

Supplementary Methods

Samples.

Microarrays (TMA) consisting of 0.6 mm core biopsies of formalin-fixed, paraffin-embedded tissues were created using a tissue arrayer (Beecher Instrument). Cohorts of 59 anonymized prostate samples present in the TMA in triplicate were analyzed. Of the 59 patients, 20 were Prostate Adenocarcinoma (PCa), 12 Benign Prostatic Hyperplasia (BPH), 13 Prostatic Intraepithelial Neoplasia (PIN), 14 Prostate Metastasis (PCaMet).

Oligonucleotide Probe Synthesis.

Four 50-mer DNA probes were synthesized for each target transcript. Sequences were designed with OLIGO-6.0 (Molecular Biology Insights) and verified for specificity through the BLAST at NCBI GeneBank. Amine-modified Thymidine bases were used as substrates for amidation using commercially available succinimidyl ester conjugates of fluorescent dyes. For experiments shown here the fluorophores were Cy3, Cy3.5, Cy5 and Cy5.5 (Amersham).

Barcoding

Each of the four dyes was viewed using the appropriate filter set (Chroma) using an Olympus BX51 with a UPlanApo 1.4 60x objective and a Roper CoolSnap camera. Each set of labeled probes was mixed to achieve the appropriate barcode: EBP4.9: Cy3 & Cy5; JAG-1: Cy3.5 & Cy5; PSMA: Cy3.5 & Cy3; AMCAR: Cy3.5 & Cy5.5; AR: Cy3, Cy3.5, Cy5. For the Chromosome FISH: Chromosome Cy3 & AR probe Cy3.5 & Cy5

Experimental Design

The inability to detect a given transcription site is a reflection of a number of components including on/off gene function, low signal to noise ratio and the qualities of the tissue section including anatomic location and section thickness..

In order to address this issue, we experimentally employed multiple probes for a single housekeeping gene / transcription site, with the intention of increasing the signal to noise ratio as well as determining the overall sensitivity and specificity of the assay.

We selected the housekeeping gene, SMG-1, a phosphatidylinositol 3-kinase-related protein kinase for these initial studies. A series of 41, 16, 8, 4 and 3 unique 50mer oligonucleotides were generated across the entire gene and individually labeled with Cy3 and Cy5, in order to generate 82, 32, 16, 8, and 6 independent probes, respectively.

Tissue-FISH.

Briefly, Tissue Microarray sections were baked in a dry oven at 65°C for 30 minutes and subsequently placed in a de-cloaking buffer used as a de-

paraffinization and antigen retrieval agent for approximately 30 minutes. After 5 minutes of cooling at room temperature followed by three washes in tap water and two in PBS, the sections were further processed through successive treatments in order to reduce auto-fluorescence: ammonia-70%alcohol for 20 min. followed by sodium borohydride for 40 min. Autofluorescence at all wavelengths was observed to be reduced by more than 50%. Pre-hybridization was performed at room-temperature in 50%formamide/2XSSC for a minimum of 15 minutes followed by hybridization with 20ng of probe for 2 hours at 37°C in a humidified chamber in the dark. Post-hybridization washes were conducted at room temperature on a shaker in the dark as follow: 50%formamide/2XSSC for 20 minutes, 2XSSC, 1XSSC and 0.5XSSC for 15 minutes each. After a brief wash with PBS the slides were counterstained with DAPI and mounted with Antifade (Vysis).

Image Acquisition and Analysis.

Three-dimensional image data were acquired using an BX61 microscope (Olympus) with a CoolSNAP HQ Charge-Coupled Device (CCD) camera (Photometrics) using IPLab Software Windows version 3.6.1 (Scanalytics). We used UPlanApo 20X, 40X 1.0 NA and 60X 1.4 NA objectives (Olympus) and HiQ Band pass filters for DAPI, FITC, Cy3, Cy3.5, Cy5, Cy5.5 (SP filter set, Chroma Technology).

Slides were evaluated at 20X magnification to review overall section quality and at 40X and 60X for the distribution of the fluorescence signal (nascent transcript spots). To be scored as a true transcription site, our experimental strategy required simultaneous detection of at least two fluorescent dyes at each site, independent hybridization events were highly unlikely to be attributable to random binding of two fluorochromes. Only sites within nuclei were counted. The number of nuclei in each field examined were also counted, including those in regions of interest where transcription sites were not detected, to determine the relative frequency of occurrence of these sites.

Transcription sites were detected with software that analyzed the image data collected in three dimensions, from six independent fluorescence channels. The data collected from the Cy3, Cy3.5, Cy5 and Cy5.5 fluorescence channels corresponded to the dyes used to label the FISH oligo probes. The data collected in the FITC channel were representative of auto-fluorescence in the specimen, as no FISH probe was labeled with a dye that emitted in this channel. The data collected in the DAPI channel provided the boundaries of nuclei in the specimen. The FITC, Cy3, Cy3.5, Cy5 and Cy5.5 image data were convolved using a classic 4x4 spatial high pass filter and then segmented using a threshold with morphological constraints on object volume and shape. The objective of

these steps was to reject fluorescing artifacts and high intensity voxel noise. The segmented data from each of these fluorescent channels were then binarized. The DAPI image data were thresholded, interior holes in segments were filled and the data were binarized, which provided a three dimensional binary map of nuclear positions. The binarized data from the Cy3, Cy3.5, Cy5 and Cy5.5 channels were masked using the binarized nuclear data multiplied by the inverted binarized FITC data. This yielded binarized image data in Cy3, Cy3.5 and Cy5 that represented nuclear FISH signal that was not coincident with autofluorescence. Using simple binary operators to interpret the FISH barcode scheme, the Cy3, Cy3.5, Cy5 and Cy5.5 resultant image data were combined into a single image volume in which each binarized object was encoded with a gene identity. The centroid of each encoded object was reported in a tabular text file and marked and numbered with a pseudo color and number assigned to each gene on a blue pseudo colored, two-dimensional maximum projection of the acquired DAPI image data.

Patient analysis by microarray.

Random fields were chosen from each individual from tissue microarray (TMA) blocks representing tissue samples derived from patients that had undergone therapeutic and/or diagnostic procedures for PCa at both Memorial Sloan-Kettering Cancer Center, NY, NY and Histopat Laboratories, Barcelona, Spain. The individual patient samples were selected from several TMAs and represent PCa progression series exhibiting a full range of clinical and pathologic patient features.