NE tissues (cohort 1) + PDX models+ NCI-H660 cell line (CRPC-NE, n=10) vs those with adenocarcinoma features (CRPC-Adeno, n=25) (cohort 1) that included miRNA isoforms. miRNAs obtained in the classifier are listed in the order of feature importance.

Table S5 MicroRNA expression in CRPC Adeno vs CRPC NE tissues (cohort 2)

Table showing significantly dysregulated miRNAs identified by sequencing of CRPC-Adeno vs CRPC-NE clinical samples in validation cohort (cohort 2).

Table S6 List of miRNAs in 'miRNA classifier' missing in Taylor *et al* dataset

To further validate our miRNA classifier obtained from sequencing of CRPC-Adeno vs CRPC-NE cases (cohort 1), we applied the classifier data to Taylor *et al* dataset [5] (Fig. S5B-D). The list of miRNAs in this table represent the miRs included in our classifier that are missing in Taylor *et al* dataset [5].

Table S7 Dysregulated miRNAs target key genes in 'NEPC gene signature' and 'androgen receptor signaling'

List of reciprocal miRNA/mRNA pairings between identified dysregulated miRNAs and NEPC/AR gene signature from Beltran *et al.* [6] study as determined by IPA.

Table S8 List of primers used for 3'UTR clonings

For *AURKA* (1 site) and *KLF4* (2 sites), 3'UTR regions containing target sequences complementary to the miR-363-3p seed sequence were cloned downstream of the luciferase gene in the pmiRGLO luciferase vector (Promega). Mutated 3'UTR sequences complementary to miR-363-3p (indicated in Fig. 9E) were cloned in the same vector.

REFERENCES

1 Hayward, S. W. *et al.* Establishment and characterization of an immortalized but non-transformed human prostate epithelial cell line: BPH-1. *In vitro cellular & developmental biology. Animal* **31**, 14-24, doi:10.1007/BF02631333 (1995).

2 Lai, S. L. *et al.* Molecular genetic characterization of neuroendocrine lung cancer cell lines. *Anticancer Res* **15**, 225-232 (1995).

3 Dylla, L. & Jedlicka, P. Growth-promoting role of the miR-106a~363 cluster in Ewing sarcoma. *PLoS One* **8**, e63032, doi:10.1371/journal.pone.0063032 (2013).

4 Ebert, M. S., Neilson, J. R. & Sharp, P. A. MicroRNA sponges: competitive inhibitors of small RNAs in mammalian cells. *Nat Methods* **4**, 721-726, doi:10.1038/nmeth1079 (2007).

5 Taylor, B. S. *et al.* Integrative genomic profiling of human prostate cancer. *Cancer Cell* **18**, 11-22, doi:10.1016/j.ccr.2010.05.026 (2010).

6 Beltran, H. *et al.* Divergent clonal evolution of castration-resistant neuroendocrine prostate cancer. *Nat Med* **22**, 298-305, doi:10.1038/nm.4045 (2016).