TITLE: Role of Mitochondrial Inheritance on Prostate Cancer Outcome in African American Men

PRINCIPAL INVESTIGATOR: Martha Mims

CONTRACTING ORGANIZATION: Baylor College of Medicine
Houston, TX 77030-3411

REPORT DATE: November 2016

TYPE OF REPORT: Addendum to Final

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland  21702-5012

DISTRIBUTION STATEMENT: Approved for Public Release;
Distribution Unlimited

The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision unless so designated by other documentation.
We are examining the hypothesis that mitochondrial inheritance plays a significant role in aggressiveness of prostate cancer in African Americans. In the first year of the project we identified 2,000 noncancerous tissues samples from African American men with prostate cancer and we have extracted DNA from ~ 1,500 of them. We have validated a robust new DNA sequencing technique developed by our collaborator using single amplicon long-range PCR that permits deep coverage (10,000-20,000X on average) of the mitochondrial genome. We have sequenced 652 samples derived from frozen fully using this technology. The additional DNA samples derived from paraffin embedded (FFPE) tissue were more challenging, but have now been sequenced. Mapping of DNA variants in our sequenced genomes to mitochondrial genes is nearly complete and examples of the data we have obtained are presented here. After months of testing, we determined that the ethidium bromide methodology for generating prostate cancer cell line cybrids was not effective and we have instead used a Rhodamine-6-G procedure. PNT1A cybrid cell lines have been generated and we have characterized these cells. PC3 cells have been more difficult, but we continue to work on this cell line. We anticipate the first manuscripts relating to the large amount of data generated in this project will be written and submitted in the spring of 2016. To our knowledge this represents the largest collection of mitochondrial sequence data from African Americans in existence.
# Table of Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Introduction</td>
<td>4</td>
</tr>
<tr>
<td>2. Keywords</td>
<td>4</td>
</tr>
<tr>
<td>3. Accomplishments</td>
<td>4</td>
</tr>
<tr>
<td>4. Impact</td>
<td>16</td>
</tr>
<tr>
<td>5. Changes/Problems</td>
<td>17</td>
</tr>
<tr>
<td>6. Products</td>
<td>17</td>
</tr>
<tr>
<td>7. Participants</td>
<td>18</td>
</tr>
<tr>
<td>8. Special reporting Requirements</td>
<td>18</td>
</tr>
<tr>
<td>9. Appendices</td>
<td>18</td>
</tr>
</tbody>
</table>
1. Introduction:
African American men are disproportionately affected by prostate cancer with increased lifetime risk, earlier onset of disease and more advanced stage at diagnosis than Caucasians. Prostate cancer is the leading cause of cancer death in African American men, with mortality more than double that observed in Caucasians. There are nearly 20 million African American men in the US, many of whom face significant risk of developing and dying from prostate cancer. Risk for and aggressiveness of prostate cancer in African American men is thought to originate in part from genetic susceptibility. Several nuclear genes and chromosomal regions have been linked to prostate cancer; however, many studies have not included African American men, and no study has linked genetic polymorphisms with clinical outcome. One factor which has not been carefully examined is mitochondrial inheritance which varies significantly between ethnic and racial groups and could explain large differences in disease characteristics. In the present study, we are examining the hypothesis that mitochondrial inheritance plays a significant role in aggressiveness of prostate cancer in African Americans. Further, we predict that adverse clinical outcomes will be reflected in dysregulation of cellular biochemical processes and in alterations in signaling pathways (Akt pathway, apoptosis). The ability to identify mitochondrial variants or haplogroups that contribute to aggressive disease will help to separate patients with indolent disease who can be spared unnecessary intervention from those who need more immediate and aggressive therapy. Identification of cellular pathways involved will help to target treatment strategies for those with cancers predicted to be more aggressive. In Aim 1 we will conduct a study of mitochondrial inheritance in 1,000 African American men with prostate cancer. We will sequence the mitochondrial genome of all 1,000 samples and determine whether particular mitochondrial variants, genes or haplogroups are associated with markers of aggressive disease (age at diagnosis, stage at diagnosis (including bone metastasis), Gleason score, PSA at diagnosis, PSA recurrence and death from disease). Findings will be replicated using an independent set of 1,000 patients from our own tissue resource with linked clinical data. Our strategy will be to sequence the mitochondrial DNA of all 1,000 patients in the discovery phase as well as in the replication group. We will control for population admixture using the Illumina African American Admixture Panel. Using cybrid technology, we will introduce our previously identified mt10398A variant and mitochondrial variants associated with highly aggressive and least aggressive disease identified in our genotyping study into prostate cell lines (derived from normal and cancer cells). Cybrids differing only in their mitochondrial composition will be examined for viability under metabolic stress, cell cycle distribution, production of reactive oxygen species, O_2 consumption, ATP synthesis, respiratory chain activity and capacity to grow in an anchorage independent manner. The effect of mitochondrial variants on nuclear gene expression will be studied using Western blotting and microarrays.

2. Keywords: Mitochondrial DNA sequencing, Prostate Cancer, Cybrid Cells, African Americans

3. Accomplishments:
In the text to follow, we provide the Aims and original Statement of Work in italics with progress on the project presented in regular text. This is an update of the work to date including this last year which was after funding for the project was completed.

Aim 1: To examine the association between mitochondrial DNA variants and clinical outcome in African American men with prostate cancer.

Task 1: Extraction of all DNA for Initial Study and Validation Set
1a. Complete extraction of DNA/ quantitation of DNA for mitochondrial sequencing
   (Extraction ongoing at present, expect more than 1,000 samples to be prepared before the project is funded. Additional 1,000 to be extracted for validation set
   IRB protocol already approved)

We have extracted DNA on more than 1500 Tissue samples from African American Men with Prostate cancer. Work was initially slower than we had initially planned because of difficulties with extracting adequate DNA from some of the FFPE (paraffin-embedded) samples. After significant technical difficulties we have now
extracted adequate quantities of DNA from more than 1000 FFPE samples and 652 Frozen tissue samples and these samples have been sent for sequencing.

**Task 2: Mitochondrial Sequencing**

2a. **Order and test overlapping mitochondrial primers with FFPE DNA**
   
   to be sure that all primers work with FFPE tissue DNA
   
   **Ongoing – Month 3**

2b. **Redesign/test and mitochondrial primers which do not give good PCR results in 2a.**
   
   **Months 4-6**

2c. **Establish Database for mitochondrial sequences**
   
   **Months 1-5**

2d. **PCR mitochondrial sequences for primary study and validation set**
   
   **Months 6-20**

2e. **Sequence mitochondrial PCR products**
   
   **Months 6-20**

   *(Illumina Admixture Genotyping to be supported by Helis funds, but performed simultaneously)*

As reported in our initial project proposal, based on published sequences for mitochondrial PCR primers\(^1\),\(^2\) we designed 61 pairs of overlapping primers to amplify the entire 16.6-kb mitochondrial genome. In order to test this technology, we used 15 matched DNA samples derived from 5 patients – frozen tissue, FFPE tissue and whole genome amplified DNA samples derived from FFPE DNA for each patient. Overall 2/3 of the amplicons had reads in both directions on all 15 samples (including WGA DNA). Five of 61 amplicons failed completely and will be redesigned. Thus, we covered 93% of the genome by reads in at least one direction in 90% of the samples, and by reads in both directions in 84% of the mt genome, an excellent outcome for a first pass analysis over a target. Using these results, we identified 165 variants as compared to the published mitochondrial sequence. Within these variants there was 96-100% concordance of calls for each patient across the three sample types. We have continued to optimize the primers for the Sanger sequencing protocols, but as we reported in our initial application, we have also continued to test new technologies for mitochondrial sequencing. As we reported last year, we have obtained excellent results using a new technique developed by one of the collaborators on this project (Dr. Lee-Jun Wong). This technique enriches the entire human mitochondrial genome by a single amplicon long-range PCR followed by massively parallel sequencing.\(^3\) This protocol utilizes less than 100 ng of tissue DNA and makes possible a one-step approach to provide quantitative base calls, exact deletion junction sequences and quantification of deletion heteroplasmy. As we reported last year, this strategy worked well for our frozen tissue DNA and we have now sequenced 652 samples using this technology.

We were not sure that this strategy will be effective in our FFPE DNA samples. Fortuitously we had a series of Frozen and FFPE samples derived from the same subject which we could use for comparison purposes. In the figures below we show the result from 4 such Frozen/FFPE pairs. To sequence FFPE samples, we used a sequence-specific capture approach followed by Illumina sequencing. We utilized a custom capture reagent (Roche NimbleGen) targeting the hg19 mitochondrial genome (16.6 Kb). Libraries of ~200 bp insert sizes were prepared from genomic DNA (500 ng) using HGSC Illumina WES protocol and co-captured at 24-48 plex using full-length blockers to capture reads at higher on-target rates. Libraries were pooled and sequenced on a single Illumina HiSeq 2500 lane (~30 Gb). We estimated that each sample would yield 312 Mb of sequence, which roughly would amount to 8000X depth across the target regions. This is a one third less average coverage to the 12000X coverage data currently generated for frozen tissue DNA samples using the long range PCR protocol, where 80 samples are sequenced per lane on Illumina HiSeq. As seen in the figures below, these estimates were relatively accurate with regard to coverage for the FFPE samples. We expected that duplicate rates, which are usually elevated in small target region captures, would not be an issue in this case, as mitochondrial DNA copy number is 1-2 log orders greater than that of the genomic DNA. Non-specific capture of Pseudogenes in the nuclear genome that share high sequence similarity with functional mtDNA genes is another valid concern when using the above described target enrichment approach. However, recent studies have shown that it is possible to bioinformatically determine the limit on heteroplasmy detection due to such contamination and largely eliminate this concern.\(^4\)
Figure 1A shows results from four samples sequenced by the long range primer technology. We now have sequenced 652 samples using this technology. As seen in figure 1A, a distinct advantage provided by amplification of the entire mitochondrial genome by long range PR with a single primer pair is the uniform coverage. This figure demonstrates that we routinely had coverage depth of 10,000-20,000-fold for the mitochondrial genome. This sequencing strategy has multiple advantages including ease of excluding nuclear DNA sequences which are nearly identical to mtDNA and more uniform coverage of the mitochondrial genome because there is much less risk of having a rare or novel variant at the primer binding site when only one set of primers is used. Deep coverage allows much more accurate base calls and improves the limit of detection for heteroplasmy from about 15% with Sanger sequencing to ~ 1-2% with the long range primer technique. Figure 1B shows sequence data from FFPE DNA samples derived from the same 4 patients; these sequences have been performed in duplicate (data not show) with similar results. Note that as we predicted, coverage depth for the FFPE DNA samples was more like 8000X across the target region. Some question has arisen as to why the depth of coverage is not uniform over the entire mitochondrial genome (sawtooth pattern). We feel this may relate to the GC content of the mitochondrial genome which demonstrates a very similar sawtooth pattern, but we do not feel it will affect our overall results.

**Figure 1: Mitochondrial DNA sequence data**

Figure 1A: Panels A-D show representative coverage for 4 different DNA samples across the entire 16,569 bp mitochondrial genome. The y axis represents fold coverage at each position (x coordinate) along the mitochondrial genome.
Using this data we were able to compare SNP counts in Frozen vs FFPE DNA samples. These data (figure 2) demonstrate that similar numbers of SNPs were called in Frozen and FFPE samples. Figure 3 shows that these SNPs were not completely concordant and we are working on this issue as noted below in the updates from this last year.

Figure 1B: Panels A-D show representative coverage for 4 different FFPE DNA samples across the entire 16,569 bp mitochondrial genome. Note that these 4 samples are matched to the Frozen DNA samples in panel 1A. The y axis represents fold coverage at each position (x coordinate) along the mitochondrial genome.
Figure 2  Mitochondrial SNP counts in Frozen and FFPE DNA samples derived from the same patient

SNP counts

Figure 3  Mitochondrial SNP comparison in Frozen and FFPE DNA samples derived from the same patient

SNP comparison (FFPE vs Frozen DNA)
Task 3: Analysis of Mitochondrial Data

3a. Analyze mitochondrial sequences and integrate with clinical data  
   Months 20-28
3b. Admixture analysis based on Illumina data  
   Months 20-28
3c. Identify mitochondrial variants of interest for cybrid study and match with lymphoblast bank  
   Months 26-28

Figure 2A demonstrates the single nucleotide variant (SNV) calls for the first approximately two hundred subjects. On average each subject has about 150 SNVs with the Cambridge mitochondrial sequence used as reference.\textsuperscript{5,6} As shown in our previous report, most of the data at each coordinate clustered at 0 or 1. Since the mitochondrial genome is essentially haploid this suggests there is little contamination with nuclear DNA sequences. Similar data for insertions and deletions is shown in figure 2B.

**Figure 2:** Single nucleotide variants in the mitochondrial genome for the first sequenced subjects.

**Figure 2A:** Single nucleotide variants in the mitochondrial genome for the first sequenced subjects.

**Figure 2B:** Insertions and Deletions in the mitochondrial genome for the first sequenced subjects.
We have now created a database of the sequence data which will ultimately include not just mtDNA sequence data, but nuclear SNPs gleaned from the Illumina Race/Ethnicity panel (see below). We have mapped these sequences to the mitochondrial genome (see figure 3 below) to demonstrate the distributions of the mutations found in the various mitochondrial genes. Mutations below the 0.01 variant ratio have been filtered out. Figure 3A shows the SNVs in the mitochondrial genes. As expected the most mutations are seen in the D-loop, but a significant number of additional mutations are found in other regions of the genome as well including the coding regions of for a number of components of the electron transport chain. Similar data in Figure 3B shows the distribution of insertions and deletions across the mitochondrial genes.

**Figure 3:** Distribution of variants found in mitochondrial genes for the initial subjects.

![Figure 3](image)

**Figure 3A:** Distribution of mitochondrial single nucleotide variants found within various mitochondrial genes for the initial subjects as compared to the Cambridge reference sequence. The Y axis represents the absolute number of variants observed within each gene.

![Figure 3A](image)

**Figure 3B** Unique insertions and deletions by gene across the initial patients as compared to the Cambridge reference sequence. The Y axis represents the absolute number of insertions/deletions observed within each gene.

![Figure 3B](image)
Although not covered by this project, we have also sent more than 300 samples for genotyping on the Illumina African American Admixture panel. Thus far all samples have been successfully genotyped.

**Aim 2: To understand the role of mitochondrial DNA variants in regulation of cellular processes important in the cancer cell including those involved with generation of reactive oxygen species and energy metabolism.**

**Task 1: Creation of Cybrid cell lines**

1a. IRB submission for creation of lymphoblastoid lines from patients  
Months 1-4

1b. Creation of immortalized lymphocyte bank for use in study  
Months 5-24

1c. Order/test primers for variable number tandem repeat studies of nuclear origin  
Months 6-12

1d. Deplete prostate cancer cells of mitochondria  
3 potential cell lines to be used: PNT1A, LNCaP or PC3  
Months 9-24

1e. Create cytoplasts and fuse with prostate cancer cells devoid of mitochondria to create cybrid cell lines  
Months 20-36

1f. Confirm nuclear origin of cybrid cell lines using variable length tandem repeat in insulin receptor and other genes  
Months 20-36

**Task 2: Testing of cybrid cells lines**

2a. Proliferation and cell cycle testing, Oxygen consumption, Production of reactive oxygen species, electron transport chain analysis, analysis of anchorage-independent cell growth.  
Months 22-40

Initially we encountered a number of technical difficulties during the generation of cybrid models using the ethidium bromide (EBr)-mediated mitochondrial DNA (mtDNA) depletion method. During the past year, we continued our original cybrid protocol using the EBr method by varying treatment conditions. Unfortunately, despite our best efforts, quantitative real time PCR (qPCR) analysis suggested a lack of complete mtDNA depletion in these cell lines (rho zero cells). Without the rho zero cells, we could not generate cybrids in prostrate cancer models using the original protocol. To overcome these difficulties, we tested multiple alternative approaches including rhodamine-6G (R6G) mediated short-term mitochondrial dysfunction in generating rho zero cells. We used the benign prostrate cell line PNT1A and the prostate cancer cell line PC3. Generation of rho zero using another prostate cancer cell line, LNCaP, was not successful due to low adherence of these cells to the bottom of cell culture plates after treatment with mtDNA depletion agents. Lack of adherence has made separation of viable rho zero cells very difficult.

Repeated efforts using R6G treatment finally succeeded in the generation of PNT1A rho zero; we have not yet been successful with the PC3 cells, which did not survive the rho zero condition. Cybrid models were then generated in the PNT1A benign prostate nuclear background. Mitochondria from enucleated metastatic prostate cancer cell line PC3 or benign prostate cell line PNT1A were fused with the rho zero PNT1A nuclear donor to form the cybrids (PC3/PNT1A and PNT1A/PNT1A respectively). We did extensive analysis of these cybrids to confirm their mitochondrial and nuclear identity. We sequenced multiple regions of mtDNA to confirm the mitochondrial identity (Figure 1). Similarly, we sequenced several nuclear SNP markers (Figure 2) as well as short tandem repeat (STR) markers (Figure 3) to confirm the nuclear identity of the cybrids. These analyses confirmed that both PC3/PNT1A and PNT1A/PNT1A cybrids contain PNT1A benign nuclear background but mitochondria from PC3 and PNT1A, respectively.

In order to confirm the role of mitochondrial retrograde regulation of cancer mitochondria in a nuclear background we analyzed the cancer property of these cybrids. We used the soft agar colony formation assay with similar number of PC3 and PNT1A parental cell lines as well as PC3/PNT1A and PNT1A/PNT1A cybrids under uniform experimental conditions. We observed that the while PNT1A/PNT1A cybrid which contain both nucleus and mitochondria from the benign cell line rarely form colonies in the soft agar like PNT1A parental
cells, PC3/PNT1A cybrid with cancer mitochondria could grow anchored dependently in the soft agar (Figure 4).

**Figure-1: Confirmation of mitochondria by sequencing mtDNA of parental cells and cybrids**

- **MTRNR2**
  - **PNT1A-WT**
  - **PC3-WT**
  - **PNT1A/PNT1A**
  - **PC3/PNT1A**

- **COX-1**
  - **PNT1A-WT**
  - **PC3-WT**
  - **PNT1A/PNT1A**
  - **PC3/PNT1A**

- **D-LOOP**
  - **PNT1A-WT**
  - **PC3-WT**
  - **PNT1A/PNT1A**
  - **PC3/PNT1A**
Figure-2: Confirmation of nucleus by sequencing random SNPs in nuclear DNA

### PNT1A-WT

<table>
<thead>
<tr>
<th>No.</th>
<th>Sample File</th>
<th>Dir</th>
<th>Gene</th>
<th>Exon</th>
<th>Start</th>
<th>End</th>
<th>Size</th>
<th>Quality</th>
<th>Mut#</th>
<th>Mutation1</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>PNT1A-InGQ-1F_2013-10-22_B01.ab1</td>
<td>1-F</td>
<td>InternalGQ-01</td>
<td>-66</td>
<td>68</td>
<td>14</td>
<td>111</td>
<td>100.00%</td>
<td>1.02&gt;G&gt;A.p.5WXX</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>PNT1A-InGQ-3F_2013-10-22_B02.ab1</td>
<td>2-F</td>
<td>InternalGQ-03</td>
<td>-40</td>
<td>68</td>
<td>17</td>
<td>111</td>
<td>100.00%</td>
<td>1.02&gt;G&gt;C.p.RMTM</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>PNT1A-InGQ-2F_2013-10-22_B03.ab1</td>
<td>3-F</td>
<td>InternalGQ-06</td>
<td>-66</td>
<td>48</td>
<td>20</td>
<td>111</td>
<td>100.00%</td>
<td>1.02&gt;G&gt;C.p.RMTM</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>PNT1A-InGQ-7F_2013-10-22_B04.ab1</td>
<td>4-F</td>
<td>InternalGQ-07</td>
<td>-66</td>
<td>48</td>
<td>20</td>
<td>111</td>
<td>100.00%</td>
<td>1.02&gt;G&gt;C.p.RMTM</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>PNT1A-InGQ-8F_2013-10-22_B05.ab1</td>
<td>5-F</td>
<td>InternalGQ-08</td>
<td>-64</td>
<td>61</td>
<td>9</td>
<td>106</td>
<td>100.00%</td>
<td>1.02&gt;T&gt;G.p.HIN</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>PNT1A-InGQ-10F_2013-10-22_B06.ab1</td>
<td>6-F</td>
<td>InternalGQ-10</td>
<td>-10</td>
<td>68</td>
<td>10</td>
<td>106</td>
<td>100.00%</td>
<td>1.02&gt;G&gt;C.p.RMTM</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>PNT1A-InGQ-12F_2013-10-22_B07.ab1</td>
<td>7-F</td>
<td>InternalGQ-12</td>
<td>-61</td>
<td>66</td>
<td>36</td>
<td>106</td>
<td>100.00%</td>
<td>1.02&gt;G&gt;A.p.1RQ</td>
<td></td>
</tr>
</tbody>
</table>

### PC3-WT

<table>
<thead>
<tr>
<th>No.</th>
<th>Sample File</th>
<th>Dir</th>
<th>Gene</th>
<th>Exon</th>
<th>Start</th>
<th>End</th>
<th>Size</th>
<th>Quality</th>
<th>Mut#</th>
<th>Mutation1</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>PC3-InGQ-1F_2013-10-22_C01.ab1</td>
<td>1-F</td>
<td>InternalGQ-01</td>
<td>-66</td>
<td>68</td>
<td>20</td>
<td>111</td>
<td>100.00%</td>
<td>1.02&gt;G&gt;A.p.5WXX</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>PC3-InGQ-3F_2013-10-22_C02.ab1</td>
<td>2-F</td>
<td>InternalGQ-03</td>
<td>-40</td>
<td>68</td>
<td>22</td>
<td>111</td>
<td>100.00%</td>
<td>1.02&gt;G&gt;C.p.RMTM</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>PC3-InGQ-4F_2013-10-22_C03.ab1</td>
<td>3-F</td>
<td>InternalGQ-06</td>
<td>-66</td>
<td>48</td>
<td>18</td>
<td>111</td>
<td>100.00%</td>
<td>1.02&gt;G&gt;C.p.RMTM</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>PC3-InGQ-7F_2013-10-22_C04.ab1</td>
<td>4-F</td>
<td>InternalGQ-07</td>
<td>-66</td>
<td>48</td>
<td>18</td>
<td>111</td>
<td>100.00%</td>
<td>1.02&gt;G&gt;C.p.RMTM</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>PC3-InGQ-10F_2013-10-22_C05.ab1</td>
<td>5-F</td>
<td>InternalGQ-10</td>
<td>-10</td>
<td>67</td>
<td>18</td>
<td>111</td>
<td>100.00%</td>
<td>1.02&gt;T&gt;G.p.RMTM</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>PC3-InGQ-12F_2013-10-22_C06.ab1</td>
<td>6-F</td>
<td>InternalGQ-12</td>
<td>-62</td>
<td>66</td>
<td>31</td>
<td>107</td>
<td>100.00%</td>
<td>1.02&gt;G&gt;A.p.1RQ</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>PC3-InGQ-8F_2013-10-22_C07.ab1</td>
<td>7-F</td>
<td>InternalGQ-12</td>
<td>-61</td>
<td>66</td>
<td>31</td>
<td>107</td>
<td>100.00%</td>
<td>1.02&gt;G&gt;A.p.1RQ</td>
<td></td>
</tr>
</tbody>
</table>

### PNT1A/PNT1A

<table>
<thead>
<tr>
<th>No.</th>
<th>Sample File</th>
<th>Dir</th>
<th>Gene</th>
<th>Exon</th>
<th>Start</th>
<th>End</th>
<th>Size</th>
<th>Quality</th>
<th>Mut#</th>
<th>Mutation1</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>PNT1A-InGQ-1F_2013-10-22_B01.ab1</td>
<td>1-F</td>
<td>InternalGQ-01</td>
<td>-66</td>
<td>68</td>
<td>14</td>
<td>111</td>
<td>100.00%</td>
<td>1.02&gt;G&gt;A.p.5WXX</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>PNT1A-InGQ-3F_2013-10-22_B02.ab1</td>
<td>2-F</td>
<td>InternalGQ-03</td>
<td>-40</td>
<td>68</td>
<td>17</td>
<td>111</td>
<td>100.00%</td>
<td>1.02&gt;G&gt;C.p.RMTM</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>PNT1A-InGQ-2F_2013-10-22_B03.ab1</td>
<td>3-F</td>
<td>InternalGQ-06</td>
<td>-66</td>
<td>48</td>
<td>20</td>
<td>111</td>
<td>100.00%</td>
<td>1.02&gt;G&gt;C.p.RMTM</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>PNT1A-InGQ-7F_2013-10-22_B04.ab1</td>
<td>4-F</td>
<td>InternalGQ-07</td>
<td>-66</td>
<td>48</td>
<td>20</td>
<td>111</td>
<td>100.00%</td>
<td>1.02&gt;G&gt;C.p.RMTM</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>PNT1A-InGQ-8F_2013-10-22_B05.ab1</td>
<td>5-F</td>
<td>InternalGQ-08</td>
<td>-64</td>
<td>61</td>
<td>9</td>
<td>106</td>
<td>100.00%</td>
<td>1.02&gt;T&gt;G.p.HIN</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>PNT1A-InGQ-10F_2013-10-22_B06.ab1</td>
<td>6-F</td>
<td>InternalGQ-10</td>
<td>-10</td>
<td>68</td>
<td>10</td>
<td>106</td>
<td>100.00%</td>
<td>1.02&gt;G&gt;C.p.RMTM</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>PNT1A-InGQ-12F_2013-10-22_B07.ab1</td>
<td>7-F</td>
<td>InternalGQ-12</td>
<td>-61</td>
<td>66</td>
<td>36</td>
<td>106</td>
<td>100.00%</td>
<td>1.02&gt;G&gt;A.p.1RQ</td>
<td></td>
</tr>
</tbody>
</table>

### PC3/PNT1A

<table>
<thead>
<tr>
<th>No.</th>
<th>Sample File</th>
<th>Dir</th>
<th>Gene</th>
<th>Exon</th>
<th>Start</th>
<th>End</th>
<th>Size</th>
<th>Quality</th>
<th>Mut#</th>
<th>Mutation1</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>PC_PNT1A-PLK01-InGQ-1F_2013-10-22_E01.ab1</td>
<td>1-F</td>
<td>InternalGQ-01</td>
<td>-66</td>
<td>68</td>
<td>28</td>
<td>111</td>
<td>100.00%</td>
<td>1.02&gt;G&gt;A.p.5WXX</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>PC_PNT1A-PLK01-InGQ-3F_2013-10-22_E02.ab1</td>
<td>2-F</td>
<td>InternalGQ-03</td>
<td>-40</td>
<td>68</td>
<td>23</td>
<td>111</td>
<td>100.00%</td>
<td>1.02&gt;G&gt;C.p.RMTM</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>PC_PNT1A-PLK01-InGQ-4F_2013-10-22_E03.ab1</td>
<td>3-F</td>
<td>InternalGQ-06</td>
<td>-66</td>
<td>49</td>
<td>21</td>
<td>104</td>
<td>100.00%</td>
<td>1.02&gt;G&gt;C.p.RMTM</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>PC_PNT1A-PLK01-InGQ-7F_2013-10-22_E04.ab1</td>
<td>4-F</td>
<td>InternalGQ-07</td>
<td>-62</td>
<td>68</td>
<td>28</td>
<td>110</td>
<td>100.00%</td>
<td>1.02&gt;G&gt;C.p.RMTM</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>PC_PNT1A-PLK01-InGQ-8F_2013-10-22_E05.ab1</td>
<td>5-F</td>
<td>InternalGQ-08</td>
<td>-60</td>
<td>61</td>
<td>21</td>
<td>104</td>
<td>100.00%</td>
<td>1.02&gt;T&gt;G.p.HIN</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>PC_PNT1A-PLK01-InGQ-10F_2013-10-22_E06.ab1</td>
<td>6-F</td>
<td>InternalGQ-10</td>
<td>-13</td>
<td>67</td>
<td>16</td>
<td>70</td>
<td>100.00%</td>
<td>1.02&gt;G&gt;C.p.RMTM</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>PC_PNT1A-PLK01-InGQ-12F_2013-10-22_E07.ab1</td>
<td>7-F</td>
<td>InternalGQ-12</td>
<td>-61</td>
<td>66</td>
<td>38</td>
<td>106</td>
<td>100.00%</td>
<td>1.02&gt;G&gt;A.p.1RQ</td>
<td></td>
</tr>
</tbody>
</table>
Figure 3  STR profiles of cybrids and parental cells

<table>
<thead>
<tr>
<th>Sample_Name</th>
<th>AMEL</th>
<th>CSF1PO</th>
<th>D13S317</th>
<th>D16S539</th>
<th>D18S51</th>
<th>D21S11</th>
<th>D3S1358</th>
<th>D5S818</th>
<th>D7S820</th>
<th>D8S1179</th>
<th>FGA</th>
<th>THO1</th>
<th>TPOX</th>
<th>vWA</th>
</tr>
</thead>
<tbody>
<tr>
<td>PNT1A-WT</td>
<td>X,Y</td>
<td>12</td>
<td>11,12</td>
<td>10,11</td>
<td>16</td>
<td>27,28</td>
<td>16</td>
<td>12</td>
<td>10,11</td>
<td>10,12</td>
<td>23,246,7,9,3</td>
<td>8,11</td>
<td>17,19</td>
<td></td>
</tr>
<tr>
<td>PC3/PNT1A-PLKO1</td>
<td>X,Y</td>
<td>12</td>
<td>11</td>
<td>10,11</td>
<td>16</td>
<td>27,28</td>
<td>16</td>
<td>12</td>
<td>10,11</td>
<td>10,12</td>
<td>24</td>
<td>6,7,9,3</td>
<td>8</td>
<td>17,19</td>
</tr>
<tr>
<td>PNT1A/PNT1A-PLKO1</td>
<td>X,Y</td>
<td>12</td>
<td>11,12</td>
<td>10,11</td>
<td>16</td>
<td>27</td>
<td>16</td>
<td>12</td>
<td>10,11</td>
<td>10,12</td>
<td>24</td>
<td>6,7,9,3</td>
<td>8,11</td>
<td>17,19</td>
</tr>
<tr>
<td>PC3-ATCC</td>
<td>X</td>
<td>11</td>
<td>11</td>
<td>11</td>
<td></td>
<td>13</td>
<td>8,11</td>
<td></td>
<td>6,7</td>
<td>8,9</td>
<td></td>
<td></td>
<td></td>
<td>17</td>
</tr>
</tbody>
</table>

Figure-4: Analysis of tumor properties in parentals and cybrids by colony formation assay

4A: Phase contrast images of soft agar colonies after two weeks of culture

4B: Number of colonies counted using Gelcount colony counter
Overall, we successfully generated cybrid models under benign prostate nuclear background and analyzed its genetic and tumor properties using different mitochondrial donors. We are continuing our efforts to generate new cybrid models with cancer nuclear background and to further characterize the already generated PNT1A cybrids.

References:


Key Accomplishments Summarized:

1. Isolation of more than 1500 DNA samples required for the project.
2. Validation of a more robust sequencing technique using single amplicon long-range PCR that permits deep coverage (10,000-20,000X on average) of the mitochondrial genome, to date 652 samples of Frozen Tissue DNA have been sequenced.
3. Sequencing of FFPE DNA samples using Sequence Capture.
4. Mapping of the DNA variants to mitochondrial genes (data shown for ~first 200 samples nearly completed for remaining samples).
5. Cybrid models of benign (PNT1A) prostate cells have been successfully created and characterized. Despite the end of the project funding, we are continuing to work with the PC3 cells with the expectation that we will be able to create PC3 cybrids and characterize them.

Opportunities for professional development:

Though not funded by the DOD directly, several students and postdoctoral fellows have interacted with the project and in the summer of 2015, an undergraduate student from an underserved minority group. This last summer the undergraduate student (under the supervision of one of a postdoctoral fellow in Dr. Wheeler’s group) worked on coverage analysis of mitochondrial samples comparing the coverages obtained from FFPE samples to those of fresh frozen samples. He determined that the overall coverage patterns of fresh frozen were different from those of FFPE and that this would require separate analysis of the two approaches. He also identified a region of the FFPE capture that was poorly covered compared with fresh frozen and attempted to map the mitochondrial genes to that area. This area is largely “uncharted territory” and allowed both the postdoctoral fellow and the undergraduate to develop new skills in analyzing sequence data. Post doctoral fellows who were part of the Houston REACH IRACDA, or Houston Research Education and Career Horizon Institutional Research and Academic Career Development Award, K12 Postdoctoral Fellowship Program have also participated in the project. This IRACDA program is a model for postdoctoral training that combines frontier-level biomedical research and training to teach research-oriented courses to undergraduates at partner institutions. The project includes under-represented post-doctoral fellows when possible, who provide excellent role models for the undergraduates.
Dissemination of data to communities of interest:
Though data from this work has not yet been published in a peer reviewed journal, portions of the data have been presented in local venues including the Human Genetics Postdoctoral seminar series and the SMART (Summer Medical And Research Training) Program Seminar series aimed at undergraduate students interested in research and medical careers.

Next Reporting period plans:

We are excited to report that we have been able to recruit an outstanding postdoctoral fellow – Dr Jing (Cindy) Dong - to work on this project in the next year. Dr. Aaron Thrift, who had intended to work on this project, will serve as one of Dr. Dong’s mentors. Funds for Dr Dong’s salary support will come from the Cancer Prevention and Research Institute of Texas via a training grant held by Dr Margaret Spitz, a BCM Faculty member. Dr. Dong is presently a postdoctoral fellow in Aging and Neuroepidemiology at the Epidemiology branch of the National Institute of Environmental Health Sciences at Research Triangle Park in North Carolina. Dr. Dong has a PhD in Epidemiology and significant experience in genome-wide association studies and genetic and environmental risk factors. Dr. Dong came to Houston in June of 2016 to interview and give a seminar. BCM faculty members were very impressed with her and offered her a position on the spot which has required that she change her visa status (which is taking months unfortunately). During that trip we met at some length to discuss the prostate mitochondrial project – Dr. Dong was extremely enthusiastic to become involved and assist with leading the data analysis. Since that time we have had a series of phone calls to plan data analysis and the decision was to obtain an IRB exemption agreement so that Dr Dong could begin data analysis even prior to joining BCM. That agreement which has just been approved contained the following statements signed by both parties:

Sender of coded specimens/data:

I, Dr. Martha Mims at Baylor College of Medicine, holder of the code-key, cipher and/or identifiers for the shared [Prostate Mitochondrial DNA Data], promise not to release individually identifiable information or about the subjects from whom the [Prostate Mitochondrial DNA Data], derive to Jing Dong, PhD, of NIEHS per the provisions of U.S. Code 45 C.F.R, 46.

Recipient of coded specimens/data:

I, Jing Dong, PhD, of NIEHS, recipient of the [Prostate Mitochondrial DNA Data], promise not to request individually identifiable information about the subjects from whom the [Prostate Mitochondrial DNA Data], derive, from Dr. Martha Mims at Baylor College of Medicine per the provisions of U.S. Code 45 C.F.R. 46.

No data has yet been sent to Dr Dong, but our intention is to initiate data analysis intensively in the next year with the expectation that we will produce a high quality publication. Dr Wheeler is fully in support of this plan.

4. Impact:
Development of the principal discipline:
This project performed mitochondrial sequencing of more than 1000 subject samples from African American men. To our knowledge this is the largest such database in existence. In concert with Dr. David Wheeler in the Human Genome Sequencing Center and Dr. Lee-Jun Wong, we have implemented a new more rapid and much more powerful technique for sequencing mitochondrial DNA using single amplicon long-range PCR. This technology permitted us to rapidly sequence 652 samples using NGS technology at a depth of coverage of 10,000-20,000X. This technology allows much more accurate base calls and improves the limit of detection for heteroplasmy from about 15% with Sanger sequencing to ~ 1-2% with the long range primer technique. This technology has permitted much more rapid generation of mtDNA sequence with much less DNA and less “contamination” by nuclear DNA. Further, we have optimized sequencing strategies for the FFPE DNA and this portion of the project has just been completed. Two strategies were necessary for sequencing these two
kinds of DNA samples and both of these strategies were either developed or perfected during the course of this project. Publication of the methodology and the data obtained will represent a huge leap forward for the field of mitochondria and cancer. Further, we have created a large database of DNA sequence combined with clinical data and we are in the process of analyzing this data (See above) to produce the first manuscript relating to this work. We envision that this annotated data will reveal that mitochondrial metabolism directly impacts the development of prostate cancer.

**Impact on other disciplines:**
Though we had not foreseen this when the project was first envisioned, we are now working with investigators in metabolomics to examine whether the mitochondrial sequence changes we identify can be correlated with changes in metabolism in vivo (urine or plasma metabolome).

**Impact on Technology Transfer:**
Nothing to report

**Impact on Society:**
The overarching goal of this project is to understand why African American men are more prone to develop prostate cancer than any other racial/ethnic group. We envision that mitochondrial inheritance plays a significant role in aggressiveness of prostate cancer in African American men. Uncovering risk alleles may impact both screening and treatment for these men.

5. **Changes/Problems:**

**Changes in Approach:**
No significant changes in objectives or scope were encountered. As detailed in Aim 1, improvements in technology developed during the project allowed us to gain much deeper sequencing information than originally envisioned. As detailed in Aim 2, the first approach to developing cybrids was not feasible and we utilized an alternative technology which was successful in the PNT1A cells.

**Problems or Delays**
Creation of the cybrid cells proved to be more difficult than we had envisioned at the outset of the project. Ultimately we were able to develop and characterize the PNT1A cybrids, but we are still struggling the PC3 cybrids making slow progress. As noted below, we will be able to share our insights in a methods paper which we hope to publish in the spring of 2017.

**Changes that had a significant impact on expenditures**
None

**Significant changes in humans subjects, vertebrate animals or select agents:**
None

6. **Products:**

**Publications:**
No completed reportable outcomes have been accomplished as yet, however; the mitochondrial database with mtDNA sequence data, African American Admixture Data and Clinical Data is being assembled now. This is a massive amount of data – more than we had originally foreseen. In order to analyze the data, we have recruited Dr. Cindy Dong (See above) to work with Dr. Aaron Thrift, in the Cancer Epidemiology program at the Dan L Duncan Cancer Center. With the help of Drs. Thrift and Dong, we will be in a position to submit the first manuscript on this data – later than we had wanted to do, but we have overcome the technical challenges inherent in sequencing the FFPE DNA samples to create a very large database of information. The cybrid cell lines were unfortunately more difficult to create than we had anticipated, but we have succeeded in creating PNT1A cybrids and we have not given up on the PC3 cybrids.
Technology for sequencing mitochondrial DNA from FFPE samples is described above. In concert with Dr. David Wheeler in the Human Genome Sequencing Center and Dr. Lee-Jun Wong, we have implemented a new more rapid and much more powerful technique for sequencing mitochondrial DNA using single amplicon long-range PCR. This technology permitted us to rapidly sequence 652 samples using NGS technology at a depth of coverage of 10,000-20,000X. This technology allows much more accurate base calls and improves the limit of detection for heteroplasmy from about 15% with Sanger sequencing to ~ 1-2% with the long range primer technique. This technology has permitted much more rapid generation of mtDNA sequence with much less DNA and less “contamination” by nuclear DNA. We have optimized sequencing strategies for the FFPE DNA and this portion of the project has just been completed. We will publish this information in a peer reviewed journal along with the actual data gleaned from these samples. The pitfalls of using DNA derived from FFPE samples as compared to DNA derived from frozen tissue will be discussed in detail as will details of how to manage these kinds of data.

Other Products:
We are in the process of creating a very large database of DNA sequence information and clinical data on African American men with prostate cancer. This information will be invaluable in terms of evaluating the role mitochondrial inheritance on prostate cancer aggressiveness. Further, we have developed a library of germline DNA which can be queried for other projects.

7. Participants
   Listed below are the participants in the project though no funding for their roles was provided in the last year.

Martha Mims, MD, PhD, PI No Change
Lee-Jun Wong, PhD, Co-Investigator No Change
David Wheeler, PhD, Co-Investigator No Change
Suzanne Leal, PhD, Co-Investigator No Change
Myoung Kwon, Research Technician, No Change

8. Special Reporting Requirements: None

9. Appendices: CV of Dr. Jing Dong proposed postdoctoral fellow
CURRICULUM VITAE
Jing Dong Ph.D.

CONTACT INFORMATION

Epidemiology Branch
National Institute of Environmental Health Sciences
BG 101 RM A359 MSC A3-05
111 T.W. Alexander Drive
Research Triangle Park, NC 27709

EDUCATION

Postdoctoral Fellow in Aging & Neuroepidemiology Oct. 2012 - present
National Institutes of Health (NIH) / National Institute of Environmental Health Sciences (NIEHS)
- Supervisor: Dr. Honglei Chen
- Research focus: Genetic and environmental risk factors for neurodegenerative disease

Ph.D. (Five-year Program) in Epidemiology Sep.2007 – July 2012
Nanjing Medical University
- Advisor: Prof. Hongbing Shen
- Research focus: Genetic and environmental risk factors for cancers
- Thesis: “Genome-wide association analysis of lung cancer in Han Chinese”

Nanjing Medical University

FUNDING


Participation, The National High Technology Research and Development Program of China (863 project), Equivalent to UO1, “Genome-wide association study of lung cancer” 2010-2012

Participation, National Science Foundation of China, Equivalent to RO1, “MicroRNA polymorphisms and NSCLC prognosis” 2010-2014

AWARDS AND HONORS

The second-class scholarship from Nanjing Medical University 2003
The first-class scholarship from Nanjing Medical University 2004
Jing Dong

The second-class scholarship from Nanjing Medical University 2005
The second-class scholarship from Nanjing Medical University 2006
The second-class scholarship from Nanjing Medical University 2008
The first-class scholarship from Nanjing Medical University 2009
The first-class scholarship from Nanjing Medical University 2010
The first-class scholarship from Nanjing Medical University 2011
Excellent student in Jiangsu Province 2012
Excellent Graduate of Nanjing Medical University 2012
Excellent Doctoral Dissertation of Jiangsu Province 2013

TEACHING AND PROFESSIONAL EXPERIENCE

**Graduate Teaching Assistant** in Epidemiology 2009-2012
Nanjing Medical University

**Graduate Research Assistant** 2010-2012
Department of epidemiology and biostatistics, Nanjing Medical University

INVITED PRESENTATIONS

“Susceptibility Loci for Pigmentation and Melanoma in Relation to Parkinson’s Disease”, Boston, ASHG 2013 annual meeting 2013


MEMBERSHIP IN PROFESSIONAL SOCIETIES

The American Society of Human Genetics 2013-

PEER REVIEWER FOR JOURNALS

Parkinsonism & Related Disorders
Oncotarget
Carcinogenesis
The Pharmacogenomics Journal
Annals of Surgical Oncology
The International Journal of Neuroscience
Lung cancer
PLOS One
Tumor Biology
Mutation Research
Journal of Bioanalysis & Biomedicine
Gene
Frontiers of Medicine


**PUBLICATIONS (Co-Author)**


**RESEARCH EXPERIENCE**

1. **Epidemiological investigations:** questionnaire design, face-to-face interviews, bio-samples collection and conservation and follow-up.

2. **Laboratory training:**
   (1) Genomic experiment, such as extraction and purification of DNA in tissue, blood and serum/plasma, and SNPs screening and genotyping techniques (including PCR-RFLP, TaqMan and Openarray);
   (2) Expression experiment, including extraction and purification of RNA (in tissue and plasma/serum) and protein, Expression analysis of mRNA/microRNA/protein by RT-PCR, Real-Time PCR and TaqMan Low Density Array or EMSA;
   (3) Molecular biological techniques, including cell culture, reporter gene assay and MTT and
trans-well assay;
(4) Tissue/serum/plasma DNA methylation analysis, such as MSP.

3. **SNP data mining:** primer design and sequencing analysis, tagging SNPs selection using haploview and tagger, dissection of LD pattern and haplotype construction using LDA, PHASE, SAS, R and haploview, and SNP functional prediction using TRANSFAC, SIFT, UCSC, SNPinfo, PolyPen and PupaSuite.

4. **MicroRNA data mining:** primer design and sequencing analysis, and target gene searching using PicTar, miTarget, Targetscan and miRTarBase.

5. **Statistical analysis:** general statistical analysis for SNP and expression data, genome-wide association analysis, gene-gene/gene-environmental interaction analysis, meta-analysis, cis-eQTL analysis, etc.

6. **Software:** statistical software (including EPIDATA, SPSS, SAS, Stata, R and Plink, etc.) and Microsoft Offices (including Word/Excel/PowerPoint and Adobe’s Reader, etc.).

7. **Professional Training:** 2010 IARC summer school in Cancer Epidemiology

---

**PATENTS IN CHINA**

1. Non-small cell lung cancer prognosis related serum/plasma miRNA markers and their applications (2009.08.28-2029.08.28  ZL200910184518.4)

2. Breast Cancer related serum/plasma miRNA markers and their applications (2010.09.08-2030.09.08  ZL 201010275557.8)

3. Cervical Cancer and CIN related serum/plasma miRNA markers and their applications (2010.09.08-2030.09.08  ZL 201010275552.5)