Sprouse, Marc L., Mitochondrial Genetics and Function among Men Screened for Prostate Cancer. Doctor of Philosophy (Biomedical Sciences, Molecular and Medical Genetics), November 2015, 178 pp., 15 tables, 33 illustrations, bibliography, 303 titles.

Genetic alterations are associated with sporadic cancer development, progression and metastasis. Until now, more was known about nuclear DNA (nDNA) mutations and their role in cancer than the types of genetic changes that occur in the mitochondrial genome (mtGenome). Changes to mitochondrial DNA (mtDNA) are frequent in prostate cancer (PCa). Further, these changes have been associated with enhanced tumorigenesis and progression to an aggressive phenotype. However, it is unclear whether changes to the mtGenome can delineate PCa progression from indolent cancer to an advanced metastatic disease. An exhaustive characterization of the mtGenome of men screened for PCa was performed to determine if genetic differences between varying PCa disease states can be measured.

Two genetic techniques (copy number and genome sequencing) were used to perform a comprehensive characterization of the changes to mtDNA in whole blood extracts from three groups of men screened for prostate cancer: normal control (NC), no evidence of disease (NED), and, biochemical recurrence/metastasis (BCR/MET). Mitochondrial DNA copies per cell and mtGenome deletion ratio (whole mtGenomes:truncated mtGenomes) were measured using a multiplex real-time quantitative PCR (qPCR) assay. Whole mtGenome sequence data were generated using a massively parallel sequencing platform, the Ion Torrent Personal Genome Machine (PGM).

Real-time qPCR revealed a higher dispersion of mtDNA copies per cell and mtGenomes harboring a large scale deletion in samples from men with advanced stages of PCa when compared to normal controls and indolent PCa. Whole mtGenome sequencing showed a higher number of genetic variants in men with PCa, some of which are predicted to be pathological. A significant positive correlation was observed between mutational load and PCa disease status. Further, three-dimensional comparative modeling evidenced the negative effect of a single mtDNA missense mutation on a protein's structural integrity. Overall, the presented data suggest there are differences in the mtGenome between men with and without PCa that are measurable in peripheral blood and may be used as a potential risk assessment tool. Future analyses with a larger sample size may lead to more compelling evidence that supports the role of changes to the mtGenome with PCa progression.

MITOCHONDRIAL GENETICS AND FUNCTION AMONG

MEN SCREENED FOR PROSTATE CANCER

Marc L. Sprouse, BS

APPROVED:

_____

Major Professor: Arthur Eisenberg, PhD

_____

Committee Member: Rhonda Roby, PhD, MPH

_____

Committee Member: Deanna Cross, PhD

_____

Committee Member: Robert Barber, PhD

_____

University Member: Rong Ma, MD, PhD

_____

Chair, Department of Molecular and Medical Genetics: Robert Barber, PhD

_____

Dean, Graduate School of Biomedical Sciences: Meharvan Singh, PhD

MITOCHONDRIAL GENETICS AND FUNCTION AMONG MEN SCREENED FOR PROSTATE CANCER

DISSERTATION

Presented to the Graduate Council of the

Graduate School of Biomedical Sciences

University of North Texas

Health Science Center at Fort Worth

in Partial Fulfillment of the Requirements

For the Degree of

DOCTOR OF PHILOSOPY

By

Marc L. Sprouse, B.S.

Fort Worth, Texas

July 2015

ACKNOWLEDGMENTS

I would like to thank and acknowledge all those who have supported me throughout this project. To my graduate committee, Drs. Rhonda Roby, Arthur Eisenberg, Bob Barber, Deanna Cross and Rong Ma, thank you for your continued support, patience and sharing of your knowledge and experience with me. Dr. Roby has been my strongest advocate. I appreciate all her contributions of time, ideas, and funding to make my Ph.D. experience productive and stimulating. Dr. Eisenberg, as my department chair and committee member, has always made my needs a priority. I thank Dr. Barber for the many engaging and informative discussions we've had. He is an excellent teacher and mentor who has taught me the value of perspective. As my graduate advisor Dr. Cross has been a tremendous help. Her contributions in the past few months have shown me the true value in mentorship and its power to elevate a student. And finally Dr. Ma who was my first mentor at UNTHSC. The amount of positive influence his guidance has had on me is insurmountable.

I also acknowledge the entire faculty of the Department of Molecular and Medical Genetics and the Center for Human Identity who took their time to assist me and played a crucial role in my continuing education. Specifically, thank you to Drs. Michael Allen, John Planz and Bruce Budowle for providing me access to their laboratories, equipment and supplies which were critical for the successful completion of the experiments performed in this project. I also

thank and acknowledge Dr. Laszlo Prokai for providing his laboratory, time and expertise which made my protein modeling study possible.

The success of this project would not be possible without the support and contributions of my collaborators at the University of Texas Southwestern Medical Center and the University of Texas Health Science Center, San Antonio. Specifically, I owe a great debt of gratitude to Dr. Jer-Tsong Hsieh and his laboratory, including Dr. Loredana Moro, at UTSW for providing samples that were used in the preliminary studies. I am also deeply grateful to Drs. Ian Thompson and Robin Leach, as well as Brandi Weaver and Desiree Wilson, at UTHSCSA for their efforts and for providing the SABOR samples which were used in the larger studies of this project.

Primarily I would like to thank my family for their unconditional support they have provided me throughout my graduate school experience. My parents, James and Renee Sprouse and my brother Randy Sprouse, who have been my strongest supporters since the beginning. Also, my sisters, Kim Higgins and Stacy Looney, who have been gracious and equally supportive of my work. To my lab-mates, colleagues and friends, thank you for all the help and support you have provided inside and outside the laboratory. Specifically, I would like to thank Dr. Nicole Phillips who has been an amazing friend and collegue during my time at UNTHSC.

And to Alessandra Alicea-Centeno, whose incredible friendship and support these past few months have been crucial for my wellbeing and the completion of my project.

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

Cancer is a group of diseases characterized by abnormal cells with a seemingly limitless proliferative potential. Collectively, cancer accounts for nearly 25% of all deaths in the United States (U.S.) making it the second-leading cause of death among Americans.[1] For men in the US, prostate cancer (PCa) is the most commonly diagnosed non-dermatologic cancer and is second only to lung cancer as the most common cause of cancer death. Its incidence and mortality rates in the U.S. are approximately fourteen percent and five percent, respectively. Consequently, a total of 220,800 new PCa cases and 27,540 deaths from the disease are estimated this year. Overall, the lifetime risk for men living in the U.S. of developing PCa is 16% whereas the risk of dying from the disease is only 2.9%.[1] Economically developed countries across the world are also burdened with a high prevalence of PCa. Globally, PCa is the second most commonly diagnosed cancer in men and the sixth leading cause of male cancer death.[2]

The greatest risk factor for developing PCa is age. A 70-year-old man has a chance of one in nine of developing PCa whereas a man younger than 50 has a chance of only one in 298. Indeed, 97% of all PCa diagnoses occur in men 50 years of age and older.[1] The second greatest risk factor for developing PCa is race and ethnicity. Within the U.S. and globally, men of African descent are affected disproportionately by the disease. The incidence rate of African Americans is 60% higher than non-Hispanic white men living in the U.S. and, globally, men of African

descent have the highest PCa incidence rates in the world.[1] Family history, which includes both environmental and genetic factors, contribute to a definitive risk for developing PCa. Men who have a first-degree relative with PCa have an increased relative risk (2.48, 95% CI 2.25–2.74) of developing the disease as compared to men who do not.[3] Additionally, risk increases if the relative has early-onset PCa or if multiple first-degree relatives are diagnosed.[4] However, an increased familial risk for PCa does not necessarily confer hereditary disease. The majority of PCa cases are sporadic adenocarcinoma whereas only 5 to 10% are characterized by Mendelian inheritance of a highly penetrant, autosomal dominant gene.[5]

One of the foremost problems in PCa management is the inability to distinguish indolent from aggressive tumors which has resulted in over-diagnosis and overtreatment of the disease. This stems from a poor understanding of the mechanisms of tumorigenesis in the prostate. Like other cancers, PCa is a complex neoplastic disease that requires genetic and nongenetic factors for its initiation and progression. However, the molecular basis of the disease includes a high degree of genetic heterogeneity which is thought to underlie its varying clinical behavior. A consequence of PCa's molecular heterogeneity has been the challenge of identifying an adequate biomarker for clinical disease. It is unquestionable that the use of biomarkers, *e.g.*, prostate specific antigen (PSA) and digital rectal exam (DRE), for PCa screening has transformed the landscape of disease detection and management but there still remains no screening test that is specific for the disease. Prostate cancer survival is contingent upon different disease elements such as the extent of the tumor(s) at the time of diagnosis. For example, the five-year

relative survival rate of men with cancer confined to the prostate (localized) or with minimal (regional) spread is 100%. But men with distant metastasis have a five-year survival rate of only 31.9%.[1] A screening program that accurately discerns indolent from aggressive PCa would reduce the number of patients unnecessarily subjected to radical treatments and the associated undesirable side effects and healthcare costs. Furthermore, identifying PCa cases likely to progress to an advanced disease at an early stage would help improve the patient's treatment strategy to one that is more efficacious and improve survival outcomes.

Structure, Function and Pathological Considerations of the Prostate

The prostate is an exocrine gland and the largest accessory organ of the male reproductive system. It is about the size of a walnut, somewhat round and conical in shape and consists of glandular (70%) and fibromuscular (30%) tissue giving it a firm but doughy consistency. The prostate is located just below the urinary bladder, behind the pubic bone and in front of the rectum. The urethra passes through the entire length of the prostate beginning at its base, just below the bladder, to its apex (Figure 1). The glandular tissue contains 30 to 50 prostatic tubuloalveolar glands which empty directly into the urethra or through independent secretory ducts. Additionally, two ejaculatory ducts, which are formed by the joining of a *vas deferens* with its corresponding seminal vesicle, pass through the prostate and open into the prostatic urethra at the *colliculus seminalis*.

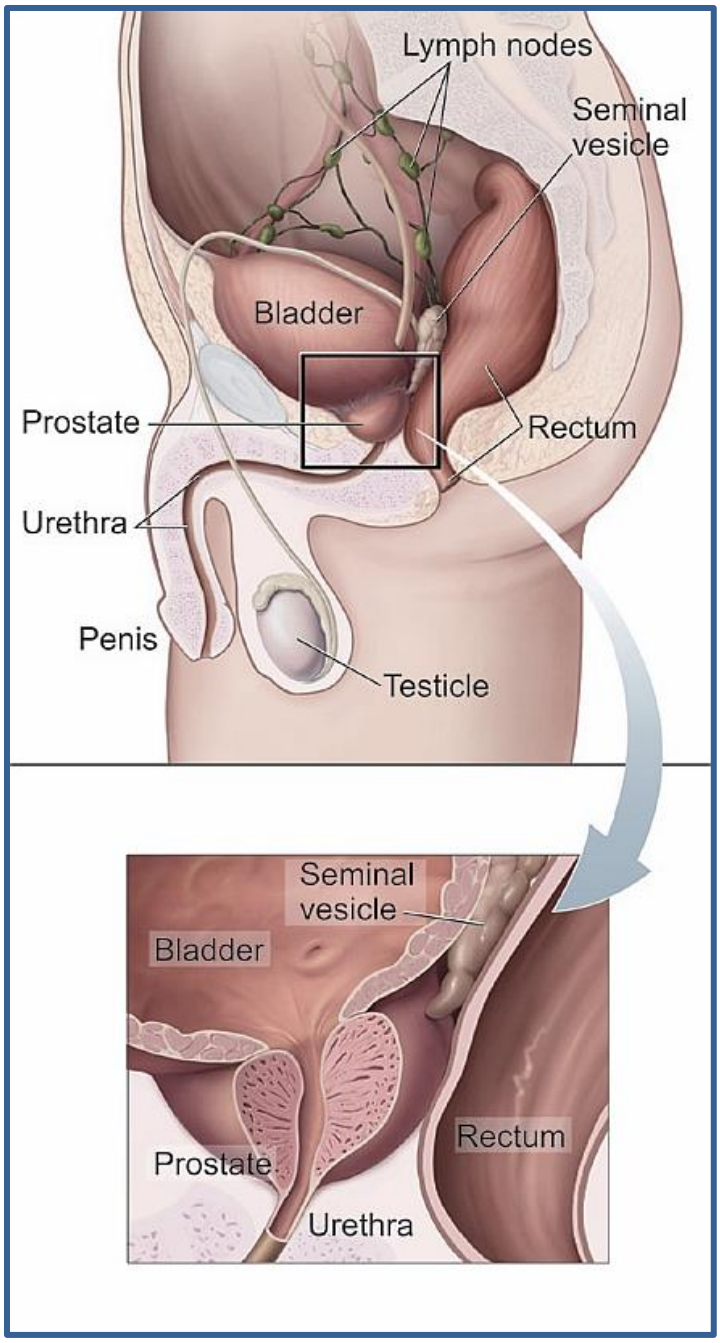


Figure 1: The prostate and nearby organs. The prostate is about the size of a walnut and somewhat round and conical in shape. It is located just below the urinary bladder, behind the pubic bone and in front of the rectum. The urethra passes through the entire length of the prostate beginning at its base, just below the bladder, to its apex.[6]

The anatomy of the human prostate is unlike other mammals because it is not distinctly lobular. Rather, it has a homogenous structure that is better defined by three glandular zones: the peripheral zone (PZ), the central zone (CZ) and the transition zone (TZ), as described by McNeal (Figure 2).[7,8] The exact origin of a prostatic tumor is often hard to identify because of hyper-proliferation and cell invasion into neighboring zones. However, approximately 70% of PCa originates in the PZ which is the largest glandular region. It constitutes approximately 70% of the prostate volume and is the area that is palpitated during a DRE.[9] The second largest glandular region of the prostate is the CZ. It accounts for 25% of the prostate volume and is the area surrounding the ejaculatory ducts. Only 2.5% to 8.0% of PCa cases originate in the CZ but those that do are comparatively much more aggressive and have a distinct route of spread through ejaculatory ducts and seminal vesicles.[9,10] The third and smallest glandular region of the prostate is the TZ which constitutes only five percent of the prostate volume. The TZ is the innermost section of the prostate and includes the area between the PZ and CZ. It constitutes the middle area of the prostate and surrounds the urethra as it passes through the gland. It is also the area of the prostate where benign prostatic hyperplasia (BPH) predominantly originates.[11,12] Despite its small volume size, a large proportion (approximately 20%-25%) of PCa cases originate in the TZ.[9,13]

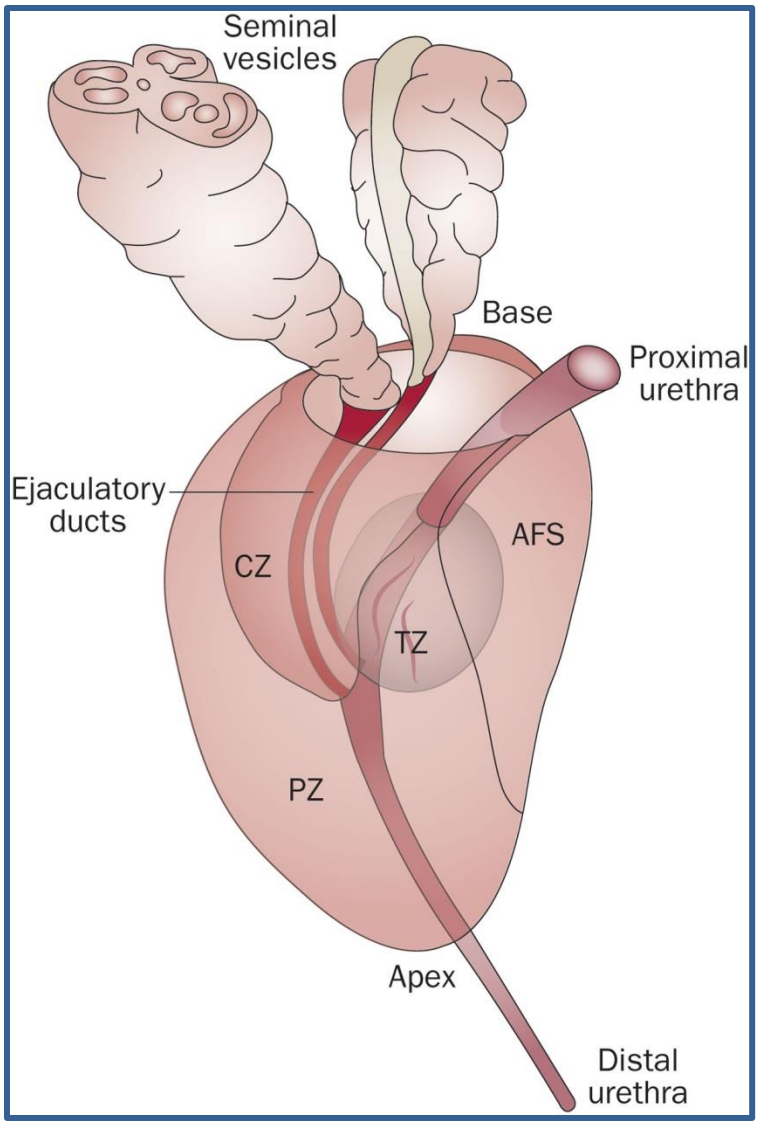


Figure 2: Zonal anatomy of the prostate. The anatomy of the human prostate is unlike other mammals because it is not distinctly lobular. Rather, it has a homogenous structure that is better defined by three glandular zones: the peripheral zone (PZ), the central zone (CZ) and the transition zone (TZ).[7]

The glandular structure of a normal, healthy prostate consists of luminal spaces embedded in fibro-muscular stroma (smooth muscle cells and fibroblasts) (Figure 3). Lining the luminal spaces are glandular epithelium which is composed of two histologically distinct cell layers, an inner layer and an outer layer. These two layers are formed by three different cell types distinguishable by their respective morphology and cell surface antigens. The three cell types that form prostate glands are luminal secretory, neuroendocrine and basal cells.

Luminal secretory cells are the major component of normal and malignant prostate gland tissue. Ninety percent of the inner, luminal cell layer is composed of tall columnar secretory cells which generate secreted proteins such as PSA and prostatic acid phosphatase (PAP).[14] These cells also express high levels of androgen receptor (AR) which mediates the production and secretion of PSA and PAP. A distinguishing characteristic of these cells is androgen-dependence; without androgen (testosterone and dihydrotestosterone), luminal secretory cells undergo apoptosis and prostate tissue regression.[15] Luminal secretory cells also express the low molecular weight cytokeratins CK8 and CK18, and the cell-surface marker CD57.[16–19]

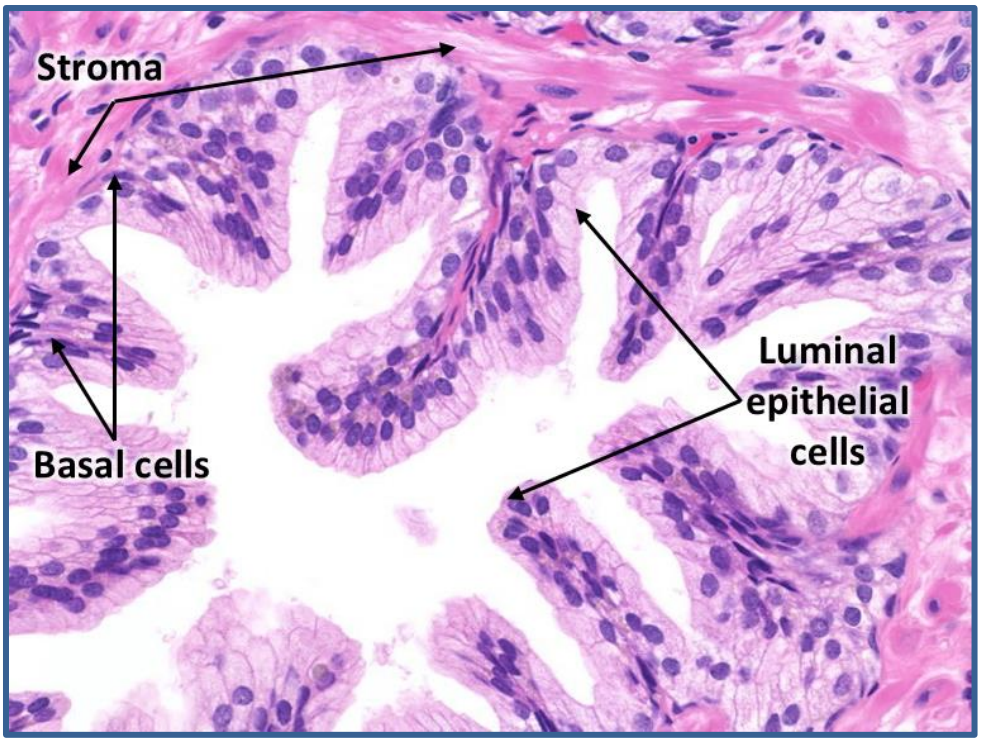


Figure 3: Glandular structure of the prostate. This figure shows an hematoxylin and eosin stained section of human prostate gland tissue. A normal, healthy prostate has an undulating luminal configuration with clusters and papillary infoldings which form an individual gland. The tall secretory epithelial cells have pale clear cytoplasm and uniform round or oval nuclei. Basal cells are cuboidal non-secretory cells which adhere strongly to, and form, a continuous layer along the epithelial basement membrane.[20]

The inner, luminal cell layer is also comprised of a small population of neuroendocrine cells which produce and secrete serotonin and other peptide hormones. Neuroendocrine cells are androgen-independent , non-proliferating, terminally differentiated cells that do not express AR or the anti-apoptotic oncogene B-cell lymphoma 2 (Bcl-2).[21–23] Although the exact role of these cells are not known, their secretory products have growth-promoting properties

that may play a role in the growth and differentiation of the developing prostate and have been implicated in carcinogenesis.[24,25]

The outer, basal cell layer is formed by cuboidal non-secretory cells which adhere strongly to, and form, a continuous layer along the epithelial basement membrane. Basal cells are easily identified by staining for high molecular weight cytokeratins, *e.g.*, CK5 and CK14, and unlike the secretory cells of the inner, luminal layer, basal cells are androgen independent and do not express AR, PSA or PAP.[19] Beyond their structural function, which is the formation of the outer-layer/base of prostate glands, very little is known about basal cell function.

Over 90% of prostatic tumors are adenocarcinomas, *i.e.*, malignant neoplasms that arise from the glandular epithelium of the prostate, but there is contention as to the origin of cancer initiating cells.[26] Conventionally, luminal-specific cells have been thought to be the source of PCa initiation given the highly characterized loss of basal cells and corresponding destruction of the basement membrane during PCa development and progression.[27–29] Some estimates report the basal cell population to be less than one percent whereas androgen dependent luminal secretory cells constitute the bulk of prostate tumor mass at greater than 99%.[30] Also, an overwhelming proportion of PCa cases exhibit a secretory phenotype.[31] However, androgen deprivation results in the majority of luminal secretory cells (90%) to undergo apoptosis. The ensuing development of castration resistant/androgen independent PCa suggests basal cells as likely progenitors. Indeed, basal cells have a higher proliferative index than luminal cells and

therefore are a stronger oncogenic target for PCa initiation.[32,33] However, there is expanding evidence that indicates a whole separate subpopulation of prostate cells that are responsible for PCa initiation. Specifically, mounting evidence has identified a rare population of prostate cancer stem cells (CSCs) that reside within luminal and/or basal compartments that have a high oncogenic potential. Under normal conditions, prostate stem cells utilize their self-renewal capacity for prostate homeostasis and regeneration. However, under the CSC hypothesis, prostate stem cells undergo cancerous transformation and serve as cells of origin for PCa following an oncogenic event. Although the PCa cell origin debate is still ongoing, general agreement exists that malignant transformation originates in a subset of primitive cells and that most cells in the prostate do not generate tumors.[34,35] Nevertheless, determining the cell type(s) and origin of PCa remains a salient topic of investigation. Identifying the cell origin of PCa would provide a better understanding of the genetic and molecular mechanisms that govern initiation and progression of the disease. The translational benefit would include development of new diagnostic tools and therapeutic strategies for more effective treatment of PCa patients.

In principle, the prostate is an accessory organ of the male reproductive tract which confers its primary functions. As an exocrine gland, the prostate secretes prostatic fluid which accounts for approximately 30% of semen volume.[36] The enzymes and electrolytes present in prostatic fluid promote fertility in a number of ways. First, prostatic fluid contains PSA, a serine protease with chymotrypsin-like activity, which is thought to have the physiological role of

lysing clotting proteins, *e.g.*, semenogelins and fibronectin. The enzymatic action of PSA liquefies seminal coagulum which is believed to increase the motility potential of spermatozoa during insemination.[37,38] Its liquefying action is also thought to facilitate entry of spermatozoa into the uterus by dissolution of cervical mucus. Prostatic acid phosphatase is another protease secreted in prostatic fluid.[39] Although its physiological function is unclear, secreted PAP is thought to provide energy substrates to spermatozoa by hydrolyzing phosphorylcholine and phosphocreatine.[40] In addition to aiding fertility, prostatic secretions are rich in ionic compounds, such as zinc, which is thought to protect the reproductive tract from infections that originate in the urinary tract.[41–43]

Physically, the muscular action of the prostate is instrumental in the formation and expulsion of semen during sexual climax.[44] The prostate serves as an axis point where the different components of semen converge and are mixed. Muscular contraction of prostatic smooth muscle forces fluid from the secretory ducts into the urethra where it combines with spermatozoa from the epididymis (by means of the *vasa deferentia*) and secretions from the seminal vesicles and bulbourethral (Cowper's) gland to form semen. Coordinated contraction and relaxation of the urethral sphincters, located at the base and apex of the prostate, prevent semen from entering the bladder while prostatic smooth muscular contraction ensures its proper transmission through the urethra during ejaculation.[44] The prostate is also involved with managing fluid output from the bladder. Urination involves relaxation of the urethral sphincters which permit urine flow from the bladder through the prostatic urethra. Simultaneously,

prostatic smooth muscle contraction seals the secretory ducts thereby confining urine to the urethra and preventing it from entering the prostate glands.

The location and anatomy of the prostate make prostatic diseases inextricably complex. The most common medical condition that affects the prostate gland is benign prostatic hyperplasia (BPH) (Figure 4).[45] This disease is characterized as a nonmalignant enlargement of the prostate gland most commonly observed in aging men. Enlargement of the prostate caused by BPH is attributed to the overgrowth of smooth muscle and glandular epithelial cells predominantly in the periurethral transition zone. Clinically, BPH manifests as a lower urinary tract condition with symptoms that affect the normal pattern of urination, *e.g.*, urgency, increased frequency, nocturia, hesitancy, weakened and/or interrupted urinary stream, straining to urinate and the sensation of incomplete bladder emptying. These symptoms are the direct result of prostate enlargement which creates constricting pressure on the urethra. In severe BPH cases, prolonged obstruction may lead to reduced compliance of the prostatic urethra, acute urinary retention, recurrent urinary tract infections, hematuria, formation of bladder trabeculae, diverticula and calculi and renal insufficiency.[45]

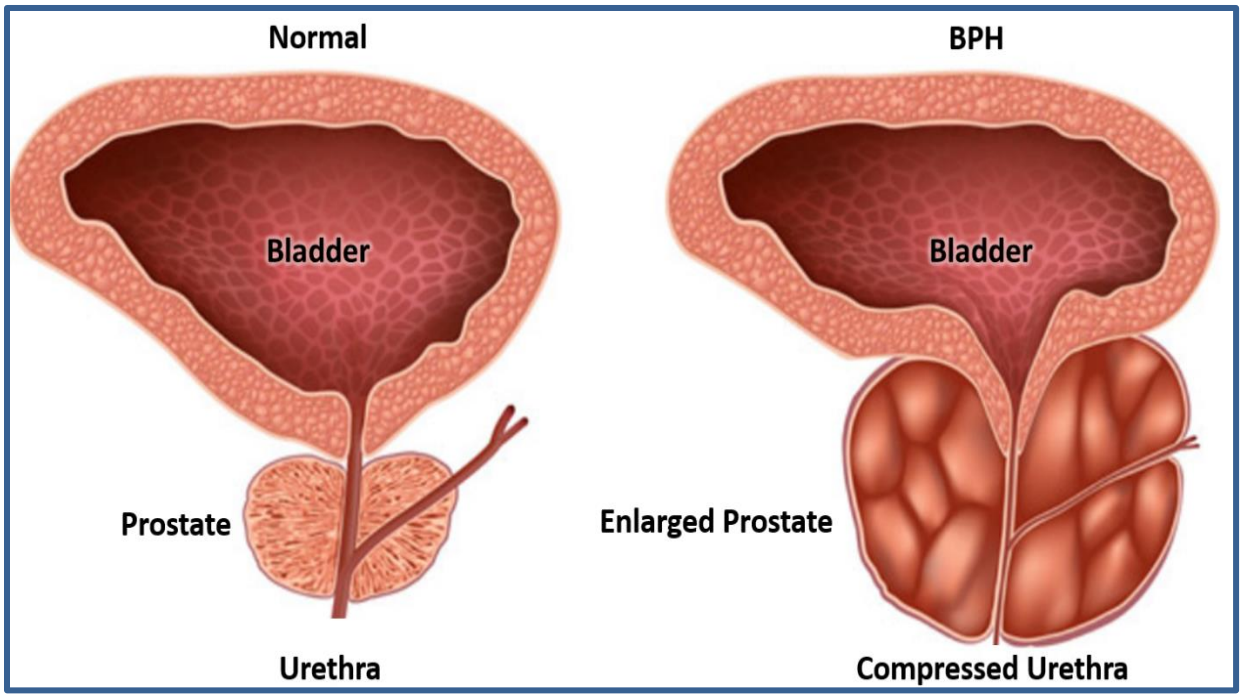


Figure 4: Normal and diseased prostate. The prostate is located just below the urinary bladder and encloses the urethra which passes through its entire length. A normal, healthy prostate (left image) is about the size of a walnut and somewhat round and conical in shape. BPH (right image) involves overgrowth of smooth muscle and glandular epithelial cells predominantly in the periurethral TZ. This results in an enlarged prostate which causes constricting pressure on the urethra.[46]

The cause of BPH is not fully understood but changes in hormone balance associated with normal aging processes appear to play a critical role in prostate enlargement. As men age, production of androgens (testosterone and dihydrotestosterone (DHT)) by the gonads and adrenals decrease at a rate of one percent per year. Further, the rate at which free testosterone decreases is more pronounced due to alterations in sex hormone binding globulin.[47] In males, estradiol-17β (E_2) is the main biologically active estrogen which is

predominantly generated through the aromatization of testosterone. Aromatase activity occurs in various regions of the male body, including liver, gonadal, central nervous system, skeletal muscle and adipose tissues. Thus, aromatization of testosterone to E_2 is often maintained in males with increasing age due to increased aromatase activity and age-associated fat mass.[48] Consequently, the age associated decrease in testosterone results in a higher proportion of E_2. The resultant imbalanced ratio of testosterone to E_2 has been shown to promote prostatic cell proliferation and may also play a role in the development of PCa. Although commonalities exist between BPH and PCa, there is no evidence that suggests men with BPH have an increased risk for PCa or that BPH is a precursor to PCa.

PCa Development and Progression

The full, natural history of PCa, beginning from its early, low grade status to its progression to an advanced, metastasized disease, has been difficult to characterize. Tissue culture models have been used successfully to glean a better understanding about the biology of PCa but they cannot accurately recapitulate the complex cellular interactions within the tumor microenvironment that play a key role in cancer initiation and progression.[49] Animal models have also been used extensively to study the etiology, prevention and treatment of human PCa. These models include genetically engineered mouse, xenograft, rat, and canine which each have their own strength and weaknesses. Undoubtedly, the use of tissue culture and animal models have greatly enhanced the understanding of PCa pathobiology.

Nevertheless, there still remains no optimal model that fully mirrors the initiation, development and progression of human PCa.[50]

Prostate cancer results from continued, unregulated proliferation of prostate cells stemming from an accumulation of abnormalities in multiple cell regulatory systems. Although the exact cause and mechanisms governing cellular dysfunction related to PCa are not fully known, key elements in the initiation and development of the disease are thought to follow the classical clonal expansion model common in most cancers (Figure 5).[51] In this model, development of PCa is a multistep process of progressive molecular alterations that transform a normal prostate epithelial cell into a tumor progenitor cell. An oncogenic event initiates this process and results in abnormal proliferation and eventual clonal expansion of tumor cells. Rapidly proliferating tumor cells incur continuous genetic events which can provide a selective advantage. Consequently, the process of clonal selection for those cells that exhibit pro-oncogenic characteristics, *e.g.*, increased growth rate, evasion of apoptosis, insensitivity to antigrowth signals, and tissue invasion and metastasis, result in further tumor growth and progression.[52]

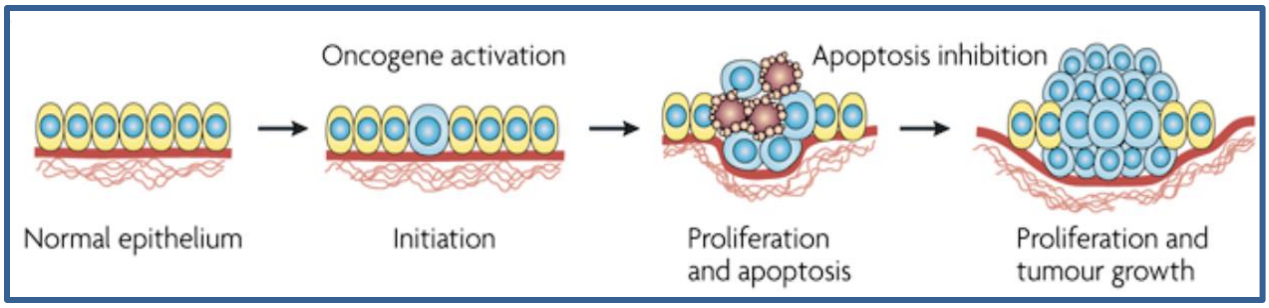


Figure 5: Stages of tumor formation. Prostate cancer initiation and development is a multistep process of progressive molecular alterations that transform a normal prostate epithelial cell into a tumor progenitor cell. An oncogenic event initiates this process and results in abnormal proliferation and eventual clonal expansion of tumor cells. In order for sustained tumor cell proliferation, defects in apoptosis are required.[53]

Most DNA mutations related to prostate cancer are acquired during a man's lifetime rather than having been inherited.[54] Further, the spectrum of genetic abnormalities in PCa is profoundly diverse evidenced by its clinical and molecular heterogeneity observed between separate PCa cases and even among different foci within the same individual. Structural mutations, or genomic rearrangements, are common in PCa and include amplification, deletion, and/or translocation of whole chromosomal segments. In contrast, point mutations which can result in missense and nonsense mutations occur less commonly. Nonetheless, driver mutations that contribute to the oncogenesis of prostate epithelial cells are generally classified as those that activate oncogenes and those that result in loss of function of tumor suppressor genes. With the exception of hereditary PCa, where several susceptibility loci have been identified, the number of recurrent mutations in specific genes are low for sporadic PCa. However, there are specific molecular pathways where genetic alterations are commonly observed. The

phosphoinositide 3-Kinase (*PI3K*) pathway is one of the most commonly altered molecular pathways in human cancer. In PCa, mutations in the PI3K pathway are observed in 25% to 70% of cases where metastatic tumors exhibit a significantly higher occurrence. Some of the most frequently mutated genes in the PI3K pathway include deletions of Phosphatase and Tensin homologue (*PTEN*) and PH domain and Leucine rich repeat Protein Phosphatase 1 (*PHLPP1*).[55–57] Ultimately, the consequence of these deletions result in the activation of the PI3K pathway which affects cell proliferation, survival and invasiveness. Other genes and pathways commonly affected by genomic alterations in PCa include mitogen-activated protein kinase (*MAPK*), the tumor suppressor p53 (*TP53*), retinoblastoma protein (*RB1*) and the transcription factor *MYC* (c-Myc). However, mutations in these genes and pathways are not unique to PCa, rather, they have been well characterized and are common in many other cancers. The question is what, if any, genetic alterations are exclusive to PCa and what unique properties of the prostate lead to its tissue specificity and the eventual manifestation of adenocarcinoma. [54]

The male sex hormones, *i.e.*, androgens, are necessary for the differentiation and maturation of male reproductive organs including the prostate.[58] Androgen dependence is also at the central axis of PCa pathogenesis.[59] Recurrent genetic alterations in the AR gene and signaling pathway further strengthen the role of androgens in the development and progression of PCa.[60] Although normal physiological levels of androgen alone do not cause PCa, it is necessary for malignant prostate cell development and proliferation.[61,62] Indeed, the androgen signaling pathway has been demonstrated as a critical factor for PCa initiation and

progression.[63,64] Prostate cancer begins as androgen-dependent; without hormones it does not grow and undergoes apoptosis which can result in clinical regression with tumor volume reduction in 60% to 80% of men.[26] As such, androgen ablation therapy is the mainstay of PCa treatment as many cases respond favorably to this modality. For some patients, approximately 30%, the cancer inevitably recurs as castration-resistant prostate cancer (CRPC) within a period of months to years indicated by an increase in serum PSA levels and resumed tumor growth. CRPC is a more aggressive type of cancer and the hallmark of PCa tumor progression. These malignant prostate cells are more resistant to apoptosis and have acquired proliferation stimulatory activity despite very low levels or absence of androgen.[65,66]

The review article by Feldman *et al.* categorizes five potential mechanisms of prostate cell androgen independence development (Figure 6).[67] These mechanisms illustrate that progression of PCa cells to androgen independence likely involve modification of the AR. Although it is regarded as an important transcription factor in the androgen signaling axis, AR relies on coactivator proteins, which may be exploited as inhibitory regulators, to activate its target genes.[68]

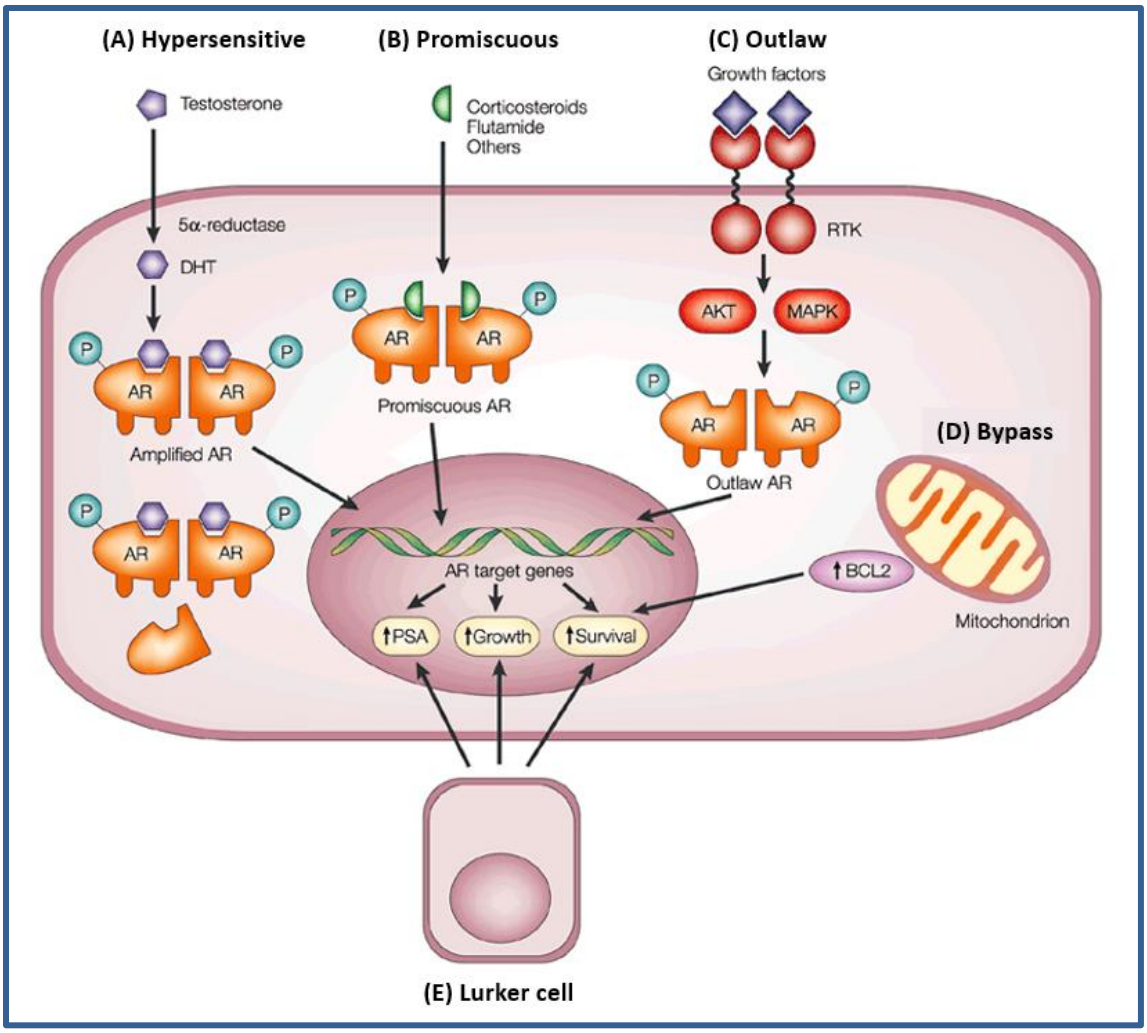


Figure 6: Five possible mechanisms of prostate cell androgen independence. A) The hypersensitive pathway involves: 1) increased production of androgen receptor (AR); 2) increased AR sensitivity; and/or, 3) more testosterone is converted to dihydrotestosterone (DHT) by 5α-reductase. B) In the promiscuous pathway, the specificity of AR is lessened where non-androgenic molecules can activate the receptor. C) In the outlaw pathway, receptor tyrosine kinases (RTKs) become activated by growth factors which subsequently results in AR phosphorylation (activation) by protein kinase B (AKT) or mitogen-activated protein kinase (MAPK). D) Independent survival pathways, *e.g.*, the anti- apoptotic protein B-cell lymphoma 2 (BCL2), bypass the need for AR. E) Androgen deprivation therapy establishes a cell selection advantage to lurker cells, *i.e.*, androgen-independent cancer cells, which are present all the time.[67]

The AR is a critical factor necessary for growth and survival of CRPC.[67–70] For example, even after ablation therapy most androgen independent PCa cases maintain high levels of AR expression.[71] Targeted AR inhibition has yielded a decrease in cell proliferation and survival in experiments using representative cell-based models of CRPC.[72–75] These findings support the proposed mechanism that AR may play a role as a cytoplasmic signaling molecule that is either ligand-independent (Figure 6.C) or non-genotropic.[76] It is not likely that malignant prostate cells use just one mechanism exclusively to maintain homeostatic regulation in the absence of androgen but instead use several mechanisms in a multi-step fashion at different stages of CRPC progression. Nonetheless, each of the pathways summarized in Figure 6 represent potential targets for therapeutic strategies that block abnormal stimulatory activity, which promote cell proliferation and inhibit cell death, and restore malignant cells back to normal cell regulation. In which case, conventional cancer treatments would be more effective against advanced PCa.

Studies have found that the tissue microenvironment (TME) is also critical for the progression of localized disease to distant metastases.[77–80] An important aspect of the TME for tumor progression includes autocrine and paracrine feedback loops, *i.e.*, cross-talk, between the PCa epithelial cells and stromal cells. The propensity of PCa to metastasize in bone tissue, which confers a microenvironment rich in growth factors and cytokines, supports these findings.[81] Not surprisingly, nearly 100% of men with advanced, high grade PCa develop metastases in regional lymph nodes and bone.[82] Paget's work may have introduced the tumor

"seed" and host "soil" hypothesis that describes the relationship allowing cancer dissemination but, more recent work has demonstrated that the bidirectional interactions between prostate cancer epithelial cells and stromal cells is much more dynamic than originally described.[83] There exists a reciprocal relationship between host and cancer cells where the two cell types can modulate the molecular function and behavior of each. Although host cells can be recruited to or transformed at the primary cancer site, it is the constant interaction with a "rich milieu" of host factors that ultimately change the cancer cells' behavior and drives migration, *i.e.*, metastasis, and invasion of host stromal cells.[78,84–89] Because survival of prostate cancer cells is supported by the host microenvironment during androgen ablation therapy, a co-targeting strategy that affects both cancer and supportive host cells may be the most effective tactic for preventing PCa progression.[78]

PCa Screening and Prognosis

Early detection and treatment of PCa can increase survival rates and extend a patient's life after diagnosis. To be effective, early screening tests need to be accurate and reliable at identifying clinically significant PCa at an early stage, *i.e.*, when treatment may be more efficacious. Although there are no known preventative measures for PCa, there are two widely used screening methods for detecting the disease: the digital rectal exam (DRE) and the prostate-specific antigen (PSA) test. Clinicians use the DRE as a relatively simple test to check the prostate for any physical abnormalities, *e.g.*, enlarge, hard and lumpy areas, and the PSA blood test. Since its approval in 1994 by the Federal Drug Administration (FDA) as a diagnostic

marker, the PSA test has become the most extensively used screening tool for PCa in the world. The success of PSA as a biomarker for PCa is due to its biology; it is expressed almost exclusively by the prostate gland and its amount in circulating blood increases as a result of disease pathogenesis.[90–92] However, its biology is also the precise reason that makes PSA an imperfect biomarker and one that may not be suitable for routine PCa screening.

Prostate-specific antigen is synthesized by the normal prostate epithelial secretory cells and released into seminal fluid but minimally in the circulatory system. The concentration of PSA is a million-fold higher in seminal fluid than in blood of healthy adult males.[93] However, the pathogenesis of PCa changes the glandular architecture of the prostate and causes PSA in the blood to increase. Specifically, PCa disrupts the basal cells and basement membrane of the prostate which allows secreted PSA in the prostatic lumen to diffuse more freely into peripheral circulation. Therefore, the relative risk of PCa increases with higher levels of serum PSA. For example, PCa has been detected in 6.6% of men with a PSA less than 0.5 ng/mL and in 27% of men with PSA between 3.1 and 4.0 ng/mL.[94] Similarly, a large European study reported a cumulative PCa detection rate of 0% to 12.3% for men with a PSA of 0 to 1.99 ng/mL, 21% to 25% at 2.0 to 2.99 ng/mL, and 33% at PSA 3.0 to 3.99 ng/mL.[95] Furthermore, advanced stages of PCa are associated with progressively worse glandular differentiation characterized by the absence of a basal cell layer and intervening stroma cells. As such, a strong positive correlation exists between serum PSA concentration and the degree of tumor differentiation based on histological grade using the Gleason system.[96] Correspondingly, PSA is associated with stage

and grade in localized disease and is higher in patients with metastatic disease.[97–99] Therefore, higher levels of PSA are generally associated with poorer PCa prognoses. However, the amount of PSA detected in the blood of affected males is highly variable due to the heterogeneous nature of PCa. Additionally, PSA levels are influenced by several factors unrelated to PCa in all men. Consequently, the effectiveness of PSA as a screening tool is very limited because of its poor sensitivity and specificity.

The potential of using PSA as a screening tool for PCa was recognized over three decades ago and was shortly followed by a generally accepted consensus to use 4.0 ng/mL as the standard reference threshold.[100] However, using a single cutoff value to decide whether or not a patient should undergo further, more invasive testing for PCa diagnosis is not pertinent. Prostate-specific antigen levels in blood are influenced by several factors, *e.g.*, age, body mass index and race/ethnicity, and can be highly variable between different individuals even in the absence of disease. [93,101] Moreover, PSA may be elevated as a result of benign pathological conditions of the prostate, *e.g.*, BPH and prostatitis.[102] Consequently, the sensitivity and specificity of PSA for detecting PCa is severely limited. For example, findings from the Prostate Cancer Prevention Trial (PCPT) affirmed that the 4.0 ng/mL cut-off point only had a sensitivity of 20.5% for PCa detection.[103] At this level, nearly 80% of PCa cases are missed using the PSA screening test. In three other large cohort studies, the range of PCa incidence among men with a PSA level above 4 ng/mL was between 27% and 44%.[104–106] Moreover, it has been shown that the specificity of the PSA test is equally low. One study reported a false positive rate of up to

70% in men beyond their fifth decade of life. [107] For these reasons, the standard 4.0 ng/mL reference threshold has been scrutinized because of its poor predictive value. Indeed, most men who undergo prostate biopsy due to an elevated PSA level in the blood do not have PCa.[108] However, a low PSA test result does not necessarily indicate an absence of PCa. For example, one study found that 17% of the men in their cohort had PCa even with low PSA (1.1 to 2.0 ng/mL) and normal DRE results.[94] Despite its poor performance as a true PCa biomarker, most men over the age of 50 have been screened using the PSA test resulting in the detection of millions of PCa cases. Approximately 90% of PCa diagnoses in the U.S. are derived from early screening and the remaining ten percent are clinically diagnosed.[109] However, three decades after its introduction, the effectiveness of the PSA test to lower PCa mortality has become a topic of much debate.[110]

The PSA test was originally approved by the FDA in 1986 but not as a diagnostic marker for PCa until 1994. It has since become the most extensively used screening test for PCa in the world. An estimated 75% of men, 50 years of age or older, in the U.S. were screened using the PSA test from the late 1980s to 2001.[111] As a consequence, a man's risk of being diagnosed with PCa nearly doubled from nine percent in 1985 to 16% in 2007.[112,113] Despite the rapid increase of PCa incidence associated with the establishment of widespread PSA testing, it, as well as mortality, have been on a steady decline. For example, incidence and mortality decreased by 2.0% and 3.1%, respectively, between the years 2006 and 2010.[1] Additionally, systematic screening for PCa has been attributed to stage migration, *i.e.*, a decline in the proportion of

metastatic and locally advanced cases of PCa detected at diagnosis.[114] However, the possibility of lead time bias may give the false impression that PSA testing results in better outcomes when in reality it may have no effect. Indeed, results of the Prostate, Lung, Colorectal, and Ovarian (PLCO) Cancer Screening Trial on PCa mortality, a large U.S. based randomized study of 76,693 men, showed that PSA and DRE testing does not reduce the death rate of PCa.[115] However, the European Randomized Study of Screening for Prostate Cancer (ERSPC) concluded the opposite: PSA testing resulted in a relative reduction in PCa mortality by 20%.[116,117] Although this may seem a large decrease, researchers from the ERSPC study estimate that about 1,410 men would need to be screened (and 48 treated) in order to prevent one death from PCa. Additionally, the moderate improvement in mortality would only benefit those between the ages 55 and 69. The disparate results of the PLCO and ERSPC studies have become the centerpiece of an ongoing debate between advocates of routine PSA screening and its critics.[110] The U.S. Preventive Services Task Force (USPSTF) issued a recommendation against PSA testing for men in the general U.S. population, regardless of age, based largely on the PLCO and ERSPC findings. The independent panel of experts concluded that there is little to no reduction in PCa mortality from PSA-based screening and that the harms related to screening and subsequent diagnostic procedures outweigh its potential benefits. [118] However, expert groups, *e.g.*, the American Cancer Society and American Urological Association, have maintained that early PCa screening saves men's lives and still recommend PSA testing. Although the PSA debate is ongoing, the discussion and controversy has shifted to a more salient concern: the era of PSA screening has resulted in an overwhelming burden of

morbidities and attendant healthcare costs indicative of overdiagnosis and overtreatment of PCa.

A substantial percentage of men who have asymptomatic PCa detected by PSA screening have a tumor that will not progress or become life threatening and represent cases of overdiagnosed disease. Men with non-life threatening, *i.e.*, indolent, PCa cannot benefit from cancer therapy. However, most men elect to treat screen-detected PCa despite receiving no clinical benefit and are unnecessarily subjected to the harms of PCa treatment. Some estimates indicate overdiagnosis of PCa is between 17% and 50%. [119] This suggests that the predictive value for PSA testing or the DRE to detect clinically significant PCa is severely limited. Indeed, no current PCa screening method can distinguish which tumors will remain indolent from those that will progress to aggressive disease. As a consequence, routine PCa screening has potentially harmful repercussions. Abnormal DRE or PSA test results usually lead to invasive prostate tissue biopsies which can cause bleeding, pain, infection, transient urinary difficulties, temporary impotence or other issues requiring clinician follow-up. [120,121] A man is much more likely to be diagnosed with PCa who undergoes early screening tests.[122] Over half of PCa diagnoses from PSA screening are localized, not clinically significant and will likely not become life-threatening. Regardless, 90% of men with screen-detected PCa will undergo immediate treatment which includes radical prostatectomy, radiation and/or androgen ablation therapy.[123,124] These types of PCa treatments are associated with unnecessary risks of urinary, sexual, and bowel dysfunction, which can adversely affect the quality of life of the patient

without any clinical benefit.[125] Given the potential risks of routine PCa screening, practitioners have adopted watchful waiting and active surveillance strategies based on recommendations from the American Cancer Society and the American Urologic Association. Theirs is a more conservative approach which considers family health history and other covariates in a risk-stratified assessment model. Identifying a patient's PCa risk status permits a more personalized health care plan that maximizes benefits over risks. Therefore, the likelihood of early screening-detected PCa for at risk men increases and men with no or low risk are spared unnecessary invasive procedures. The same watchful waiting and active surveillance approach can be applied to men diagnosed with a specific type of PCa. Cases that are localized to the prostate, low-grade and slow progressing may benefit from careful observation and may never require treatment. However, most men in the U.S. with low-risk PCa will receive up-front, aggressive treatment without gaining any improvement in survival or decreased morbidity.[126,127] Indeed, over 50% of screening-detected and diagnosed PCa cases in U.S. men over the age of 65 are excessive.[128] Therefore, the overdiagnosis and overtreatment of PCa emphasize the need for improved screening methods that can distinguish between indolent cancer and PCa cases that will become life threatening.

Many blood and urine biomarkers have been developed for PCa detection but only PSA is routinely used by clinicians. Most PCa biomarkers that are reported to be biomedically and statistically significant are largely not reproducible by independent researchers and never gain FDA approval.[129] Nevertheless, the current PCa overdiagnosis and overtreatment crisis

underscores the great need for a noninvasive biomarker that can accurately detect PCa and differentiate between indolent cases and those that are likely to become clinically significant. Currently, the strongest prognostic predictor for PCa progression relies on histologic grading of the prostatic tumor using the Gleason system. However, obtaining tissue specimens for histological examination requires the patient to undergo an invasive and potentially harmful procedure, *i.e.*, needle biopsy or radical prostatectomy. For men who have indolent PCa that will not progress to an advanced stage, invasive diagnostic procedures are extraneous. A PCa biomarker found in peripheral blood is highly desirable as it is minimally invasive. In order for it to have a high diagnostic accuracy, specificity and sensitivity, the PCa biomarker detected in blood must be produced by the tumor or by the body in response to the tumor. Mitochondrial dysfunction may be a potential marker for age-related diseases such as cancer. [130] Specifically, changes to the mitochondrial genome (mtGenome), as a function of disease pathogenesis, may be measured and used as a molecular marker. It is clear that mitochondrial dysfunction is a component of PCa and recently has been studied with great interest as potential biomarker.

Mitochondrial Dysfunction in PCa

Mitochondria are a specialized organelle critically important for key cellular processes involved in homeostatic regulation and overall health of the organism. Specifically, mitochondria are an integral unit in the processes of cell signaling, synthesis of iron-sulfur clusters and heme prosthetic groups, cellular proliferation and, conversely, apoptosis.[131–137] Most importantly, mitochondria are the cellular powerhouse that generates adenosine

triphosphate (ATP), the fundamental unit of energy required to fuel all cellular processes and functions. The final step of ATP production via aerobic respiration is oxidative phosphorylation (OXPHOS), a complex process which utilizes carbohydrate, lipid and protein metabolites and oxygen in a series of redox reactions that drive ATP synthesis.

Unlike other organelles, mitochondria contain their own circular genome which is approximately 16.5 kb (kilobases) (Figure 7).[138,139] The number of mitochondrial organelles present in a cell varies greatly, up to an eight-fold difference, and is largely dependent on the energetic demands of the cell type. Furthermore, each mitochondrion contains between one and ten copies of mtDNA.[140,141] Consequently, a single cell can contain hundreds of thousands of mtDNA copies; a mature human oocyte can have over 150,000 copies of mtDNA whereas human spermatozoa contain only 323 - 1,111 copies.[142,143] Although the mechanism(s) involved with mitochondria and mtDNA copy number regulation are not clearly understood, it is evident the process is dynamically coordinated, in a cell-specific manner, to maintain energy homeostasis.[144,145] The number of mtDNA copies per cell is an index for mitochondrial biogenesis and is closely regulated in normal functioning mitochondria.[140,144] As such, cell maintenance of mtDNA copy number is critical for normal cellular function. Mitochondrial DNA depletion syndromes (MDSs) are a heterogeneous group of severe disorders characterized by reduced mtDNA content associated with impaired energy production.[146] Mitochondrial DNA copy number dysregulation has also been implicated in age-related diseases including

neurodegenerative diseases and many cancers. Specifically, decreased mtDNA content has been measured in gastric, renal, and hepatocellular tumors.[147–150]

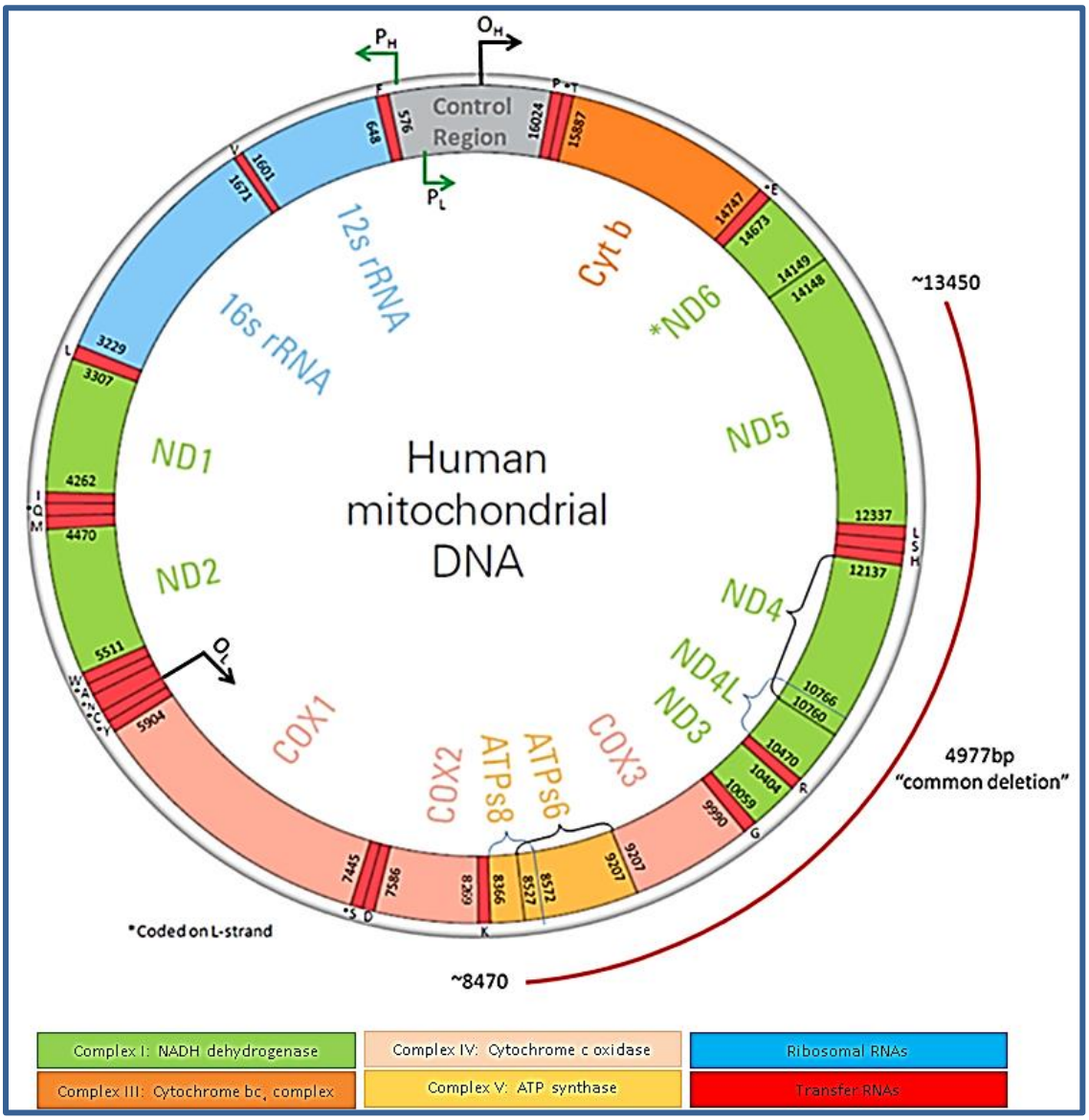


Figure 7: Schematic of the human mitochondrial genome. Heavy strand origin of replication (O_H), heavy strand promoter (P_H), light strand origin of replication (O_L) and light strand promoter (P_L). The mitochondrial genome encodes 13 genes which are critical components of the OXPHOS complexes: seven subunits of Complex I (ND1, ND2, ND3, ND4, ND4L, ND5, ND6), one subunit of Complex III (Cyt b), three subunits of Complex IV (COX1, COX2, COX3), and two subunits of Complex V (ATPs6, ATPs8). Reported break points for the 4,997 base pair (bp)

"common" deletion vary, therefore approximate positions have been indicated. All 23 amino acids encoded by the mitochondrial genome are labled by their one letter abbreviation. All genes are encoded on the heavy strand with the exception of ND6.

A study by Mizumachi *et al.* used qPCR and laser microdissection to quantify the amount of mtDNA in individual PCa cells and adjacent normal cells from cancer tissue specimens.[151] Their results showed that cancer cells exhibited a higher degree of variance in mtDNA content and the total amount of mtDNA was greater in 78% of the PCa cells compared to normal tissue cells. Results of a different study demonstrated that a decrease in mtDNA and respiratory function transforms androgen-dependent cells to androgen-independent cells; a hallmark of advanced PCa progression.[152] In this study, the quantity of normal mtDNA was greatly reduced and large deletion mutant DNA was greatly increased following androgen ablation of an androgen-dependent PCa cell line (LNCaP). When mtDNA-depleted clones were reconstituted with normal mtDNA, androgen-dependence was reversed. Although the scope of this study did not investigate the difference in mtDNA copy number between PCa cells and normal prostate cells, it is interesting that, in contrast to the previous study, a decrease in mtDNA content was associated with PCa progression from androgen-dependence to androgen independence. These results suggest a troubling notion that androgen ablation treatment may actually contribute to the progression of PCa from androgen-dependent to CRPC.

Large-scale deletions of the mtGenome result in the loss of gene products critical for OXPHOS. The resultant organellar dysfunction impairs energy production and causes mitochondrial diseases such as Kearns-Sayre syndrome (KSS). Large mtDNA deletions have also

been associated with normal aging and age-related diseases.[153–157] Indeed, a "common" deletion of 4,977 bp, originally identified in KSS, has been observed at varying levels in post-mitotic, aging tissues (Figure 7).[158,159] Interestingly, large-scale deletions are predominantly located within the major arc of the mtGenome and are flanked by direct or indirect sequence repeat motifs at the deletion junctions.[160] Although the cause of large-scale deletions is unclear, sequence repeat motifs have been implicated in two different biological mechanisms: one involving deficient mtDNA replication and another involving mtDNA damage repair.[160–163] Nonetheless, large-scale deletions cause cellular pathology due to an OXPHOS and energy deficiency. However, the multi-copy nature of mtDNA requires that the proportion of mtDNAs harboring large-scale deletions exceed a critical threshold in order to elicit mitochondrial dysfunction.[164] The compensatory effect of wild-type (whole) mtDNAs allows low levels of mtDNA deletions to be tolerated. Therefore, the longitudinal accumulation of large-scale mtDNA deletions is likely a characteristic of the mitochondrial disease etiology.

The 37 genes encoded by the mtGenome include protein subunits of the electron transport chain (ETC) and are essential for ATP production via OXPHOS. This process involves the transfer of electrons through a series of coupled redox reactions which ultimately results in the final reduction of molecular oxygen to water. A consequence of this process is the inevitable generation of reactive oxygen species (ROS), primarily superoxide, from complexes I and III of the ETC. Superoxide in the mitochondria matrix is converted to hydrogen peroxide (H_2O_2) by manganese super oxide dismutase (Mn-SOD) which is a less reactive and a more

stable ROS. However, the hydroxyl radical, the most reactive ROS, is also formed. The Fenton reaction is one mechanism that is well characterized and involves the iron-catalyzed reaction of superoxide and H_2O_2. Reactive oxygen species in the mitochondria matrix react directly with several biomolecules thus causing damage to enzymes, fatty acid side chains of membrane lipids and nucleic acids.[165–169] Different ROS produce many types of DNA damage including base modifications, abasic sites, strand breaks and deletions.[170] Mitochondrial DNA is susceptible to ROS induced damage primarily due to its close proximity to the ETC which is the major site of ROS production. Also, unlike the nuclear genome, mtDNA is not associated with protective histones and does not have a significant DNA repair system. Therefore, mitochondria lack the ability to repair all the different forms of mtDNA damage.[171] For these reasons, the mutation rate of mtDNA is estimated to be approximately 10-fold greater than that of nuclear DNA.[172] If not repaired, ROS induced damage to mtDNA can accumulate overtime and cause mitochondrial dysfunction such as decreased energy metabolism. Another consequence of accumulated mtDNA damage and mitochondrial dysfunction is the perpetual production of ROS which, over time, leads to a persistent and modestly progressive increase in oxidative stress. This process constitutes the fundamental basis of the free radical/mitochondrial theory of aging and has also been implicated in carcinogenesis and cancer progression .[173–175]

Research investigating mitochondrial dysfunction has focused on whether or not oxidative stress is associated with prostate cancer, its role in the development and progression of the disease and specifically what are the mechanisms involved that cause increased ROS

generation. It is very clear that oxidative stress is involved with PCa etiology. Higher levels of oxidative stress have been measured in the prostate epithelium of men with PCa than men without the disease. Figure 8 illustrates some of the proposed mechanisms that lead to increased ROS production and the corresponding response by prostate cells. This diagram also suggests that PCa, like many other age-related diseases, may be a consequence of the progressive increase of endogenous ROS and resultant somatic mitochondrial alterations, *i.e.*, lipid, protein and DNA damage. Cellular ROS can also function as a signaling molecule to induce cell toxicity and death or cell growth and proliferation.[176–178] Activation of the signaling pathways involved in proliferation, senescence and apoptosis are thought to be induced by intracellular changes in ROS levels.[179] Specifically, increased oxidative stress may play an important role in the neoplasmic transformation, aberrant growth and proliferation necessary for PCa development.[180,181]

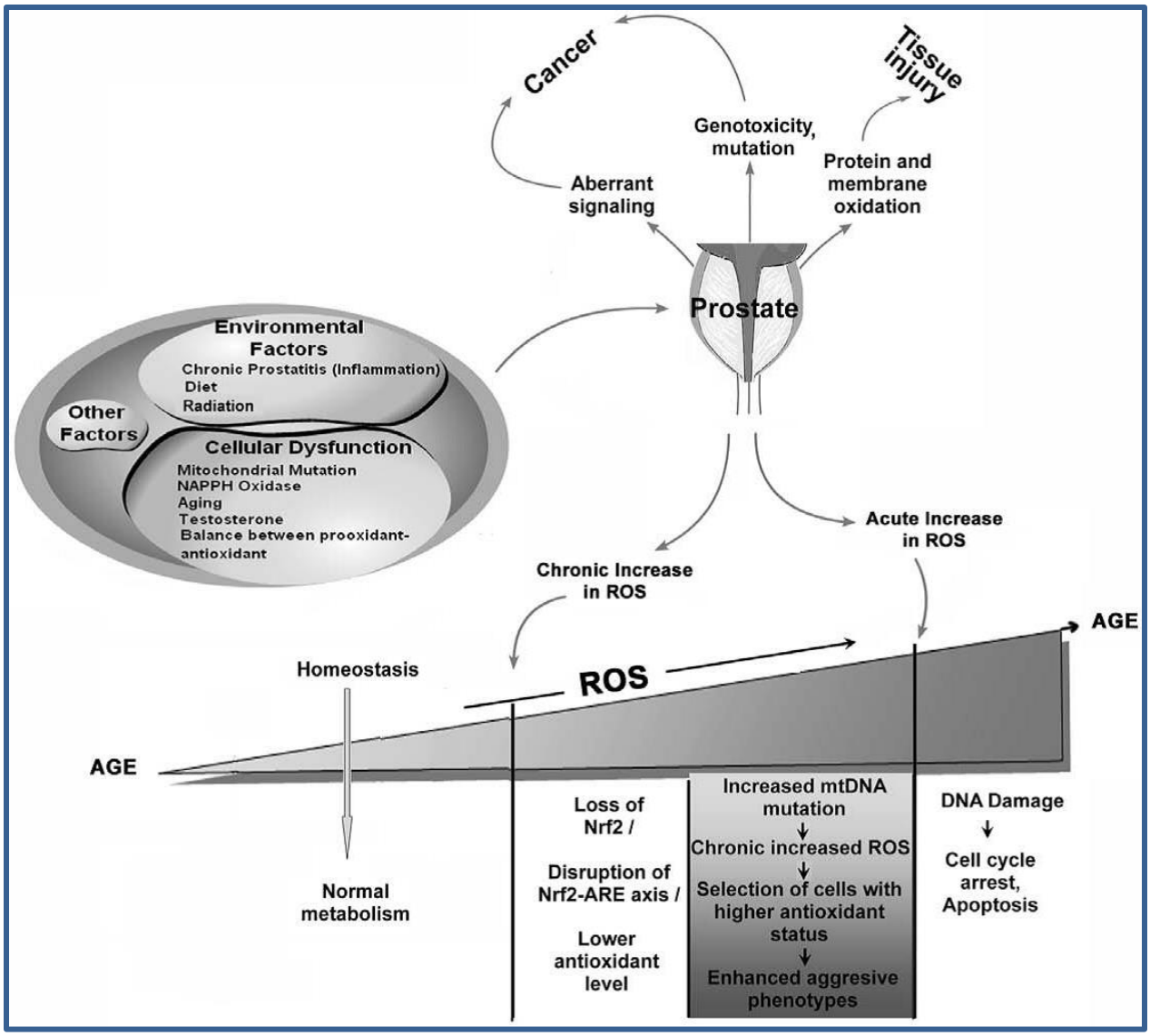


Figure 8: Possible mechanisms of ROS production and prostate cellular response. Several factors can lead to ROS production in the prostate. However, in normal functioning cells an endogenous antioxidant defense system counteracts and regulates ROS to maintain physiological homeostasis. Lowering ROS below the homeostatic threshold may depress normal metabolism and affect cell proliferation. In contrast, an increase in ROS may exhaust the endogenous antioxidant level (by disrupting Nrf2-antioxidant response element axis), increase mtDNA mutation and aggressive phenotypes, and cause DNA damage.[182]

A number of mtDNA mutations have been found in various cancers and seem to alter mitochondrial metabolism, enhance tumorigenesis and permit cancer cell adaptation to changing environments.[183–186] The tissue-specific manifestation of PCa may be the pathophysiological consequence of systemic mitochondrial dysfunction where retrograde signaling and metabolic shift promote tumor initiation and progression by driving cellular proliferation and inhibiting apoptosis.[187] Mutations in mtDNA are frequent in PCa and seem to occur early during prostate malignant transformation.[188,189] A study by Petros *et al*. reported that mutations in the ETC Complex IV (cytochrome c oxidase) gene are significantly more frequent in prostate tissue biopsies from PCa patients than no cancer controls.[190] Cybrid experiments have shown that PCa cells harboring a mutation in the ATP synthase F0 subunit 6 (ATP6) gene have increased levels of ROS, an increased rate of tumor growth and generate tumors seven times larger than wild-type in nude mice.[190] These studies support the posit that specific mtDNA mutations manifest into deficient ETC complex subunits, cause mitochondrial dysfunction and result in increased ROS production. Indeed, increased ROS has been implicated as a factor in tumor formation through a signaling transduction pathway that inactivates p16Ink4A and p53.[191] These findings are supported by a more recent article that reviews studies showing mitochondrial oxidative stress drives tumor progression and metastasis.[192] A proposed mechanism, "The Autophagic Tumor Stroma Model of Cancer Metabolism," states that increased levels of ROS drive mitophagy resulting in substrate level phosphorylation, *i.e.*, aerobic glycolysis, in fibroblasts. The byproducts of aerobic glycolysis, such as lactate, pyruvate,

ketones and glutamine, fuel adjacent cancer cells thereby promoting tumor growth and protection from apoptosis.[193]

Project Overview

Molecular and genetic alterations are associated with PCa development, progression and metastasis; the further the progression, the further these molecular changes can occur that help define the cancer. Specifically, cancer cells undergo a metabolic shift known as the "Warburg effect" associated with a seemingly limitless proliferative potential.[194,195] This suggests that a bioenergetic component may have a role in the pathogenesis of the disease. Prostate cancer development has been associated with mitochondrial dysfunction where the delayed onset and progression of the disease may be the result of an accumulation of mutations to the mtGenome. Several studies have reported somatic changes in mtDNA of PCa patients in both prostate tumor tissue and bodily fluids. As there is a critical need to develop improved detection methods for PCa that can distinguish indolent from aggressive disease, mtDNA may be a suitable molecular marker. For instance, mtDNA is potentially a far superior disease screening biomarker due to its increased sensitivity; cells can contain a thousand-fold more copies of mtDNA than nuclear DNA. However, genetic alterations to mtDNA and their influence on the development and progression of PCa are not clearly understood. This is in large part due to the limited scope of research from previously published literature. Furthermore, using changes to mtDNA as a molecular marker to predict PCa progression from an indolent stage to a more aggressive phenotype has not been fully explored.

The purpose of this research project is to address these key gaps in existing knowledge by performing an exhaustive characterization of the changes to mtDNA associated with PCa development and progression of a scale larger than has been previously reported. A unique comprehensive strategy, one that combines multiple aspects of the mtGenome, is used to achieve two specific research goals: (1) demonstrate marked differences in the type and amount of changes to the mtGenome among different stages of PCa; and (2) show associations between measures of changes to the mtGenome and PCa progression and thus establish a basis for using mtDNA as a molecular marker for disease prognosis. The project's impetus is based on the following hypothesis:

> ***Mitochondrial dysfunction contributes to the development of PCa where systemic, longitudinal changes to the mtGenome occur that delineate PCa progression. Consequently, increased alterations to mtDNA are associated with PCa progression as compared to mean age-matched normal controls (NC). Changes in mtDNA can be detected in peripheral blood which may be used as a biomarker for predicting PCa progression to a more aggressive phenotype.***

This hypothesis is tested using two independent specific aims that focus on: (1) group differences in the mtGenome population, e.g., copy number expansion, depletion, and/or change in variance, as well as the ratio of whole or "wild-type" mtGenomes to those that

harbor a large-scale deletion in the major arch; and (2) whole mtGenome sequencing to determine group differences in overall mutational load with an emphasis on missense mutations that are predicted to be pathogenic.

A series of studies are presented that use genetic and bioinformatic techniques to characterize mtGenome differences in peripheral blood of men screened for PCa. The scope of Chapters II and III address the first specific aim whereas Chapters IV and V pertain to the second specific aim. Chapter II reports on an internal validation and a small, preliminary study that includes measuring the concentration of mtDNA using qPCR. The results of these two studies establish the feasibility of using qPCR to measure the quantity of mtDNA in blood samples for subsequent PCa studies. More importantly, Chapter II identifies strengths and weaknesses of qPCR when measuring mtDNA which were applied in the design and development of a much larger qPCR study (presented in Chapter III). Using knowledge gained from the studies presented in Chapter II, a novel multiplex qPCR assay was developed specifically to assess mtDNA copy number and the ratio of whole mtDNA to mtDNA harboring a large-scale deletion. Chapter III briefly introduces the development of this assay and then presents a study that uses it to assess the differences between three separate groups of men screened for PCa; a normal control (NC) group, an indolent PCa group, designated “no evidence of disease” (NED), and an advanced PCa group, designated “biochemical recurrence and metastasis” (BCR/MET). Next-generation sequencing is performed on the same cohort to generate whole mtGenome sequence haplotypes for each sample in the NC, NED and BCR/MET groups. In Chapter IV, an in-

depth characterization of the observed genetic variants, as well as differences in mutational load between groups, are presented. And finally, Chapter V details the last study of this research project which expands on the results obtained in Chapter IV. Whole mtGenome sequence data and bioinformatics are used to generate three-dimensional (3D) models of a wild-type and mutant mitochondrial protein which are then tested *in silico* to determine the mutational effect on structural energy states resultant of a single mtDNA variant.

The innovative aspects and relevance of this research project distinguish it from other PCa studies. The qPCR assay developed for this project is novel in that it provides the same amount of information in a single multiplex as what normally requires more than one qPCR assay. Further, appending genetics with proteomics to translate the mutational effect of an observed mtDNA variant, to the extent presented here, has not been accomplished before. But most distinctive is the research strategy: a comprehensive approach was applied to assess multiple aspects of the mtGenome among PCa subjects representative of different developmental stages of the disease. Specifically, mtDNA copy number, deletion ratio, and mutational load are used as indicators of mitochondrial dysfunction in blood and, collectively, may prove useful as a noninvasive molecular marker to predict PCa progression. Although mitochondrial dysfunction associated with PCa has been researched for a decade or more, the extent of mtDNA alterations associated with the disease that are present in blood cells is not known. Also, a very important and unresolved problem in PCa is the ability to differentiate between indolent and aggressive disease using a molecular marker. By addressing these two

topics, the studies presented here are relevant to the overarching challenges in PCa research today and contribute an incremental advancement to the existing, limited body of knowledge regarding the role of mitochondrial dysfunction in PCa development and progression.

CHAPTER II

QUANTIFICATION OF MITOCHONDRIAL DNA: PRELIMINARY STUDIES

Mitochondrial dysfunction causes a broad spectrum of disorders and is associated with aging and age-related diseases. Specifically, its role in cancer has become increasingly significant as our understanding of mitochondrial involvement in the cell cycle, gene expression, metabolism, cell viability, and other aspects of cell growth and stress responses become clearer. The amount of mtDNA in a cell represents an index for mitochondrial biogenesis whereby a change may indicate organelle dysfunction. Likewise, a high proportion of mtDNAs harboring large-scale deletions may be indicative of an impaired OXPHOS system. Collectively, mtDNA copy number dysregulation and the accumulation of large-scale mtDNA deletions are potential indicators of mitochondrial dysfunction and may signify a pathological state. Experiments using qPCR were performed to measure the amount of mtDNA and the proportion of mtDNAs harboring large-scale deletions in biological specimens derived from PCa patients and matched normal controls. In this chapter, two studies are presented that involve a mtDNA qPCR assay developed by Kavlick *et al*.: (1) an internal validation and, (2) a small, preliminary study.[196] Lessons learned from these two studies were applied to the development of a more informative qPCR assay which was then used in a larger, more comprehensive study (Chapter III).

Study 1: Internal Validation of a mtDNA qPCR Assay

Real-time qPCR is a reliable, highly specific, flexible and sensitive method to quantify DNA.[197] Several qPCR assays have been developed specifically for mtDNA quantification which have many different biomedical applications.[196,198–201] An internal validation was performed on the assay described by Kavlick *et al*. following the quality assurance standards used in forensic DNA testing laboratories.[202,203] The purpose of this study was to demonstrate that the assay performed as expected, *i.e.*, as described by Kavlick *et al*., and is suitable to be used to measure the concentration of mtDNA in sample preparations for PCa research. A report by Sprouse *et al*. details the internal validation in its entirety but presented here are three components of the study: 1) reproducibility and precision; 2) sensitivity; and 3) contamination assessment with a known inhibitor.[203]

Materials and Methods

Samples

Human genomic DNA from cell line HL-60 (ATCC, CCL-240D) was purchased from American Type Culture Collection (Manassas, VA). The primary stock of HL-60 DNA was reconstituted with TE^{-4} (10 mM Tris-HCl, pH 8.0, 0.1 mM EDTA).

Following approval from the Office for the Protection of Human Subjects-Institutional Review Board (Project #2010-120; "Quantification and Analysis of Human mtDNA from

Different Sample Types"; UNT Health Science Center, Fort Worth, TX), DNA from peripheral blood of a single individual was extracted using an organic protocol.[204] A total of 100 µL of whole blood was incubated with 500 µL of stain extraction buffer (10 mM Tris, 100 mM NaCl, 39 mM DTT, 10 mM EDTA and 2% SDS) and 5 µL of proteinase K (20 mg/mL) for 2 hours at 56˚C. Following incubation, the blood sample was extracted with an equal volume of phenol/chloroform/isoamyl alcohol (PCIA) (25:24:1, v/v) and purified by ethanol precipitation. DNA was reconstituted with molecular biology grade water at a final recovery volume of 100 µL and stored at 4˚C.

Real-Time qPCR

All qPCR experiments were performed on two 7500 Real-Time PCR Systems over a 9-month period. Data were analyzed with Sequence Detection Software (SDS) version 1.2.3 (Applied Biosystems, Foster City, CA). Quantification of nuclear DNA (nDNA) was accomplished using the Quantifiler® Human DNA Quantification Kit (Applied Biosystems) following the manufacturer's recommended protocol.[205]

Mitochondrial DNA was quantified using a previously described assay with minor modifications.[196] The mtDNA amplification target of this assay is a 105 bp region within the NADH dehydrogenase subunit 5 (ND5) gene and corresponds to positions 13,288 to 13,392 of the revised Cambridge Reference Sequence (rCRS) (rCRS-GenBank: AC_000021 gi: 115315570). The polyacrylamide gel electrophoresis (PAGE) purified synthetic oligonucleotide

standard (Integrated DNA Technologies, Coralville, IA) was diluted using TE^{-4} to generate standards 1 through 8, consisting of 1,000, 100, 10, 1, 0.1, 0.01, 0.001 and 0.0001 pg/µL, respectively. Amplification reactions consisted of 2 µL of template DNA in a final volume of 25 µL. The thermal profile of the 7500 Real-Time PCR System was 50˚C for 2 min, 95˚C for 10 min, followed by 40 cycles of 95˚C for 15 sec and 60˚C for 1 min in 9600 Emulation Mode. All qPCR experiments consisted minimally of: each quantification standard in duplicate; a TE-4 no template control (NTC); and an HL-60 DNA positive control. Three different concentrations of positive control, based on nuclear (n) or mitochondrial (m) DNA quantification, were used: 1.0 ng/µL (n), 0.10 ng/µL (n) and 0.12 pg/µL (m).

Inhibition

A stock concentration of hematin (Sigma Aldrich, St. Louis, MO) was prepared by mixing 0.5 mM hematin with 0.1 N sodium hydroxide (NaOH). Re-purification experiments were performed on mtDNA samples using Microcon® YM-100 centrifugal filter devices (EMD Millipore Corp.) following the manufacturer's recommended protocol.

Each inhibition experiment included treated (with inhibitor) and non-treated (without inhibitor) samples where two variables were tested: inhibitor concentration and amount of mtDNA template. Nine concentrations of hematin and four different amounts of mtDNA template were tested: 0.001, 0.01, 0.1, 0.25, 0.5, 1, 5, 15 and 30 µM; and, 20, 2, 0.002 and 0.0002 pg/µL, respectively. For treated samples, the amplification master mix was spiked with

an inhibitor to achieve the desired concentration. Extracted DNA from peripheral blood of a single individual was used as the mtDNA template for treated samples and non-treated controls.

Inhibition of the mtDNA qPCR assay is expected to increase the quantification cycle (Cq) values for treated samples relative to that for non-treated controls. To assess the degree of PCR inhibition, the Cq value of the mtDNA target was compared between treated samples and non-treated controls respective to the amount of mtDNA template tested. A Cq value, designated as the inhibition assessment margin, was determined for each amount of mtDNA template using the average Cq plus 3 standard deviations (SD) for non-treated controls. Raw Cq values were not used but instead were linearly transformed (2^{-Ct}); the average, SD, and inhibition assessment margin were calculated on the transformed values.[206] A Cq value for the inhibition assessment margin was then calculated using a logarithmic transformation. Partial PCR inhibition was defined as a Cq value greater than the inhibition assessment margin and complete PCR inhibition was indicated by no Cq result.

Results and Discussion

Data generated for all eight quantification standards (Table 1), controls and linear regression of the standard curve (Table 2 and Fig. 9) from 18 separate qPCR runs demonstrated similarly high reproducibility, precision, and sensitivity as reported by Kavlick *et al*.[196] As

expected, standard 8 (0.0001 pg/µL) exhibited the highest degree of variation with an average Cq of 35.63 (SD±0.39) and a range of 1.52 cycles. However, standard 8 was detected in each experiment thereby demonstrating the minimum detection threshold of the assay is 0.0001 pg/µL. Given the total molecular weight of the human mitochondrial genome, approximately 1.6998×10^{-5}, the minimum detection threshold of this qPCR assay is equivalent to approximately twelve human mtDNA copies.[139]

Standard	mtDNA (pg/ µL)	Average	Standard Deviation	Minimum	Maximum	Range
1	1000	11.99	0.29	11.57	12.75	1.18
2	100	15.57	0.32	15.10	16.07	0.97
3	10	19.18	0.20	18.78	19.59	0.81
4	1	22.62	0.27	22.19	23.19	1.00
5	0.1	26.09	0.25	25.56	26.56	1.00
6	0.01	29.35	0.26	28.88	30.00	1.12
7	0.001	32.62	0.31	32.16	33.44	1.28
8	0.0001	35.63	0.39	34.79	36.31	1.52

Table 1: Quantification Standard Cq data from 18 separate qPCR runs.[203]

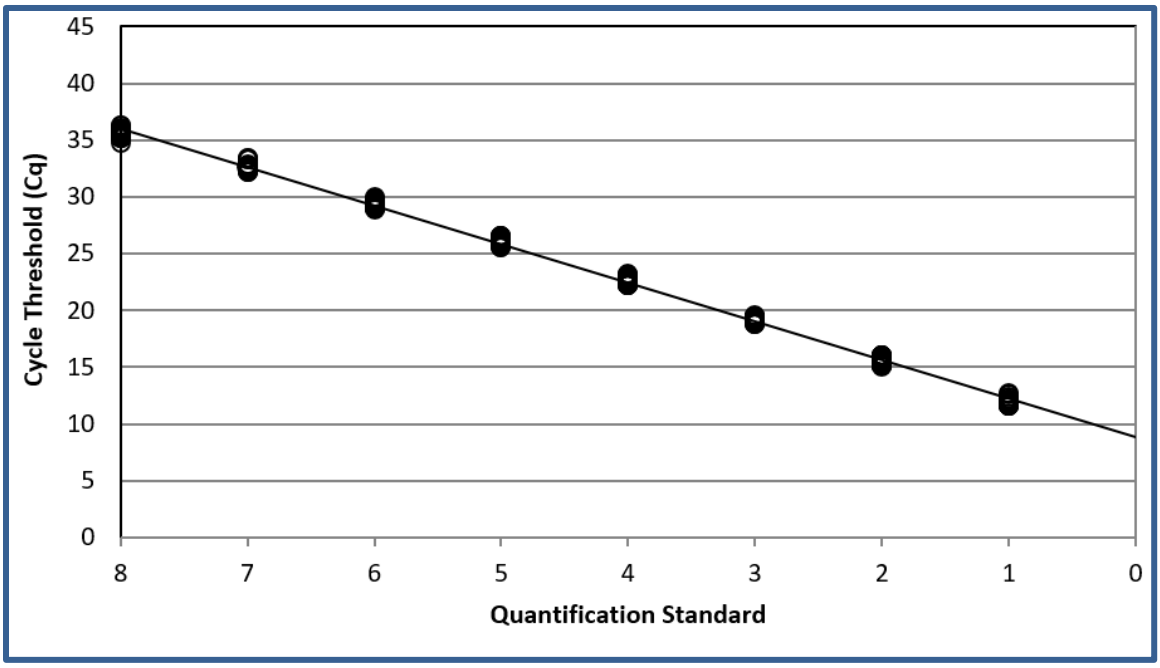


Figure 9: Reproducibility of standard curve. The plot represents an average standard curve generated from Cq values, corresponding to each quantification standard, of the mtDNA target. Data are from 18 separate qPCR runs where each quantification standard was assayed in duplicate. The trend line, representing the average Cq values, has an R^2 of 0.9993 and a slope of -3.39 which corresponds to an amplification efficiency of 97.2%. The ranges for individual quantification standards varied from 0.81 Cq (standard 3, or 10 pg/µL) to 1.52 Cq (standard 8, or 0.0001 pg/µL).[203]

Run	Slope	Amplification Efficiency	R^2	Y-Intercept
1	-3.44	95.26%	0.9987	22.90
2	-3.44	95.30%	0.9990	22.37
3	-3.42	96.07%	0.9989	22.16
4	-3.42	96.23%	0.9980	22.33
5	-3.40	97.01%	0.9981	22.43
6	-3.43	95.60%	0.9995	22.49
7	-3.39	97.43%	0.9986	22.59
8	-3.37	97.91%	0.9982	22.29
9	-3.34	99.44%	0.9977	22.71
10	-3.33	99.68%	0.9997	22.70
11	-3.37	98.14%	0.9989	22.50
12	-3.37	97.96%	0.9984	22.06
13	-3.42	95.89%	0.9997	22.15
14	-3.40	96.87%	0.9991	22.38
15	-3.37	98.15%	0.9993	22.35
16	-3.42	96.15%	0.9995	22.50
17	-3.40	96.74%	0.9985	22.46
18	-3.40	96.87%	0.9991	22.63
Average	-3.40	97.04%	0.9988	22.44
Standard Deviation	0.03	1.31%	0.0006	0.21
Coefficient of Variation	0.01	0.01	0.0006	0.01
Minimum	-3.44	95.26%	0.9977	22.06
Maximum	-3.33	99.68%	0.9997	22.90
Range	0.11	4.43%	0.0020	0.84

Table 2: Linear regression data from 18 separate qPCR runs.[203]

Three concentrations of highly purified DNA from cell line HL-60 were included as a positive control during each run of the mtDNA qPCR assay. These positive controls tested the assay's functionality and were used to monitor reproducibility and precision. Positive results were obtained for each HL-60 control in every qPCR run. All three positive controls exhibited low Cq and quantity estimate variation between runs. Furthermore, the human mtDNA qPCR assay estimated the mtDNA concentration of positive controls with 1 ng/µL and 0.1 ng/µL consistent with a 10-fold dilution.

A TaqMan® Exogenous Internal Positive Control (IPC) is integrated in this assay to detect inhibition of the PCR as opposed to a true negative sample result. The IPC components are incorporated into the assay's amplification master mix at a fixed concentration; the IPC primers are intentionally limited in order to minimize competitive inhibition of the mtDNA target. Consequently, the IPC is expected to amplify equally in every reaction, barring inhibition, and detected at a consistent Ct. Ideally, optimal performance of the IPC is expected from the NTC. Analysis of 30 NTCs resulted in an average Cq of 28.39 (SD±0.35) and a range of 1.31 cycles for the IPC. Based on these data, partially or moderately inhibited samples are defined as those having an IPC Cq value greater than 29.44. No detected Cq for a sample indicates complete amplification failure and suggests a high degree of inhibition.

Analysis of treated samples revealed that the mtDNA qPCR assay functioned normally despite the presence of a known PCR inhibitor up to a specific concentration termed the "tolerance threshold" (Table 3). The tolerance threshold represents the concentration at which partial or complete inhibition of the human mtDNA qPCR assay is expected. The Cq values remained stable up to 0.50 µM of hematin in the treated samples. However, increased Cq values were observed with higher concentrations of inhibitor suggesting the mtDNA qPCR assay was inhibited in a dose dependent manner. The degree of inhibition, none, partial or complete, was proportional to the concentration of inhibitor. Additionally, complete inhibition for higher concentrations of mtDNA template (20 and 2 pg/µL) required higher concentrations of inhibitor compared to lower amounts (0.002 and 0.0002 pg/µL). For example, 1.0 µM of hematin produced complete inhibition at 0.0002 pg/µL of mtDNA template but only partial inhibition with higher amounts of mtDNA template. Data generated from the inhibition experiments suggest the tolerance threshold for hematin is 1µM.

Increased IPC Cq values, however, can occur in the absence of inhibitors due to PCR selection, *e.g.*, inhibition by competition.[207–209] Similar to results reported by Kavlick *et al*., amplification bias of mtDNA over the IPC DNA was consistently observed in each experiment performed in the validation study where standards 1 through 5 (1,000 to 0.1 pg/µL) exhibited either increased IPC Cq values or failed IPC amplification. These results suggest that partial or complete inhibition of the IPC for samples with a mtDNA concentration greater than standard 5 (0.1 pg/µL) may not be due to the presence of inhibitors. Therefore, interpreting the IPC

requires careful consideration of the mtDNA concentration when determining whether or not an inhibitor is present in a sample.

Data generated from the inhibition experiments clearly demonstrate the mtDNA qPCR assay can function despite being challenged with a known PCR inhibitor. However, at certain concentrations of inhibitor, complete PCR inhibition occurred, suggesting a specific tolerance threshold value for hematin (Table 3). The highest amount of template tested (20 pg/µL) required a higher amount of hematin to elicit complete qPCR inhibition compared to the lowest amount of template tested (0.0002 pg/µL). Our results support a previous study which suggests hematin is a T*aq* polymerase inhibitor.[210] Theoretically, higher amounts of template and/or T*aq* polymerase can overcome, to some degree, PCR inhibition. It seems contrary then that reactions with lower amounts of template (0.002 and 0.0002 pg/µL) were not partially inhibited whereas those with higher amounts of template (20 and 2 pg/µL) were. For example, hematin at 1 µM had no inhibitory effect with a template amount of 0.002 pg/µL, but exhibited partial inhibition at 20 and 2 pg/µL. This disparity is due to our assessment margin calculation; the degree of variation in Cq values for the lower template amounts were such that partial inhibition determination was not possible. Complete inhibition is defined as a Cq greater than 40, representing the approximate cycle at which one mtDNA copy is detected given the very high efficiency of the assay. Therefore, increasing the number of cycles beyond 40 would not reconcile the ability to establish partial inhibition assessment for the 0.002 and 0.0002 pg/µL

template amounts. Furthermore, Cq values beyond 40 cycles exhibit a level of variation that precludes an appreciable level of certainty and therefore would be unreliable.

mtDNA Template (pg/µL)	[Hematin] µM								
	0.001	0.01	0.10	0.25	0.50	1.00	5.00	15.00	30.00
20	✓	✓	✓	✓	✓	!	×	×	×
2	✓	✓	✓	✓	✓	!	×	×	×
0.002	✓	✓	✓	✓	✓	✓	×	×	×
0.0002	✓	✓	✓	✓	✓	×	×	×	×

Table 3: Degree of qPCR inhibition for treated samples. Level of inhibition is represented by: no inhibition (✓) defined as a Cq value equal to or less than the inhibition assessment margin; partial inhibition (!) defined as a Cq value greater than the inhibition margin; and, complete inhibition (×) defined as no Cq result. [203]

Study 2: Mitochondrial DNA Concentration of PCa and BPH Blood and Prostate Tissue Samples

Altered mtDNA content may indicate mitochondrial dysfunction associated with cancer development and progression. Circulating peripheral blood leucocytes (PBLs) are the predominant source of mtDNA in blood and may prove to be a viable noninvasive biomarker for cancer. Indeed, multiple studies have investigated the association of mtDNA content in PBLs

with the risk of different cancers. However, the relationship between mtDNA content and PCa is not clear. Different studies have found both an increase and decrease in mtDNA content associated with PCa and its progression to androgen-independence.[151,152] These studies did not test peripheral blood but instead examined PCa cells which can exhibit major genomic variation. Nevertheless, these studies established a basis for altered mtDNA content associated with PCa. A small preliminary study was performed using a validated qPCR assay to measure the concentration of mtDNA in sample extracts derived from two different patient groups: men with PCa and men with BPH. The purpose of this study was to evaluate group differences in mtDNA concentration between PCa and BPH and also between tissue and blood specimens within the same group.

Materials and Methods:

Samples

A total of 21 de-identified biological samples from 12 men, who underwent a radical prostatectomy, was obtained from the University of Texas Southwestern Medical Center with approval from the Office for the Protection of Human Subjects-Institutional Review Board (Project #2012-080; "Damage to Mitochondrial DNA in Age-Related Diseases: Alzheimer's and Prostate Cancer"; UNT Health Science Center, Fort Worth, TX). Eighteen of the samples were matched prostate tissue and blood samples from nine men diagnosed with PCa. The remaining three samples were prostate tissue from men with BPH and used as age-matched controls. The prostate tissue samples were assessed histologically for PCa or BPH; PCa samples were scored

and given a Gleason rating and tumor, node and metastasis (TNM) cancer staging rating (Table 4). DNA from blood and prostate tissue samples were extracted and purified using the DNeasy Blood and Tissue Kit (Qiagen, Valencia, CA). DNA from blood and prostate tissue samples were extracted and purified using the DNeasy Blood and Tissue Kit (Qiagen, Valencia, CA) following the manufacturers recommended protocol.

Positive and negative controls were assayed with the unknown samples. The positive control was human genomic DNA from cell line HL-60 at 0.24 pg/µL and the negative control was TE^{-4}.

BPH Patient #	Gleason	Stage	Age
1350	-	-	55
1365	-	-	60
1391	-	-	61
PCa Patient #			
1727	9	T3bN0Mx	61
1838	8	T2cN0Mx	74
2114	7	T3aNxMx	58
2160	7	T3NxMx	62
2164	9	T3bN0Mx	54
2167	9	T3aN0Mx	76
2178	7	T2cNxMx	60
2208	7	T2cNxMx	62
2238	7	T2cNxMx	59

Table 4: Clinicopathological information of prostatectomy patients. Gleason rating, TNM cancer staging and age for the 12 patients used in the preliminary study. The age of the individuals were 54 to 76 years old at the time samples were collected. For PCa patients the Gleason rating ranged from moderate to high (7 to 9) as well as the size of the tumor, *i.e.*, moderate to large (T2 and T3). None of the tumors exhibited local lymph node or metastatic invasion.

Real-Time qPCR and Data Analysis

Mitochondrial DNA was quantified using an internally validated real-time qPCR assay previously described by Kavlick *et al*. with minor modifications.[196,203] Quantification experiments were performed on a 7500 Real-Time PCR System and analyzed using Sequence Detection Software (SDS) version 1.2.3 (). Pearson correlation was performed on sample variables to test for linear associations and mtDNA concentration (copies per µL) data were analyzed with the Student's t-test; significance was indicated by a p-value <0.05.

Results

Absolute quantification of mtDNA was successful using the validated mtDNA qPCR assay. The standard curve for the assay run exhibited a slope of -3.6191 which corresponds to an amplification efficiency (E_x) of 88.93%. The standard curve also had a Y-intercept of 22.40 and a coefficient of determination (R^2) of 0.9997. Both controls passed and the exogenous IPC was detected for all 21 samples with an average Cq of 28.48 (SD±0.09) indicating no inhibition.

Twenty-one DNA extracts derived from 12 radical prostatectomy patients were tested: nine PCa patients (experimental group) and three BPH patients (control group) (Figure 10). For the PCa group matching blood and prostate tissue were collected totaling 18 samples (9 blood and 9 prostate tissue) whereas three prostate tissue samples were collected for the BPH group. The mtDNA concentration for the PCa blood and tissue samples averaged 777,076 copies per µL (SD ±472,414) and 2,081,944 copies per µL (SD±958,682), respectively. The three BPH tissue

samples averaged 2,484,167 copies per µL (SD±1,537,341). These estimates suggest the average mtDNA concentration in the PCa group is lower than the BPH group. However, a high level of dispersion and small sample size affected the calculated probability ($p = 0.64$) indicating no significant difference between groups (Figure 11). In contrast, the mtDNA concentration of PCa tissue and blood samples was determined to be significantly different ($p = 0.0047$). The tissue samples exhibited a mean mtDNA concentration (2,081,944 copies per µL) (SD±958,682) nearly three times greater than that of the blood samples (777,076 copies per µL) (SD±777,076). However, only a slight positive correlation ($r = 0.13$) was observed between them.

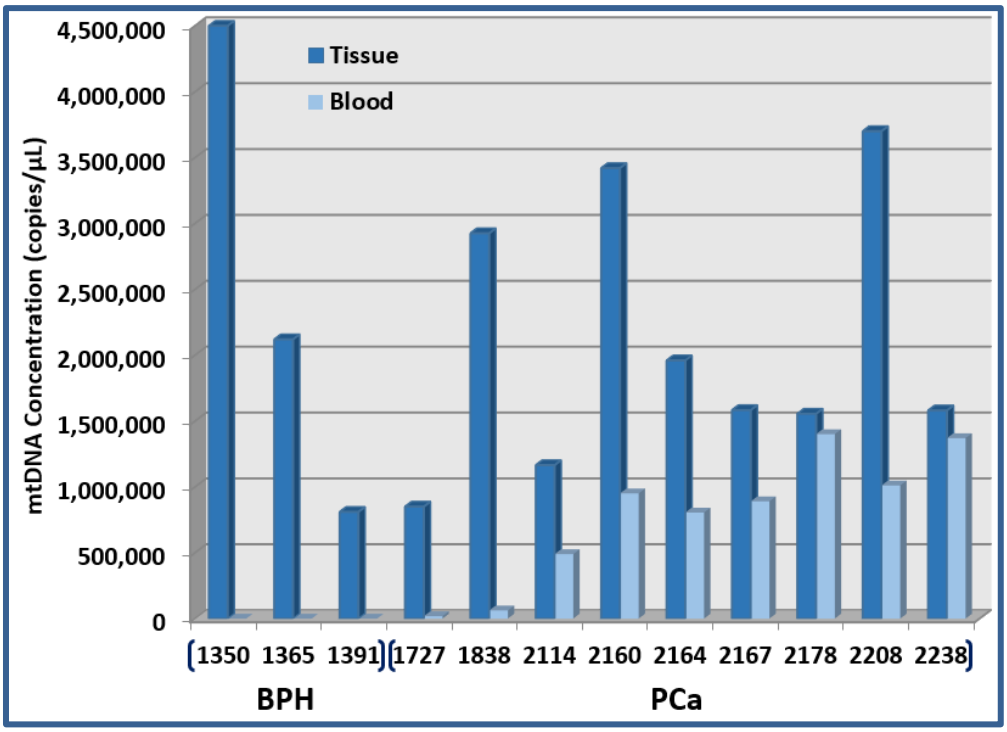


Figure 10: Mitochondrial DNA concentration of blood and prostate tissue extracts from 21 prostatectomy patients. The mtDNA concentration of twelve prostate tissue samples (dark blue bars) is highly variable within BPH and PCa: no significant difference (p=0.64) was observed between groups. The mtDNA concentration of nine (PCa only) blood samples (light blue bars) are also highly variable. However, a significant difference (p=0.0047) and a slight positive correlation (r=0.13) was observed between sample types, *i.e.*, tissue and blood.

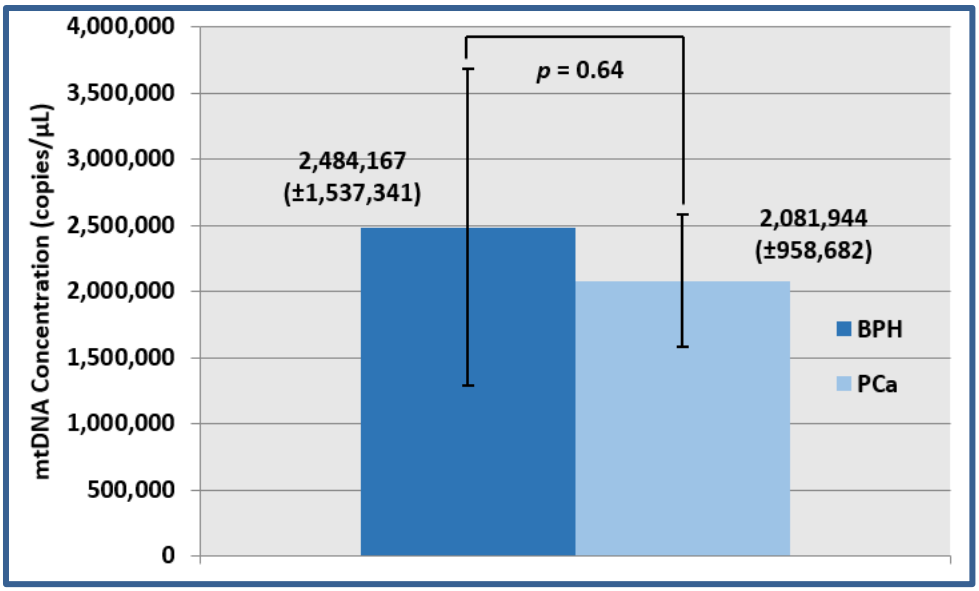


Figure 11: Mean concentration of mtDNA in PCa and BPH prostate tissue samples. Although the mtDNA concentration of tissue samples was higher in BPH than PCa, no significant difference (p=0.64) was observed between groups. PCa group (n=9) and BPH group (n=3)

Conclusion and Discussion

A decline in mtDNA copy number in peripheral blood occurs with age and has been associated with age-related diseases. Conversely, high mtDNA copy number in blood is associated with better health among aged individuals and may convey a beneficial effect to longevity.[211,212] No association (r = -0.047) was observed when mtDNA concentration of all twelve tissue (PCa and BPH) samples were tested with age. When tested separately, a marginal positive correlation (r = 0.19) was observed between the mtDNA concentration of PCa tissue and age, and, an inverse correlation (r = -0.28) between mtDNA concentration of PCa blood and

age was observed. Therefore it is uncertain whether an age-effect, if any, contributes to a change in mtDNA concentration.

As an age-related disease, it is expected that sporadic PCa progression, *i.e.*, higher TNM stage, increases with age. Surprisingly, the opposite was observed in this study (r = -0.15). The extent of PCa cases used in this study were limited to stage II and stage III. Further, the most progressed cases had a primary tumor grade (T3) but no regional or distant spread (N0 and MX). Therefore, a lack of samples representative of low and progressed PCa grades is likely the reason for the observed discrepancy between age and TNM stage. However, when Gleason scores and age were tested a positive correlation (r = 0.28) was observed and not surprisingly, Gleason scores and TNM staging demonstrated the strongest correlation (r = 0.59) among all the variables tested.

Overall, the mtDNA concentration data obtained from blood and prostate tissue samples of prostatectomy patients is inconsistent with the expected outcome based on a recent study by Zhou *et al*.[213] They found that the mtDNA copy number of peripheral blood is significantly greater ($p < 0.001$) in PCa individuals than age-matched healthy controls.[213] In contrast, data generated in this study show that PCa samples are marginally lower than the BPH group. An important distinction that may explain the discrepancy between studies is different measurement strategies for mtDNA were performed. The validated human qPCR assay developed by Kavlick *et al*. used in this study provides only a concentration estimate (copies per

µL). Considering that DNA from all 21 samples were eluted at the same volume using the DNeasy Blood and Tissue Kit, differences in the starting amount of prostate tissue or peripheral blood extracted would affect the final mtDNA concentration. The coefficient of variation (CV) for mtDNA concentration of the PCa tissue and blood samples were 46% and 61%, respectively. These high CV values may reflect sampling imprecision regarding the amount of tissue or blood used during DNA extraction. Therefore, normalizing mtDNA estimates to account for differences in the total number of nucleated cells that are extracted would correct for this effect. The quantification strategy used by Zhou *et al*. also used a real-time qPCR assay but it included two amplification targets: the mitochondrial-encoded NADH dehydrogenase 1 (*ND1*) gene and the human globulin (*HGB*) nuclear gene.[213] This approach effectively measures the mtDNA copy number per cell by calculating the ratio of mtDNA to nDNA concentrations. The benefit of using this type of strategy is that it normalizes the mtDNA content estimates and eliminates any sampling variability between specimens.

The two studies presented here detail the earliest experiments performed for this research project which focus on measuring mtDNA using qPCR. Data generated from the internal validation study demonstrate that qPCR can measure the concentration of mtDNA in different sample types with a high level of reproducibility, precision and sensitivity. However, when it was used to measure mtDNA in peripheral blood and prostate tissue samples, a severe shortcoming of the validated qPCR assay was revealed: it only provides a mtDNA concentration estimate. As such, the starting amount of tissue, *i.e.*, the number of nucleated cells, that is

extracted is not accounted for and may contribute to an inaccurate assessment of mtDNA differences between sample types and sample groups. Therefore, a qPCR assay that measures the amount of mtDNA per cell would be more informative and more appropriate for making estimate comparisons. The primary method for quantifying mtDNA copy number is qPCR. To achieve a per cell estimate, a nDNA target and a mtDNA target are quantified together. The two quantity estimates are then used to create a ratio which represents the number of mtDNA molecules per nDNA molecule, *i.e.*, the mtDNA copy number per cell. In the following chapter, a novel qPCR assay is developed and used to measure mtDNA copy number per cell as well as another type of mtDNA alteration that has been associated with PCa progression.

Chapter II Results at a Glance

1) A validated mtDNA qPCR assay was implemented in a preliminary study to evaluate the concentration of mtDNA in blood and prostate tissue samples from prostatectomy patients.
2) The concentration of mtDNA was marginally lower in PCa than BPH samples but small sample size and variance prohibited significance.
3) A multiplex qPCR assay that estimates mtDNA copy number per cell is more advantageous than the assay used in this study as it would normalize mtDNA content and decrease sampling variability.

Chapter III

QUANTIFICATION OF MITOCHONDRIAL DNA: DNA COPY NUMBER AND DELETION RATIO

Mitochondrial dysfunction is implicated in a vast array of age-related pathologies such as neurodegenerative diseases and cancer. Assessing changes to mtDNA may provide insight into the processes that contribute to mitochondrial dysfunction and its role in disease development and progression. The ratio of mtDNA to nDNA is often used as an estimate for the number of mtDNAs per cell, or mtDNA copy number. This is a high-level indicator of mitochondrial biogenesis as the mtGenome encodes critical enzymatic subunits of OXPHOS. Perturbation of mtDNA content can result in a mitochondrial bioenergetic deficiency and have pathophysiological consequences. Indeed, reduced mtDNA content has been shown to decrease OXPHOS activity and has been correlated with tumor progression.[214,215] Studies involving PCa cell lines have shown depletion of mtDNA is linked to becoming androgen-independent, progression to an invasive phenotype that is resistant to conventional chemotherapies, and induction of epithelial-mesenchymal transition leading to cancer metastasis.[152,216,217]

Large scale deletions in the mtGenome accumulate in aged, post-mitotic tissues and can cause mitochondrial dysfunction and disease.[153–157,159] Specifically, large-scale deletions found in the mtDNA major arc, which cause mitochondrial disease, have been associated with various

cancers.[218–222] In PCa, the number of large mtDNA deletions in malignant prostate tissue increases with advanced age.[223] Further, one group claims that a specific 3.4 kb mtDNA deletion is associated with PCa and may have clinical value as a disease biomarker.[224,225]

A unique multiplex qPCR assay was developed and validated for the purpose of assessing mtDNA changes associated with PCa progression. The assay design provides simultaneous assessments of mtDNA copy number per cell ($mtDNA_{CN}$) and the proportion of mtDNAs with common large deletions ($mtDNA_{DR}$) by targeting two mitochondrial sites and one nuclear locus (Figure 12). Following development and validation of the multiplex qPCR assay, 134 normal control and PCa patient samples from a nested case-control cohort were tested for differences in $mtDNA_{CN}$ and $mtDNA_{DR}$. In addition, $mtDNA_{CN}$ and $mtDNA_{DR}$ are tested for correlation with PCa progression.

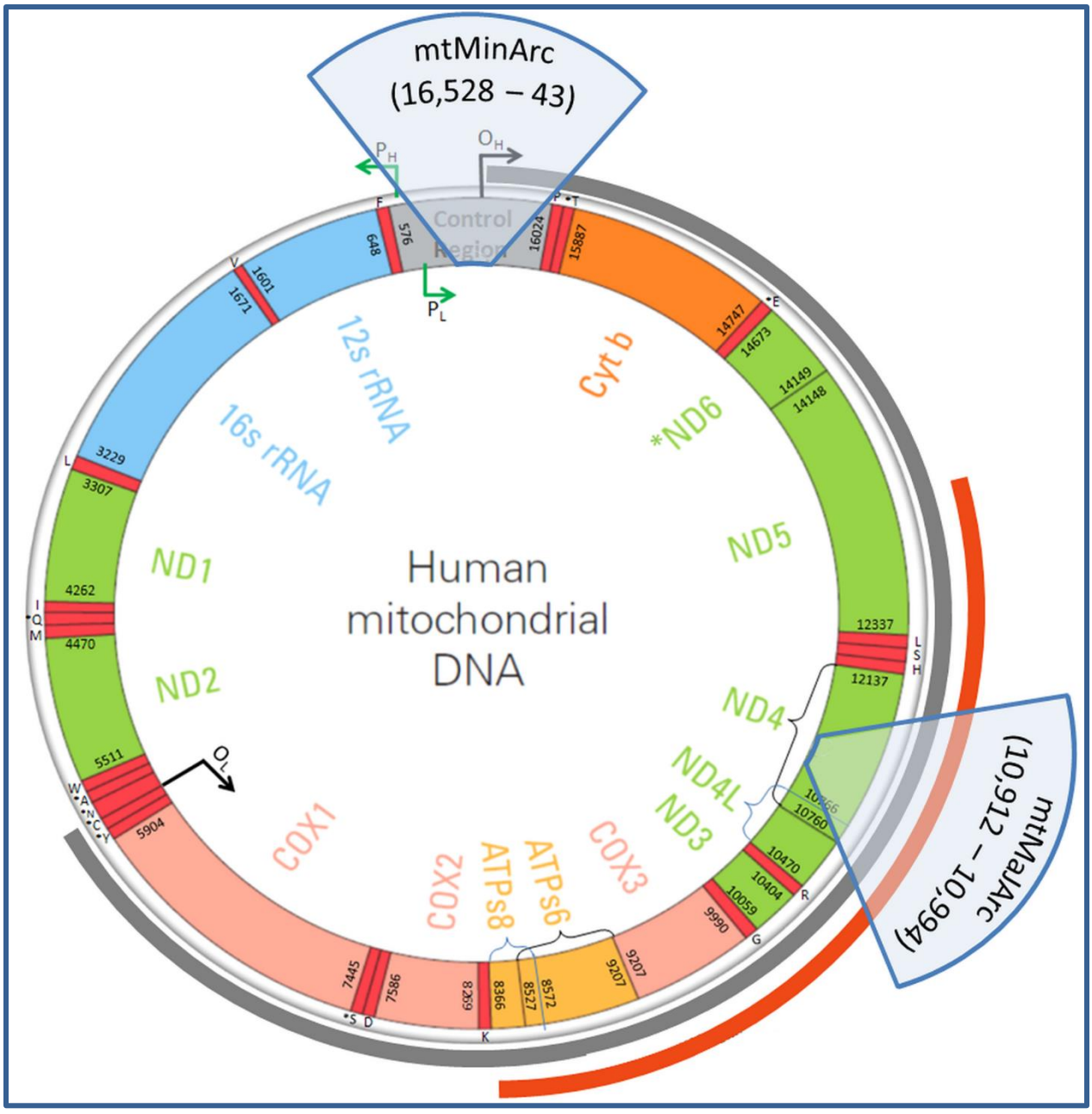


Figure 12: Mitochondrial DNA amplification targets for qPCR. The mtMajArc target is located within the major arc of the mtGenome (gray arc between the origins of replication, heavy strand (O_H) and light strand (O_L)) and spans 84% of reported large deletions, including the 4,977 bp "common" deletion (red arc between positions 8469-13447). The mtMinArc target is located in the control region of the mtGenome where no common deletions of more than two bases have been reported.

Materials and Methods

Biological specimens used for this study were obtained with approval from the Office for the Protection of Human Subjects-Institutional Review Board (Project #2012-080; "Damage to Mitochondrial DNA in Age-Related Diseases: Alzheimer's and Prostate Cancer"; UNT Health Science Center, Fort Worth, TX).

Samples

Samples used for this study were obtained from San Antonio's Center of Biomarkers of Risk for Prostate Cancer (SABOR), a National Cancer Institute and Early Detection Research Network sponsored Clinical Validation Center which is comprised of more than 3,700 male participants from the San Antonio area. Participation in SABOR involves an annual screening that consists of a serum PSA test and digital rectum exam (DRE). Individuals with a serum PSA level above 2.5 ng/mL, a suspect DRE and/or family history of PCa are referred for a 12-core ultrasound guided biopsy.

One hundred and thirty four whole blood DNA extracts were provided for this study by SABOR. Nested case-control samples from the SABOR cohort are divided among three groups based on disease status and are mean age-matched as well as proportionally race and ethnicity matched; normal control (NC) group, indolent PCa and aggressive PCa. The NC group consists of 46 samples from study participants who have at least five years follow up with no PCa detection (based on screening negative, not necessarily biopsy confirmed negative). The indolent PCa

group, no evidence of disease (NED), comprises 46 confirmed PCa cases that tested negative for biochemical recurrence within five to ten year follow-up after PCa treatment. Alternatively, the aggressive PCa groups, biochemical recurrence (BCR) and metastasis (MET), comprise 42 confirmed PCa cases that tested positive for biochemical recurrence or distant metastasis within five to ten year follow-up after PCa treatment. Phenotypic measures associated with each PCa case includes individual Gleason grades, Gleason score, serum PSA, PCa tumor, node and metastasis (TNM) staging as well as associated demographic information, *e.g.*, race, ethnicity and age (Table 5).

	NC N = 46	NED N = 46	BCR N = 42	*p* -value
Age (At Blood Collection)				
Mean Years ±SD	57 (7.2)	67 (8.5)	68 (7.6)	<0.05
Min-Max	43 - 70	49 - 85	55 - 84	
Race/Ethnicity				
White	23	23	21	>0.05
Hispanic	23	23	21	
Gleason Score				
Left				
≤ 6		23	5	
7 (3 +4)		7	7	
7 (4 + 3)		0	3	
≥ 8	NA	1	15	
Right				
≤ 6		22	6	
7 (3 +4)		5	9	
7 (4 + 3)		0	4	
≥ 8		3	15	
Cancer Stage				
Localized	NA	39	23	
Regional/Distant Spread		1	8	
Initial PCa Treatment				
Prostatectomy	NA	34	25	
Other		12	17	

Table 5: SABOR cohort used in study. The mean age in the Normal control (NC) is significantly lower (p<0.05) than the mean ages of the no evidence of disease (NED) and biochemical recurrence (BCR) groups. No significant difference (p>0.05) was observed in the number of White or Hispanic subjects in all three groups.

Real-Time qPCR

Briefly, three amplification targets were chosen for the multiplex: a mtDNA target in the minor arc, a region where large deletions are rare (mtMinArc); a mtDNA target in the major arc, a region where age- and disease- related deletions are common (mtMajArc); and a single copy target in the nuclear gene, beta-2 microglobulin (β2M) (Figure 12). The $mtDNA_{CN}$ per cell is represented by the ratio of the mtMinArc target to the nuclear target. The mtDNA deletion ratio, or $mtDNA_{DR}$, is represented by (mtMinArc-mtMajArc)/mtMinArc. The mitochondrial targets are novel and were carefully selected with regard to the location of reported deletions within the mtGenome. The mtMinArc target spans the heavy strand origin of replication and is not affected by any of the reported large deletions. The mtMajArc target is located within the "common" deletion as well as within 84% of the reported large deletions in the major arc (Figure 12). Primers and probes were designed using Primer3 (Whitehead Institute for Biomedical Research, Cambridge, MA), Primer-BLAST/BLAST (National Center for Biotechnology Information, Bethesda, MD) and Oligo Analyzer 3.1 (Integrated DNA Technologies Inc., Coralville, IA). Using these bioinformatic tools, the primers and probes were tested for compatibility as well as non-specific amplification. Additionally, the mtMinArc primers were screened for common population polymorphisms using the European DNA Profiling (EDNAP) Group's mtDNA Population Database (EMPOP). EMPOP is a large database of partial or whole mtGenome haplotypes collected from all over the world.[226] The β2M primers and probe sequences have been previously described.[227]

Real-time qPCR of the three targets was performed using a probe-based multiplex assay on the 7500 Real-Time PCR System using the following thermal profile: 95°C for 10min; 40 cycles of 95°C for 15s, 55°C for 15s; and 60°C for 1min. The reaction components are as follows (given in final concentration): 22.5µL TaqMan® Universal PCR Master Mix, No AmpErase® UNG (Applied Biosystems), each mitochondrial amplification primer at 50nM (FmtMinArc, RmtMinArc, FmtMajArc, RmtMajArc), each nuclear DNA amplification primer at 1250nM (Fβ2M, Rβ2M), and each dual hybridization probe at 250nM (PmtMinArc, PmtMajArc, Pβ2M) (TaqMan® MGB Probe, Applied Biosystems). The final reaction volume is 25µL.

Absolute quantification was obtained using a standard curve generated from a reference DNA sample from a healthy, 30-year-old male. Whole blood was extracted using the QIAmp® DNA Blood Mini Kit (Qiagen, Inc.) and the respective DNA concentrations were then measured independently as follows: mtDNA was quantified using a validated mtDNA qPCR assay (from Chapter II) and nuclear DNA was quantified using Quantifiler® Duo DNA Quantification Kit (Applied Biosystems).[196,203] The sample was concentrated to generate the first standard, Std1, with a target concentration of approximately 8000 cells/µL and 2.4 million mtDNA copies/µL (exact quantification values used for absolute quantification). A two-fold dilution series of Std1 was used to create the subsequent standards (Std2 through Std8). Standard curves were analyzed at a quantification cycle (Cq) of 0.04 using the automatic baselining algorithm. The R^2, Y-intercept and amplification efficiency values were analyzed for all standard curves prior to absolute quantification of unknown samples.

Statistical Analysis

Statistical analyses were performed using IBM© SPSS™ Statistics 21 (IBM Corporation, Armonk, NY) and Microsoft Excel® (Excel) Data Analysis Tools (Microsoft Corporation, Redmond, WA). Data sets were tested for normality (Shapiro-Wilk test), equality of variances (Levene's test) and comparison of means. Group differences were tested using ANOVA or Kruskal-Wallis when the assumption of equal variances was not met. Pearson correlation was performed on sample variables to test for linear associations where linear regression analysis was used to determine predictor status. Tests were two-tailed and significance was assessed at an alpha level of 0.05.

Results

Appropriate age and race/ethnicity matching between groups was tested using analysis of variance (ANOVA) and the Chi-square statistic, respectively. Age matching between groups was not equal ($p < 0.05$); Tukey's HSD post-hoc analysis revealed significant differences of mean age between the NC and NED ($p < 0.05$) as well as NC and BCR/MET ($p < 0.05$) but no significant difference between NED and BCR/MET ($p = 0.733$). As an age-related disease, a man's risk of developing PCa increases as he gets older. Indeed, 97% of all PCa diagnoses occur in men 50 years of age and older.[1,228] Not surprisingly then, the mean age of the NC group (57.1 ± 7.23) was approximately ten years less than the NED (66.7 ± 8.52) and BCR/MET (68.0 ± 7.64) groups. Age may have a confounding effect on observed differences in the $mtDNA_{CN}$ and $mtDNA_{DR}$ measures. Pearson tests showed a negative correlation between age and $mtDNA_{CN}$ (r = -0.250,

p = 0.005) and between age and $mtDNA_{DR}$ (r = -0.215, p = 0.014). Further, linear regression analysis using age as the predictor variable for $mtDNA_{CN}$ or $mtDNA_{DR}$ revealed significance in the models but both lack appreciable precision due to high variability (Figures 13 and 14).

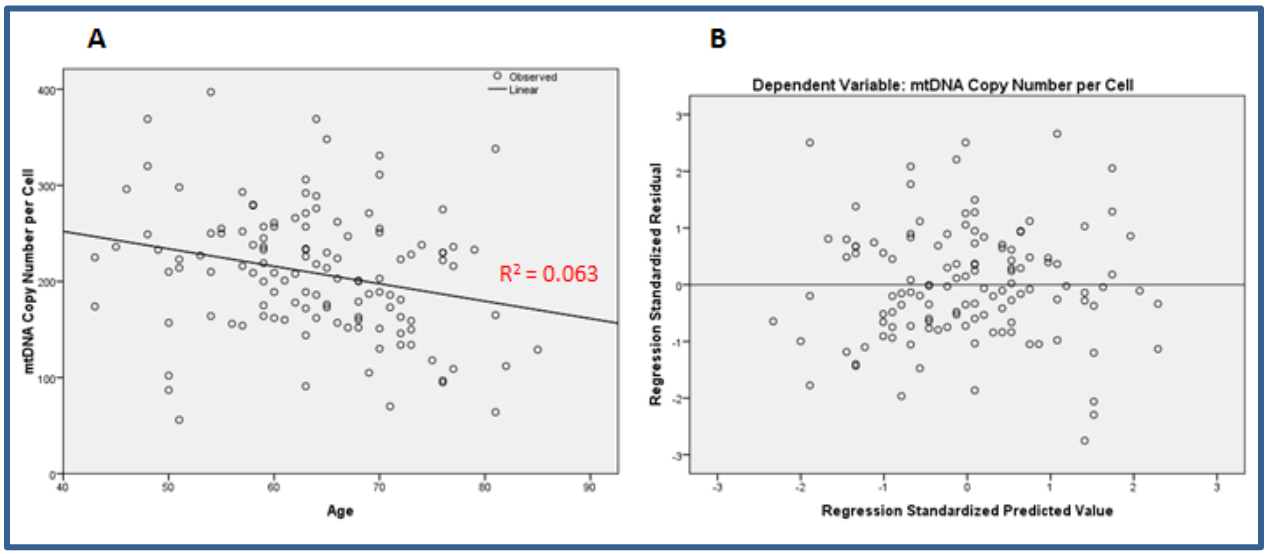


Figure 13: Linear regression for age and $mtDNA_{CN}$. A) The linear regression line shows that as age increases, copy number decreases. Although significant (p=0.005), the R^2 value is low indicating high variability and low precision in the model. B) Residual plot shows the residuals are normally distributed and centered on zero thus the linear regression is an appropriate model.

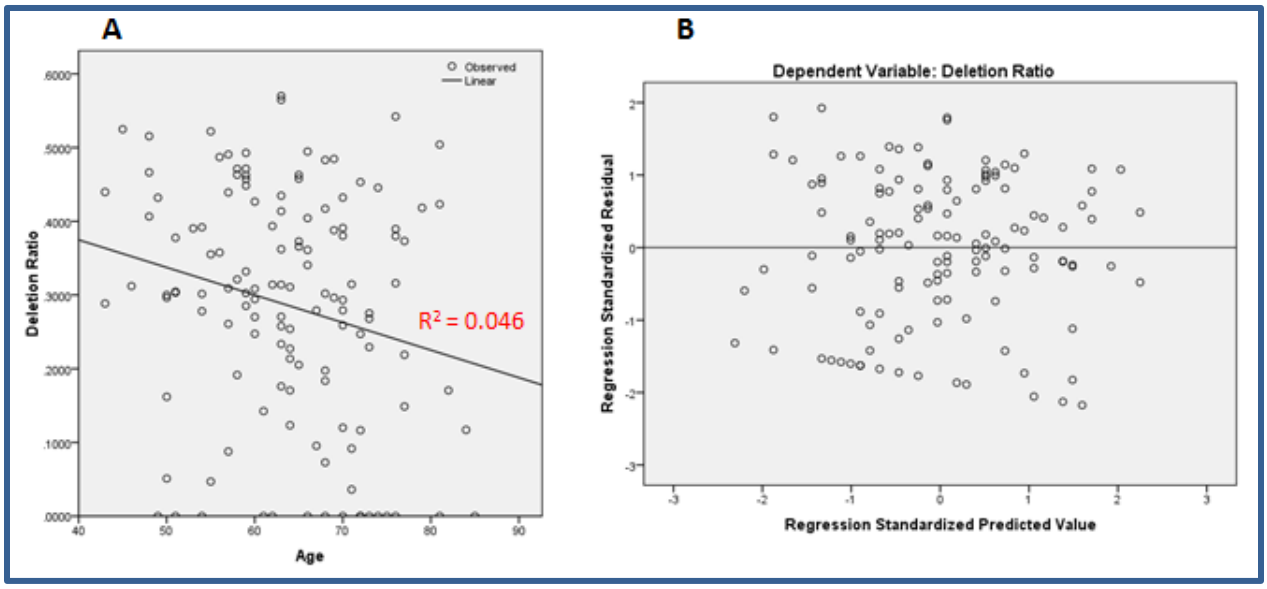


Figure 14: Linear regression for age and deletion ratio. A) The linear regression line shows that as age increased, deletion ratio decreased. This means the proportion of mtDNAs harboring large-scale deletions decreased, *i.e.*, improved, with age. As was seen in the previous figure (13-A), age as a predictor variable is significant (p = 0.014) but the R^2 is low indicating high variability and low precision in the model. B) Residual plot shows the residuals are normally distributed and centered on zero thus the linear regression is an appropriate model.

A total of six qPCR experiments were performed; R^2, slope, amplification efficiency, and Y-intercept were within the expected range (data not shown). All 134 samples were assayed in duplicate. Absolute quantification estimates for all three targets (mtMinArc, mtMajArc, β2M) were tabulated for each sample and used to calculate $mtDNA_{CN}$ and $mtDNA_{DR}$. Duplicates for each sample were averaged, to give a single $mtDNA_{CN}$ and $mtDNA_{DR}$ value. Data sets were then subjected to outlier analysis, stratified by disease status, using Tukey's method. A total of nine outliers were identified, NC=2, NED=4, and BCR/MET=3, in the $mtDNA_{CN}$ data set and three, NC=2, NED=0 and BCR/MET=1, in the $mtDNA_{DR}$ data set. After outliers were removed,

descriptive statistics were generated for $mtDNA_{CN}$ and $mtDNA_{DR}$ (Table 6). The mean $mtDNA_{CN}$ was marginally greater in the BCR/MET (214.23±70.20) group as compared to the NC (208.73±63.08) and NED (201.64±66.35) groups but no significant difference (p=0.564) was observed. Further, analysis of $mtDNA_{CN}$ variance showed homogeneity among all three disease groups.

		N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
						Lower Bound	Upper Bound		
mtDNA Deletio Ratio	Normal	44	.3082	.1454	.0219	.2640	.3524	.0000	.6094
	NED	46	.2715	.1498	.0221	.2271	.3160	.0000	.5647
	BCR and MET	41	.2866	.1910	.0298	.2263	.3469	.0000	.5701
	Total	131	.2886	.1619	.0141	.2606	.3165	.0000	.6094
mtDNA Copy Number per Cell	Normal	44	208.73	63.08	9.51	189.55	227.90	56	369
	NED	42	201.64	66.35	10.24	180.97	222.32	70	397
	BCR and MET	39	214.23	70.20	11.24	191.47	236.99	64	369
	Total	125	208.06	66.12	5.91	196.36	219.77	56	397

Table 6. mtDNA qPCR assay descriptive statistics. The $mtDNA_{DR}$ (mean scores) decreased with PCa progression, *i.e.*, the number of whole mtDNAs to truncated DNAs was higher in the NC group than the NED group and even more than the BCR and MET groups. Also, the level of variance, given by standard deviation and 95% CI, increased with PCa progression suggesting a higher level of copy number control dysfunction associated with disease status. Although the variance in $mtDNA_{CN}$ exhibited a similar trend (increased with disease progression) as with $mtDNA_{DR}$ variance, there was no clear tendency with respect to the mean values. Instead, the BCR and MET group had the highest $mtDNA_{CN}$, the NED group the lowest, and the NC group in between.

The NC group had the highest mean $mtDNA_{DR}$ (0.3082±0.1454) of the three groups. The NED group exhibited the lowest mean $mtDNA_{DR}$ (0.2716±0.1498) but the BCR/MET group's

mean $mtDNA_{DR}$ (0.2866±0.2004) was nearly the same (Figure 16). Again, no significant difference was observed between means of the three groups (p=0.564) However, a test for homogeneity of variances was significant (p=0.06). Performing post-hoc pairwise F-tests revealed the difference to be between NC and BCR/MET (p=0.041). A Kruskal-Wallis test was performed since the assumption of equal variances was not met but the adjusted p-value was still not significant (p=0.498).

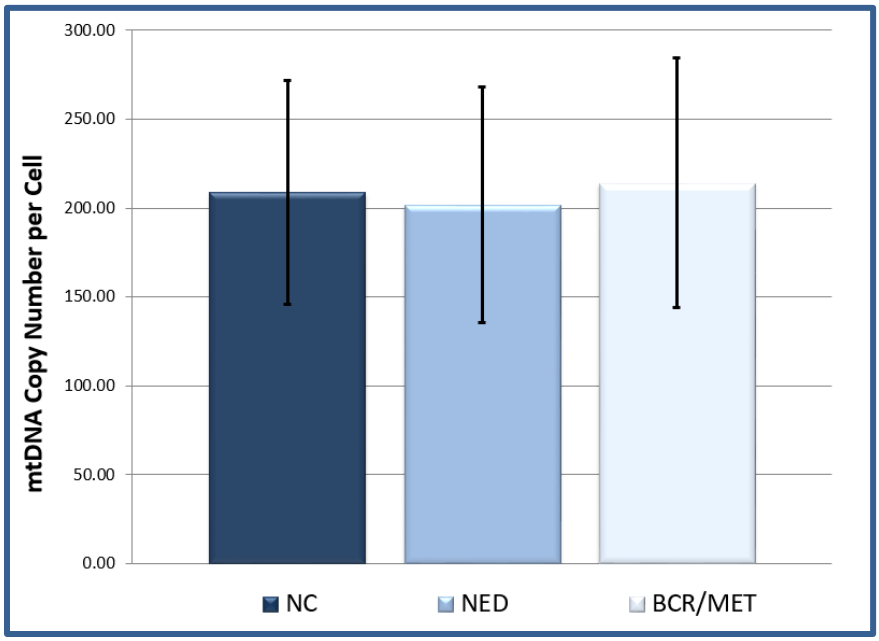


Figure 15: Mean mtDNA copy number per cell by disease status. No significant difference (p=0.694) was observed between the normal control (NC), no evidence of disease (NED) and biochemical recurrence/metastasis (BCR/MET) groups. Further, no difference in variances between all three groups was detected.

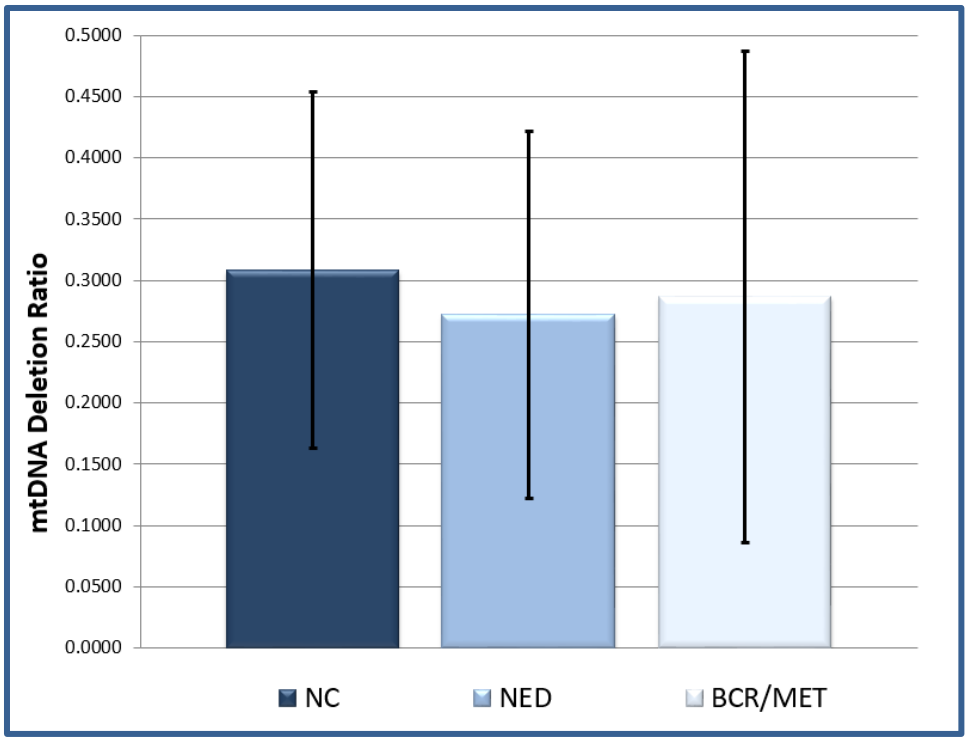


Figure 16: Mean mtDNA deletion ratio by disease status. No significant difference (p=0.498) was observed between the normal control (NC), no evidence of disease (NED) and biochemical recurrence/metastasis (BCR/MET) groups. However, pairwise F-test for two-sample variances indicates a significant difference (p=0.041) between the NC and BCR/MET groups but not NC and NED or NED and BCR/MET.

Additionally, correlation analysis was performed with clinicopathological characteristics of the PCa samples to determine if $mtDNA_{CN}$ or $mtDNA_{DR}$ associate with PCa progression. Results showed a positive correlation (r=0.286, p=0.025) between Gleason score and $mtDNA_{CN}$ which supports findings from a recent study.[213] As expected, a very strong correlation (r=0.457, $p<0.01$) was observed between Gleason score and tumor stage. However, $mtDNA_{CN}$ only had a weak correlation (r=0.178, p=0.149) with tumor stage despite its significant correlation with

Gleason score. Additionally, no significant correlations were observed between $mtDNA_{DR}$ and any of the clinicopathological variables tested.

Conclusion and Discussion

Prostate cancer is an age-related disease and it is also well-established that changes to mtDNA accumulate with age. Therefore, age is expected to be a strong confounder in this study. While matching is intended to eliminate confounding effects, it is also beneficial as it improves the efficiency in estimating the effect an independent variable has on the measurable outcomes of interest. Given $mtDNA_{CN}$ is an indicator of mitochondrial biogenesis, it follows then that a measurable difference is expected between the NC and PCa groups if mitochondrial dysfunction is associated with PCa pathology. The same is true for $mtDNA_{DR}$. However, without proper age matching among the normal control group and PCa groups, bias due to the confounding age effect decreases the ability to accurately assess if PCa has an effect on $mtDNA_{CN}$ and $mtDNA_{DR}$.

Age is the strongest risk factor for developing PCa. Therefore, frequency or mean matching for age within a research cohort is normally performed in PCa studies. For this project, age matching men from the SABOR cohort proved difficult and an age difference of approximately 10 years was observed between the NC group and the PCa groups. Furthermore, age was shown to have significant negative correlations and to be a significant predictor for $mtDNA_{CN}$ and $mtDNA_{DR}$ based on linear regression modeling. It was expected then that a

measureable difference in $mtDNA_{CN}$ and $mtDNA_{DR}$ would be observed between the NC group and the NED and BCR/MET groups, if only because of the age effect. However, no significant difference was observed in either dependent variable between the three groups or when grouped as NC or PCa ($mtDNA_{CN,}$ p=0.935 and $mtDNA_{DR}$ p=0.326).

Another potential confounding factor in this study is the amount of time between age of PCa diagnosis and age of the SABOR participant when the blood sample was collected. On average the age difference was 3.48 (SD±3.05) and 6.17 (SD±6.22) years for NED and BCR/MET, respectively. Per the stated hypothesis, PCa etiology involves changes to mitochondria, primarily through a ROS mediated axis, that accumulate with disease progression and therefore can delineate indolent from aggressive cases. Given that mitochondria are highly dynamic organelles which change in response to the cellular environment, *e.g.*, energetic demands; in absence of disease, mitochondria will adjust towards a homeostatic state. Indeed, this has been demonstrated in prostate tumor cell line experiments.[152] It is reasonable then that the observed $mtDNA_{CN}$ and $mtDNA_{DR}$ may not be truly representative of the disease state, *i.e.*, indolent (NED) or aggressive (BCR/MET) PCa. This may be especially true in instances where several years have transpired between the date of PCa diagnosis and the date when the blood sample was collected from the SABOR participant. The longest amount of time between PCa diagnosis and sample collection for the NED and BCR/MET group was 12 and 21 years, respectively. Linear regression revealed that indeed, the amount of time (years) between PCa diagnosis and sample collection was a significant predictor for $mtDNA_{CN}$ (p=0.027) (Figure 17).

However, low precision of the linear model indicates time difference does not explain all the observed variability in the measured outcomes.

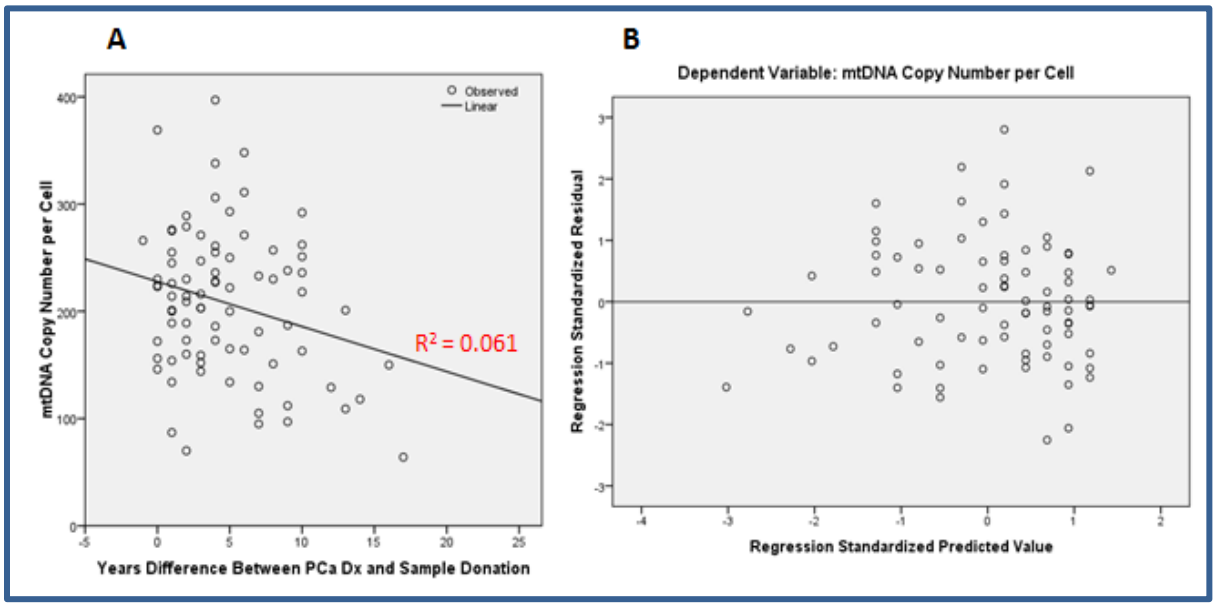


Figure 17: Linear regression for time difference and $mtDNA_{CN}$. A) The linear regression line shows that as time difference between PCa diagnosis and sample collection increases, $mtDNA_{CN}$ decreases. Time difference as a predictor variable is significant (p=0.027) but the R^2 is low indicating high variability and low precision in the model. B) Residual plot shows the residuals are normally distributed and centered on zero thus the linear regression is an appropriate model.

Results of the qPCR assay indicate that $mtDNA_{CN}$ of peripheral blood from the BCR/MET group is marginally greater than the NC group. This is consistent with previous findings where PCa tumor cells and peripheral blood from PCa patients exhibit a higher $mtDNA_{CN}$ than normal prostate tissue cells and men without PCa.[151,213] The study by Zhou *et al.*, was conceptually very similar to the study performed here.[213] However, in their study significance was observed at

$p<0.001$ for a difference in $mtDNA_{CN}$ between PCa and controls. Not only was their cohort larger (196 age-matched subjects in each group) but 84% of their PCa cases were newly diagnosed and none of the PCa cases had received any previous PCa-related treatment.

Differences between this study and that performed by Zhou *et al*., highlight the probable factors affecting the observed variable distribution. Figures 15 and 16 clearly show that variance in the $mtDNA_{CN}$ and $mtDNA_{DR}$ measures were large in all three groups and, consequently, contributed to the low statistical power. It is likely the large variances observed are due to the small sample size of each group. Using the group means and standard deviation of $mtDNA_{CN}$ and $mtDNA_{DR}$, a theoretical sample size was calculated to achieve a five percent confidence level. According to the power calculation, a sample size of approximately 440, in each group, would be needed to achieve significance at the p=0.05 level. Therefore, a larger sample size would provide a better resolution for discerning significant differences in $mtDNA_{CN}$ and $mtDNA_{DR}$ between the study groups.

In addition, the normal control samples selected from the SABOR cohort are nested case-controls. In this type of experimental design, a subject chosen to be a control for a case at a given age may not necessarily be excluded from the set of controls because of subsequent development of disease. Thus, a normal control that subsequently develops disease can also serve as a case. Enrollment into the SABOR program is not exclusive to new cases. Indeed, the SABOR cohort used in this study includes individuals who: 1) screened negative for PCa at the

time of enrollment, 2) developed PCa during the course of SABOR enrollment, and 3) were previously diagnosed with PCa before enrollment. Together, the nested case-control design and inclusion of prospective and retrospective PCa cases may introduce an element of selection bias in the sampling distribution. Although its effect is likely very minimal, selection bias in conjunction with the other previously discussed confounding factors affected the stability of the measured $mtDNA_{CN}$ and $mtDNA_{DR}$ outcomes.

Chapter III Results at a Glance

1) Marginal differences in $mtDNA_{CN}$ and $mtDNA_{DR}$ were observed between the NC, NED and BCR/MET groups.
    A. The BCR/MET group had the highest $mtDNA_{CN}$ which may be a compensatory response to disease pathology.
    B. The PCa groups had a lower $mtDNA_{DR}$ which indicates a higher proportion of large deletions in the mtGenome as compared to the NC group.
2) A significant correlation was observed between increasing Gleason score and increasing $mtDNA_{CN}$ which indicates a possible role in disease progression.
3) Larger sample size, properly age-matched controls and samples collected at a time point more representative of health status are recommended for future studies.

Chapter IV

PATHOLOGICAL MITOCHONDRIAL DNA VARIANTS IN PROSTATE CANCER SAMPLES

Prostate cancer results from continued, unregulated proliferation of prostate cells stemming from an accumulation of abnormalities in multiple cell regulatory systems. Mitochondria have been implicated in the process of carcinogenesis because of their vital role in energy production, cell signaling, proliferation and, conversely, apoptosis.[131,132,134,136] Indeed, mtDNA somatic and germline mutations have been detected in mtDNA of many different cancers.[183,229–233] Specifically, somatic mtDNA mutations are frequent in PCa and may occur early during prostate malignant transformation.[188,189] Damaged mtDNA molecules result in defective proteins which can alter mitochondrial metabolism, enhance tumorigenesis and permit cancer cell adaptation to changing environments.[184–186]

Mutations in the cytochrome c oxidase subunit I (COI) gene are significantly more frequent in prostate tissue biopsies from PCa patients than no cancer controls.[190] In addition, PCa cells harboring a mutation in the ATP6 gene have increased levels of reactive oxygen species (ROS), an increased rate of tumor growth and generate tumors seven times larger than wild-type controls.[190] These studies support the posit that specific mtDNA mutations manifest into deficient ETC complex subunits, cause mitochondrial dysfunction and result in increased ROS production. Increased ROS has been implicated as a factor in tumor formation through a

signaling transduction pathway that inactivates p16Ink4A and p53.[191] These findings are supported by several independent studies that have shown that mitochondrial oxidative stress drives tumor progression and metastasis.[192] Furthermore, the autophagic tumor stroma model of cancer etabolism states that increased levels of ROS drive mitophagy resulting in substrate level phosphorylation, *i.e.* aerobic glycolysis, in fibroblasts. The byproducts of aerobic glycolysis, such as lactate, pyruvate, ketones and glutamine, fuel adjacent cancer cells thereby promoting tumor growth and protection from apoptosis.[193]

Despite the growing body of evidence that supports the association of genetic alterations in mtDNA and cancer, its influence on the development and progression of PCa are not clearly understood. Studies have reported somatic changes in mtDNA of PCa patients in both prostate tumor tissue and bodily fluids but have failed to demonstrate a causal role related to neoplastic transformation. Here, mtDNA from extracted peripheral blood samples of normal controls and PCa patients were sequenced using the traditional Sanger method and a massively parallel sequencing platform. Mutational load was assessed for each group using a phylogenetic approach in order to distinguish ancestral gene variants from those that may be associated with disease. Finally, the predicted structural and functional consequence of amino acid substitutions were estimated for each missense mutation.

Study 1: Mitochondrial DNA variants of PCa and BPH Blood and Prostate Tissue Samples

Prior to sequencing the entire mtGenome on a large study cohort, a small preliminary study was performed to assess the success of the research strategy in which only two functional locations of the mtGenome were sequenced: (1) the hypervariable control region; and (2) the NADH dehydrogenase subunit 5 (ND5) gene. The hypervariable segments HV1 and HV2 were sequenced for haplogroup designation and the ND5 gene was sequenced because of its critical importance to the function of NADH dehydrogenase (Complex I) and its involvement in ROS production.

Materials and Methods:

Samples

Biological specimens described in the second study of Chapter II were also assayed in this preliminary sequencing study (Table 4 and Table 10). Briefly, 21 blood and prostate tissue samples from 12 men, who underwent a radical prostatectomy, were obtained from the University of Texas Southwestern Medical Center. Eighteen of the samples were matched prostate tissue and blood samples from nine men diagnosed with PCa. The remaining three samples were prostate tissue from men with BPH and used as age-matched controls. The samples were obtained and used with approval from the Office for the Protection of Human Subjects-Institutional Review Board (Project #2012-080; "Damage to Mitochondrial DNA in Age-Related Diseases: Alzheimer's and Prostate Cancer"; UNT Health Science Center, Fort Worth, TX).

Mitochondrial DNA Sequencing

The Sanger method was used to sequence HV1, HV2 and the ND5 gene using five primer pairs (Table 7 and Figure 18). Amplification reactions were performed on a 96-Well GeneAmp® PCR System 9700 (Applied Biosystems) and electrophoresed on a 3130*xl* Genetic Analyzer (Applied Biosystems). The sequencing strategy for ND5 involved three overlapping fragments using primer pairs $ND5_A$, $ND5_B$ and $ND5_C$, respectively. $ND5_A$ was designed using Primer3 (Whitehead Institute for Biomedical Research, Cambridge, MA), Primer-BLAST/BLAST (National Center for Biotechnology Information, Bethesda, MD), and Oligo Analyzer 3.1 (Integrated DNA Technologies Inc., Coralville, IA).[234] Using these bioinformatic tools, primers were tested for compatibility as well as specificity to minimize unintended co-amplification of mitochondrial pseudogenes present in the nuclear genome, *i.e.*, nuclear-encoded mitochondrial pseudogenes (NUMTS). Primer pairs $ND5_B$ and $ND5_C$ were selected from Kumar *et al*.[235]

	Primer	Binding Site Position		Sequence (5' - 3')
Forward	R1	15,910	- 15,931	CACCAGTCTTGTAAACCGGAGA
	C1	29	- 48	CTCACGGGAGCTCTCCATGC
	$ND5_AF$	11,900	- 11,922	GTGCTAGTAACCACGTTCTCCTG
	$ND5_BF$	12,551	- 12,571	AAACAACCCAGCTCTCCCTAA
	$ND5_CF$	13,319	- 13,338	ACATCTGTACCCACGCCTTC
Reverse	R2	564	- 545	CTTTGGGGTTTGGTTGGTTC
	B1	16,410	- 16,391	GAGGATGGTGGTCAAGGGAC
	$ND5_AR$	12,808	- 12,786	ATCAACTGATGAGCAAGAAGGAT
	$ND5_BR$	13,526	- 13,507	TCGATGATGTGGTCTTTGGA
	$ND5_CR$	12,808	- 12,786	AGAGGGGTCAGGGTTGATTC

Table 7: Sanger sequencing primers. Five primer pairs were used to sequence HV1, HV2 and ND5 of the mtGenome. Binding site position refers to the start and stop base positions on the rCRS**.**

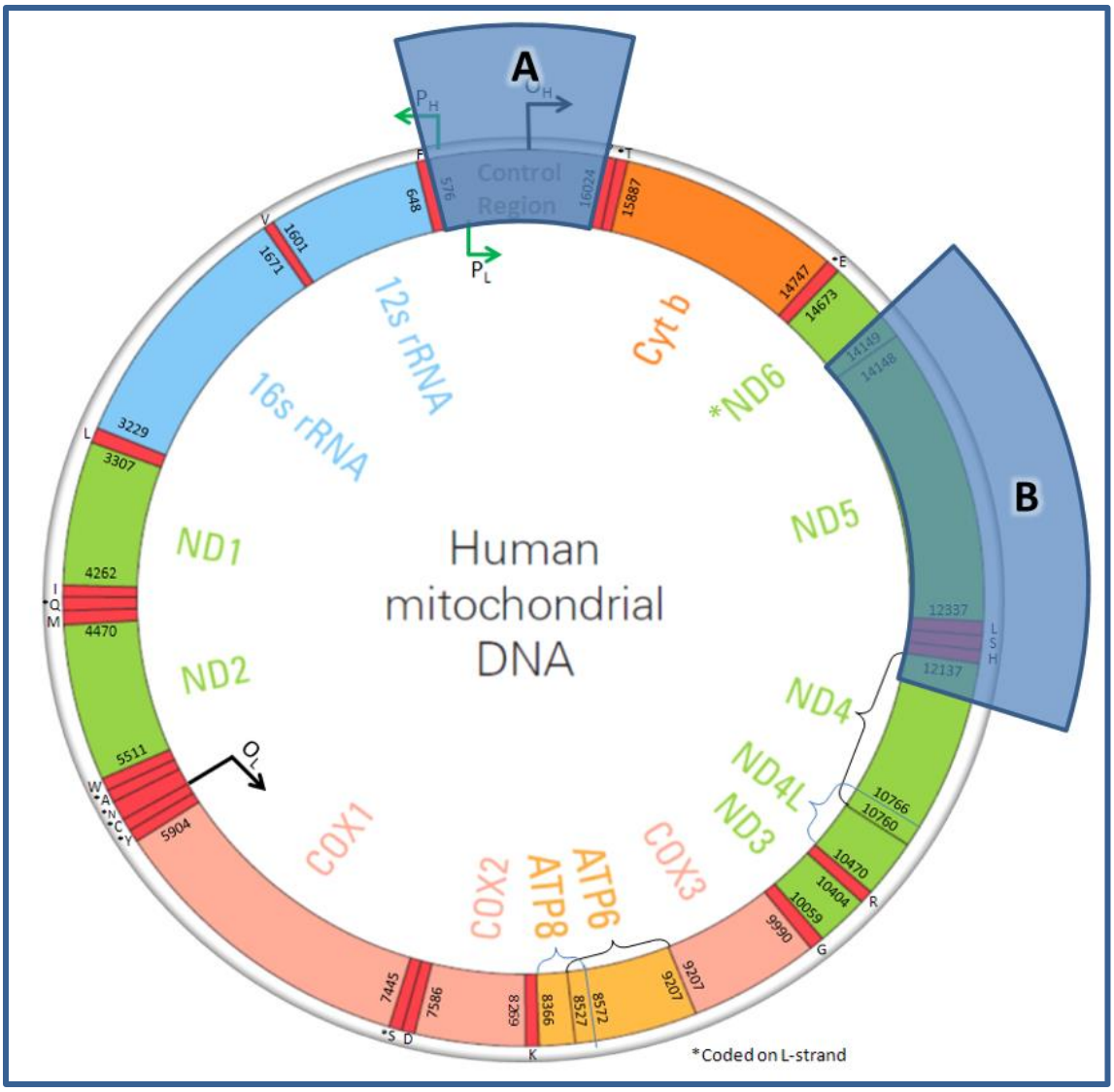


Figure 18: Sequenced regions of the human mtGenome. Regions sequenced using the Sanger method are shaded blue. A) HV1 and HV2, located within the control region, were amplified and sequenced using primer pair combinations of R1, B1, R2 and C1 (Table 7). B) Primer pairs $ND5_A$, $ND5_B$ and $ND5_C$ were used to sequence the ND5 gene in overlapping segments. A portion of the sequenced coding region also spans NADH dehydrogenase subunit 4 (ND4), tRNA histidine (H), tRNA serine 2 (S), tRNA leucine 2 (L) and NADH dehydrogenase subunit 6 (ND6) genes.

A total of 0.34 pg of template mtDNA was used in a 25 µL amplification reaction.[236] PCR master mix consisted of 12.5 µL of AmpliTaq Gold® PCR Master Mix (Applied Biosystems), 76.8 ng/µL BSA, and 180 nM each of the forward and reverse primers. The amplification thermal cycling parameters were as follows: 95°C hold for 11 minutes; 28 cycles of 95°C for 10 seconds, 60°C for 45 seconds and 72°C for 1 minute; a 15°C hold for 10 minutes; and a final 4˚C hold.

Post-PCR product visualization on an agarose yield gel was performed to ensure amplification was successful before proceeding with downstream sequencing. Following amplification, PCR product was purified using ExoSAP-IT® (USB Corp) and then cycle sequenced with BigDye® Terminator version 1.1 (Applied Biosystems). Cycle sequencing reactions were performed without post-PCR quantification and included 1 µL of non-normalized PCR product. BigDye® XTerminator™ (Applied Biosystems) was used for post-cycle sequencing purification. Sequence quality was assessed using eFAST™ Software (UNTHSC, Ft. Worth, TX). Sequence electropherograms were aligned to the rCRS (NC_012920.1) and analyzed with MTexpert™ Software (MitoTech LLC, Santa Fe, NM).[139]

Bioinformatic and Statistical Analysis

Haplogroup designation was assigned to each mtDNA haplotype using HaploGrep (Medical University Innsbruck, Innsbruck, Austria) (http://haplogrep.uibk.ac.at/), a web-based application for automated haplogroup classification.[237,238] Sequence variants were evaluated independently and characterized based on the type of genetic alteration, *i.e.*, neutral or

deleterious, using a phylogenetic approach.[239] Also, each divergent base call from the rCRS was screened against the human mitochondrial genome database (http://www.MitoMap.org) to identify novel or unreported mtDNA variants.[240,241] Non-synonymous genetic variants were further investigated with the open-source bioinformatics tools PolyPhen-2 (http://genetics.bwh.harvard.edu/pph2/index.shtml) and PMut (http://mmb.pcb.ub.es/PMut/).[242,243] These two tools use straightforward physical and comparative considerations to predict the possible impact of amino acid substitutions on the structure and function of human proteins, *i.e.*, their damaging affect and pathological potential. Finally, Geno3D (https://geno3d-prabi.ibcp.fr/cgi-bin/geno3d_automat.pl?page=/GENO3D/geno3d_home.html) was used for protein molecular modeling to generate a 3D representation of the target protein of interest, *i.e.*, ND5, for a perspective analysis.[244]

Group differences were evaluated using the student's T-test and ANOVA. Linear regression was applied to analyze the relationship between several independent variables where Gleason score and TNM staging were used as dependent variables. Statistical analyses were performed using IBM© SPSS™ Statistics 20 (IBM Corporation) and Excel. All tests were two-sided and significance was set at a $p < 0.05$.

Results

Two specific regions of the mtGenome were sequenced for 12 prostatectomy patients with BPH or PCa. Appendix A lists the sequence profiles of each individual for HV1, HV2 and ND5, respectively. Combined, a total of 110 sequence variants at 62 different positions were detected. Two overall observations worth noting are: (1) all observed variants are homoplasmic; and (2) no somatic mutations were observed, *i.e.*, no differences were observed between tissue and blood samples among the nine PCa subjects tested.

The non-coding region, spanning positions 15,392 to 16,390 and 49 to 544 for HV1 and HV2, respectively, generated a total of 92 sequence variants at 48 different positions when compared to the rCRS. Transitional events constituted 69.5% of the variants observed where 40 were pyrimidine transitions (C>T or T>C) and 26 were purine transitions (A>G or G>A). In total, only three transversions were observed. Interestingly, all three are the same A>C variant at position 16,183 which was detected once in a BPH sample (1365) and twice in a PCa sample from the same individual (blood and tissue from 2208). Seven of the twelve samples exhibited an insertion at position 309 (309.1C and 309.3C) and all the samples exhibited the common 315.1C insertion. The 309 and 315 insertions are very common mtDNA polymorphisms and are designated as "hot spot" mutations by HaploGrep. An additional insertion at position 16,193 (16,193.1C) and three deletions (249delA, 291delAA and 16,189delT) were observed in the

control region. Interestingly, two of these three deletions were observed in a single PCa patient (2164t and 2164b) (Appendix A).

A 2,388 bp section of the mtGenome, encompassing all of TRNH, TRNS2, TRNL2, ND5 and portions of the ND4 and ND6 genes, was sequenced in the mitochondrial coding region. A total of 18 variants at 14 different positions, as compared to the rCRS, were observed (Table 8 and Appendix A). One variant was observed in the TRNL2 gene, eleven in the ND5 gene and two in the ND6 gene. Eleven of the observed variants are synonymous transitions and the remaining three are non-synonymous transversions (12,631T>G, 12,713T>G and 13,759G>A) resulting in the following ND5 amino acid substitutions, p. Ser(1)99Ala, p. Ile126Ser(2) and p. Ala475Thr, respectively.

Position (rCRS)	Gene	Transcript Location	Mutation		Effect
*12,308	TRNL2	tRNA - leucine	Transistion	NA	NA
*12,372	MT-ND5	12th residue at codon position 3	Transition	S	p. Leu12Leu
12,612	MT-ND5	92nd residue at codon position 3	Transition	S	p. Val92Val
12,631	MT-ND5	99th residue at codon position 1	Transversion	NS	p. Ser(1)99Ala
12,705	MT-ND5	123rd residue at codon position 3	Transition	S	p. Ile123Ile
12,713	MT-ND5	126th residue at codon position 2	Transversion	NS	p. Ile126Ser(2)
13,263	MT-ND5	309th residue at codon position 3	Transition	S	p. Gln309Gln
13,368	MT-ND5	344th residue at codon position 3	Transition	S	p. Gly344Gly
13,434	MT-ND5	366th residue at codon position 3	Transition	S	p. Met366Met
13,759	MT-ND5	475th residue at codon position 1	Transition	NS	p. Ala475Thr
14,016	MT-ND5	560th residue at codon position 3	Transition	S	p. Lys560Lys
14,145	MT-ND5	603rd residue at codon position 3	Transition	S	p. Thr603Thr
14,221	MT-ND6	151st residue at codon position 3	Transition	S	p. Trp151Trp
14,233	MT-ND6	147th residue at codon position 3	Transition	S	p. Asp147Asp

*Observed in three different individuals
Not applicable (NA), synonymous (S) and nonsynonymous (NS)

Table 8: Observed mtDNA variants. Fourteen different mtDNA variants were observed in a 2,388 bp sequenced segment of the mitochondrial coding region. Each variant was fully

characterized to determine the affected gene, its transcriptional location, type of mutation and its effect on the translated protein.

A phylogenetic association was performed using HaploGrep where each sample was classified based on mtDNA polymorphisms indexed by Phylotree (http://www.phylotree.org) (Table 9).[238] The major haplogroups represented in these samples are H, T, U, J and C (listed in order based on evolutionary distance from haplogroup L, where H is the furthest and C is the closest) which all originate from the West and East Eurasian groups. The global private mutations identified by HaploGrep were investigated further to identify their prevalence or novelty. Each variant was cross referenced with MitoMap and the Single Nucleotide Polymorphism database (dbSNP) (www.ncbi.nlm.nih.gov/snp). Although each variant observed within the control region has been previously reported, 11 are mutations that have been found in prostate tumors and/or associated with PCa (Table 10). A T-test, however, did not reveal any significance ($p>0.05$) between the number of variants observed at these 11 positions and the BPH or PCa groups. Within the coding region, three variants have also been previously reported and associated with PCa (Table 10).

	Sample ID	Gleason Grade	Stage	Age	Haplogroup
BPH	1350	-	-	55	H20
	1365	-	-	60	H2a2a
	1391	-	-	61	T2b6b
PCa	1727	9	T3bN0Mx	61	J1c2o
	1838	8	T2cN0Mx	74	H11a2a
	2114	7	T3aNxMx	58	H6
	2160	7	T3NxMx	62	H2a2a
	2164	9	T3bN0Mx	54	C1b8
	2167	9	T3aN0Mx	76	U4a3
	2178	7	T2cNxMx	60	U5a1b+16362
	2208	7	T2cNxMx	62	H1c3
	2238	7	T2cNxMx	59	U5a1f1

Table 9: Demographic, clinicopathological and haplogroup status of PCa and BPH patients. Haplogroup designation was used in a phylogenetic approach to screen the observed mtDNA variants against known ancestral polymorphisms according to Phylotree. Therefore, mtDNA variants associated with the major groups H, T, U, J and C, were not considered for further characterization.

Position (rCRS)	Location	Variant	Observed Frequency			Reference
			BPH	PCa	Total	
16,183	HV1	A>C	33%	11%	17%	Chen, *et al* . (2002)
16,189	HV1	T>C	0%	11%	8%	Chen, *et al* . (2002)
16,193.1	HV1	C ins	0%	11%	8%	Chen, *et al* . (2002)
16,249	HV1	T>C	0%	11%	8%	Chen, *et al* . (2002)
16,304	HV1	T>C	33%	11%	17%	Chen, *et al* . (2002)
16,311	HV1	T>C	0%	22%	17%	Chen, *et al* . (2002)
73	HV2	A>G	33%	33%	33%	Chen, *et al* . (2002)
146	HV2	T>C	33%	0%	8%	Jeronimo, *et al* . (2001)
195	HV2	T>C	0%	33%	25%	Chen, *et al* . (2002)
489	HV2	T>C	0%	22%	17%	Chen, *et al* . (2002)
499	HV2	G>A	33%	0%	8%	Chen, *et al* . (2002)
12,308	tRNA Leu	A>G	0%	33%	25%	Jeronimo, *et al* . (2001)
12,372	ND5	G>A	0%	33%	25%	Jeronimo, *et al* . (2001)
12,705	ND5	C>T	0%	11%	8%	Jeronimo, *et al* . (2001)

Table 10: Observed sequence variants previously associated with prostate cancer. Of the 63 different variants observed, 14 (listed here) have been previously associated with PCa. Observed frequency was calculated by dividing the number of times the variant was observed by the number of samples per group.

PolyPhen-2 was then used to analyze the three missense mutations to determine if the resultant change in amino acid residues would negatively impact the structure and function of the mature ND5 protein. For analysis the Ensembl protein identifier (ENSP00000354813) for ND5 was used and individual queries were performed for amino acid positions 99 (S>A), 126 (I>S) and 475 (A>T). According to the PolyPhen-2 results, all three missense mutations were determined to be benign and would not have a functional effect. A similar analysis was performed using PMut but with an exhaustive query approach. Every position of the ND5 primary amino acid sequence was mutated individually to determine which mutations would

have a consequential effect on the structure and function of the protein. Similar to the PolyPhen-2 results, the analysis using PMut determined that the p. Ser(1)99Ala and p. Ala475Thr mutations are neutral and not likely to cause a functional change in ND5. However, according to PMut, an isoleucine to serine mutation at position 126 of ND5 (p. Ile126Ser(2)) may cause a functional change in the ND5 subunit thereby making it a pathological mutation. A 3D structure of ND5 was generated using Geno3D automatic modeling and viewed using the molecular graphics visualization tool RasMol v. 2.7.5.2 (http://www.openrasmol.org/) (Figure 18).[245] According to this model, amino acid position 126 is located within α-helix number five and may be sensitive to residue substitutions that disrupt its secondary structure.

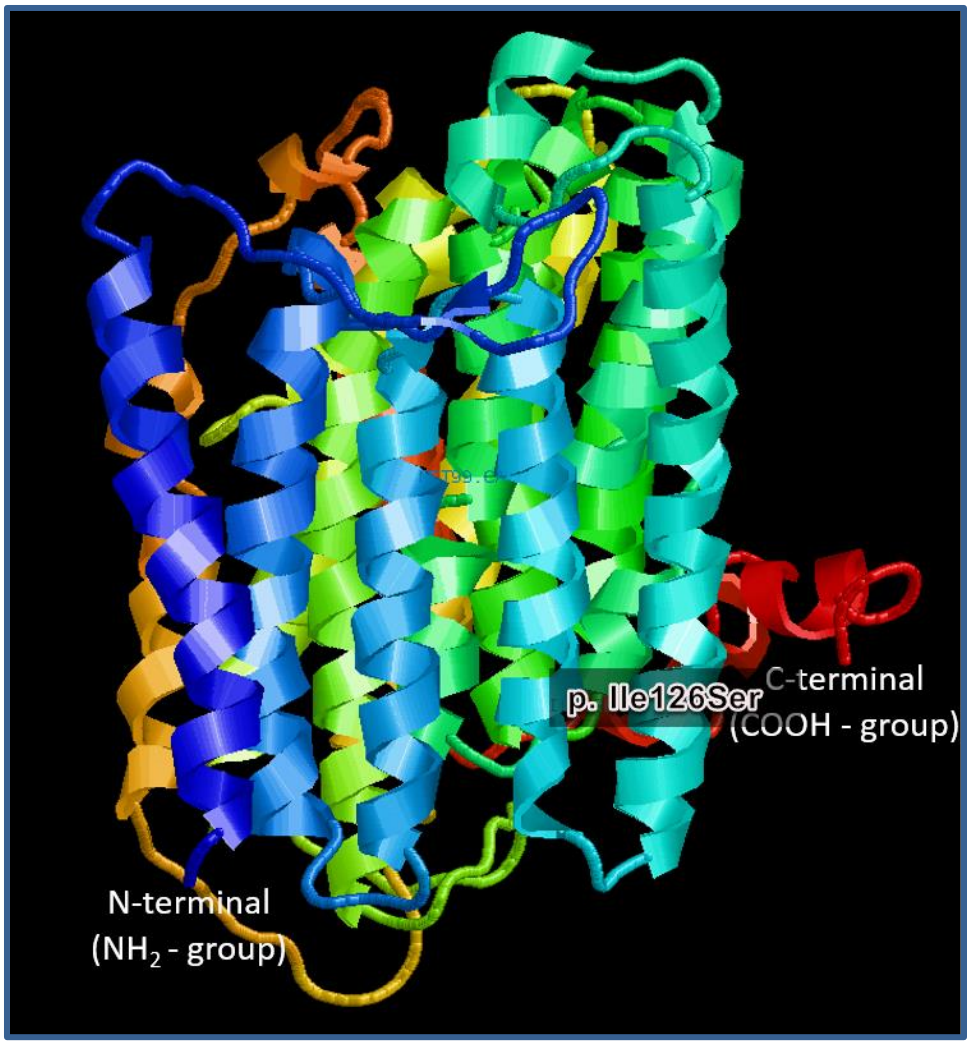


Figure 19: Three dimensional model of ND5. The ND5 protein is presented here in its predicted tertiary structure as an isolated subunit with the first five α-helices (beginning from the N-terminal group) in the foreground. According to this model, amino acid position 126 is located within helix number five. The PMut software predicts that the 12,713T>G transversion is pathological; it results in a non-synonymous mutation from isoleucine to serine at amino acid position 126 (p. Ile126Ser(2)). Model was generated using Geno3D's automatic modeling program and viewed in RasMol v. 2.7.5.2.

Conclusion and Discussion

A total of 110 mtDNA variants (at 62 different positions) were observed in three sequenced regions of the mtGenome among 12 prostatectomy patients: three BPH and nine PCa. To avoid making erroneous conclusions and false disease associations due to confounding mitochondrial population stratification, each samples' haplogroup status was taken into consideration in a phylogenetic approach for evaluating the observed mtDNA variants.[246,247] All twelve samples were classified as West and East Eurasian haplogroups and phylogenetic consistency was met. As such, polymorphisms associated with each sample's designated haplogroup were not considered further and only variants not described in Phylotree (global private mutations) were. Proportionally, the PCa samples exhibited a slightly higher mutational load, *i.e.*, higher number of mtDNA variants, than the BPH group although no significant difference ($p>0.05$) was observed between them. Also, no significant associations ($p<0.05$) were detected between any of the observed mtDNA variants and Gleason score or TNM staging. Of the 14 different coding region variants observed, three are missense mutations (12,631T>G, 12,713T>G and 13,759G>A) located in the ND5 gene and cause the following amino acid substitutions: p. Ser(1)99Ala, p. Ile126Ser(2) and p. Ala475Thr, respectively. Two of the missense mutations (12,631T>G and 13,759G>A) were observed in the PCa group whereas only one (12,713T>G) was observed in BPH control group.

Many *in silico* tools have been developed to predict the pathogenicity of missense variants and are broadly classified into three categories: (1) sequence and evolutionary conservation-based; (2) protein sequence and structure-based; and (3) supervised learning-based methods.[248] Bioinformatic analyses using PolyPhen-2 (a protein sequence and structure-based tool) and PMut (a supervised learning-based tool) were performed on the three missense mutations to determine if the resultant change in amino acid residues would negatively impact the structure and function of the mature ND5 protein. Results of PolyPhen-2 and PMut were concordant save for one prediction. Both PolyPhen-2 and PMut determined that the p. Ser(1)99Ala and p. Ala475Thr mutations are neutral but PMut designated the p. Ile126Ser(2) mutation as potentially pathological whereas PolyPhen-2 did not. The consequence of a missense variant on the function of a protein depends on several physicochemical properties of the affected amino acid. For example, divergence between the original (wild-type) and substitute (mutant) amino acids can be tolerated if their functional groups have similar size, charge and solubility. However, amino acid residues that have wholly different side-chain properties are more likely to have a greater effect. As a protein sequence and structure-based tool, PolyPhen-2 assesses the degree of similarity between the amino acid residues in question as well as phylogenetic and structural information to characterize the substitution and predict its functional significance.[242] Whether or not a missense mutation is pathogenic or neutral can depend on a number of different factors that cannot be captured using basic protein sequence and structural information alone. Alternatively, PMut uses neural networks trained on a large database of disease-associated and neutral mutations to achieve a high level of accuracy to

predict disease-associated mutations.[243] It is unclear why PolyPhen-2 and PMut generated disparate predictions of pathogenicity for the 12713T>G variant but it may be due to alignment diversity in their respective analyses. Both the number and type of orthologues used in multiple sequence alignments can affect the accuracy of pathogenicity predictions.[249] An extraneously diverse compendium of ortholgues used in the PolyPhen-2 analysis may have resulted in its non-pathogenic prediction. Nonetheless, perspective analysis on ND5 was performed based on PMut's pathogenic prediction despite the conflicting report by PolyPhen-2.

The 3D model of ND5 (Figure 19) generated by Geno3D shows that the location of the 12,713T>G variant causes an isoleucine to serine substitution within the fifth helix of the folded protein. The helical propensity of isoleucine (0.41 Kcal/mol) is more favorable than that of serine (0.50 Kcal/mol).[250] In addition, isoleucine is bulky and strongly hydrophobic which normally relegates it to the interior core of a protein fold. In contrast, serine is a smaller, more flexible amino acid that is hydrophilic making it indifferent to the interior or surface of a protein. These physical, chemical and structural differences are likely to have a regional effect on the secondary structure of ND5 which may subsequently change its tertiary and quaternary structure and affect the overall functionality of Complex I in the ETC. Prostate cells harboring this variant may exhibit mitochondrial dysfunction characterized by increased oxidative stress. Although this study was limited by a small number of samples, it demonstrated distinct differences between the mtGenome of prostate tissue from PCa and BPH patients. By expanding sample size and sequence analysis to include the entire mtGenome, it is expected

that results will be consistent and may lead to more compelling evidence that supports the role of changes to the mtGenome with PCa progression.

Study 2: Whole mtGenome Sequencing of Normal and PCa Blood Samples

Results of the preliminary study are consistent with a previous report where mtDNA mutations detected in prostate tissue are also found in the serum of the same PCa patient.[251] Homoplasmy between prostate tissue and bodily fluid specimens is likely due to maternal inheritance or germline mutational events but may also be indicative of a systemic component to PCa.[252] Therefore, specific mtDNA variants and/or mutational load observed in peripheral blood may serve as a PCa marker that correlates with disease progression. Here, the entire mtGenome of a large PCa study cohort is sequenced in order to evaluate group differences in overall mutational load with an emphasis on missense mutations with a predicted functional impact.

Materials and Methods:

Samples

The University of North Texas Health Science Center Department Molecular and Medical Genetics, the Texas Center for Health Disparities and the Methodist Health Systems' Prostate Screening and Awareness Program, Methodist Hospital Dallas have collaborated to build Texas' largest known repository of sera and DNA with a corresponding database of ethnicity, age, and PSA levels pre-diagnosis of PCa. Thirty genomic DNA extract samples from the in-house

repository were used as training set to assess the performance of the proposed next-generation sequencing strategy. These 30 samples were collected from men who screened negative for PCa based on normal serum PSA (<4.0 ng/mL) and DRE. For the larger study, the same NC, NED and BCR/MET samples from the SABOR cohort used in the Chapter III study are tested (Table 5). Biological specimens used for this study were obtained with approval from the Office for the Protection of Human Subjects-Institutional Review Board (Project #2012-080 and #2013-116; "Damage to Mitochondrial DNA in Age-Related Diseases: Alzheimer's and Prostate Cancer" and "Evaluation of Biomarkers in Prostate Cancer Patients; UNT Health Science Center, Fort Worth, TX).

Mitochondrial DNA Sequencing

Whole mtGenome sequencing was performed for the 134 samples in the NC, NED and BCR/MET groups using a massively parallel sequencing platform, the Ion Torrent PGM (Life Technologies). However, prior to sequencing the SABOR cohort samples, a training set was tested to assess the performance of the proposed next-generation sequencing strategy. Mitochondrial DNA enrichment is a required first-step in order to generate a suitable amount of amplifiable template for the sequencing reactions. Multiple displacement amplification (MDA) is an enrichment strategy and was used to amplify the 30 sample training set. MDA reactions are isothermal and based on the naturally occurring rolling circle amplification used for replication of circular DNA molecules, *e.g.*, mtDNA. The ϕ29 DNA polymerase used for MDA applications is highly processive, *i.e.*, reproduces DNA in long fragments, typically greater than

10 kb, and has a robust proof-reading capability which ensures high sequence fidelity.[253] For this study, a commercially available MDA kit, the REPLI-g® Human Mitochondrial DNA Kit (Qiagen, Hilden, Germany), was used. This kit's targeted enrichment is effective at amplifying mtDNA without co-amplifying NUMTS. Five nanograms of template mtDNA was added to the enrichment reactions using the REPLI-g® Human Mitochondrial DNA Kit following the manufacturer's protocol.[254]

A second enrichment strategy was selected for the 134 SABOR cohort samples. Mitochondrial DNA amplification was performed in two overlapping fragments using two primer pairs originally reported by Gunnarsdóttir *et al*. (Table 11).[255] Two separate long-range PCR amplifications were performed: amplicon one (Amp 1) spans positions 644 to 8,982 and amplicon two (Amp 2) spans positions 8,789 to 877 of the rCRS (Figure 20). Long-range PCR reactions were carried out using the PrimeSTAR® GXL DNA Polymerase reagent kit (Takara Bio Inc. Otsu, Shiga Japan). PrimeSTAR GXL DNA Polymerase allows amplification of products equal to or greater than 30 kb in length while maintaining exceptionally high fidelity and is suitable for GC-rich templates. Ten nanograms of template mtDNA was added to the reaction mix for a final volume of 50 µL and ran on a 96-Well GeneAmp® PCR System 9700 following the manufacturer's recommended protocol.[256]

	Primer	rCRS Position	Sequence (5'-3")
Forward	mtG1F	622-644	GACGGGCTCACATCACCCCATAA
	mtG1R	9,003-8,982	GCGTACGGCCAGGGCTATTGGT
Reverse	mtG2F	8,764-8,789	GCCACAACTAACCTCCTCGGACTCCT
	mtG2R	898-877	GGTGGCTGGCACGAAATTGACC

Table 11: Long-range PCR primers. MtG1F and mtG1R were used for amplicon one (8.3 kb) and mtG2F and mtG2R were used for amplicon two (8.6 kb).

.

In order to accurately normalize the enriched extracts for library preparation, samples were quantified using the Qubit® 2.0 Fluorometer with the dsDNA BR Assay (Thermo Fisher Scientific Inc., Carlsbad, CA) following the manufacturer's protocol.[257,258] This method accurately estimates the concentration of double-stranded DNA (dsDNA) between 100pg/µL to 1µg/µL. Based on the Qubit® results, the PCR products from Amp 1 and Amp 2 were mixed to form a single equimolar concentration of "combined" PCR product for each sample. Enriched product from MDA or long-range PCR was then purified using Agencourt AMPure® XP Beads (Agencourt Bioscience Corp., Beverly, MA) to remove unincorporated PCR components and very short DNA fragments.

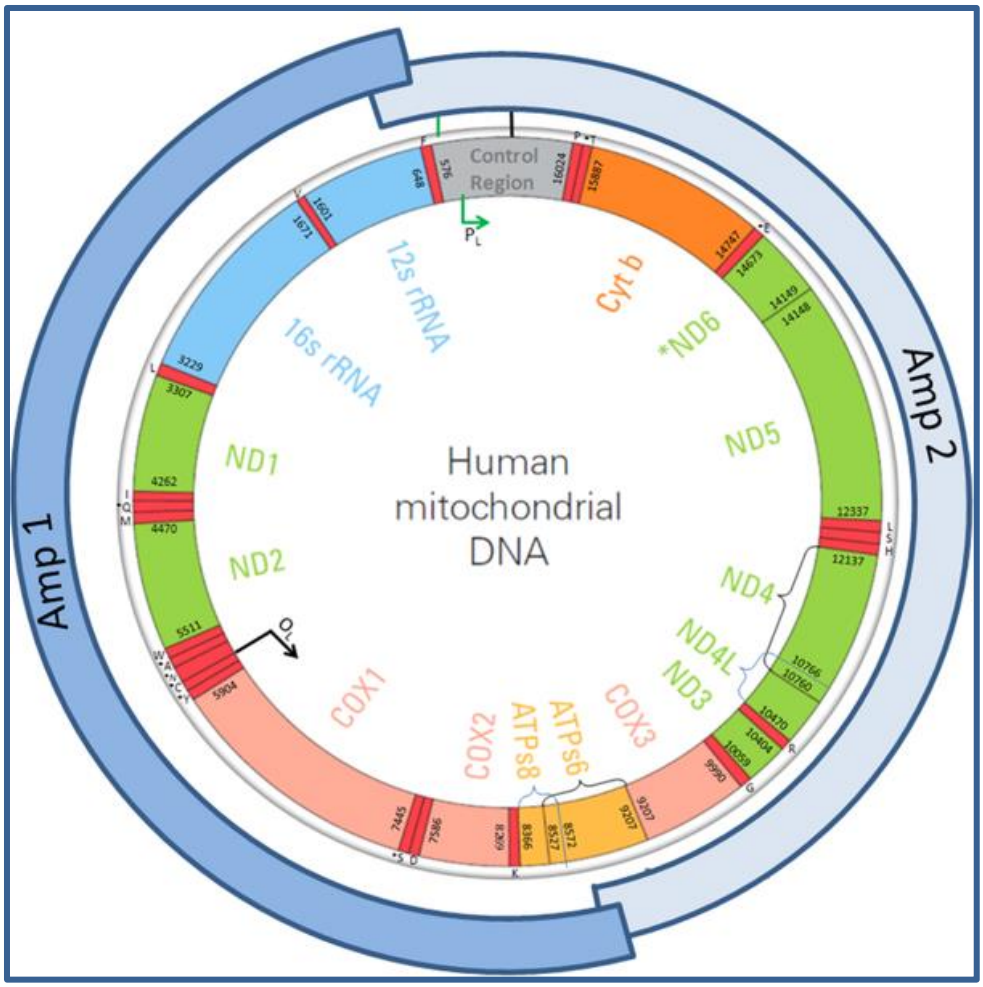


Figure 20: Whole mtGenome enrichment strategy. For mtDNA enrichment, long-range PCR of two large amplicons was used. Approximately 200 bp of overlap between alternate amplicon ends ensured complete coverage of the whole mtGenome. Amplicon 1 (Amp 1) spans positions 644 to 8,982 and Amplicon 2 (Amp 2) spans positions 8,789 to 877.

The enriched and purified mtDNA samples were then prepared for sequencing using the NEBNext® Fast DNA Fragmentation & Library Prep Set (New England Biolabs, Ipswich, MA). Following fragmentation and end repair of the mtDNA, library adaptors were added using NEXTflex™ DNA Barcodes (Bioo Scientific Corp., Austin, TX). The cleanup of adaptor-ligated DNA

and size selection were performed simultaneously using Agencourt® AMPure® XP Purification system (Agencourt Bioscience Corp.). Dual bead-based fragment size selections were performed at a 60% and 15% bead to DNA ratio for a target insert size of 400 bp. A short 10-cycle PCR amplification of adaptor-ligated and sized-selected DNA was performed following the manufacturer's recommended protocol followed by post-PCR cleanup using the Agencourt® AMPure® XP Purification system. The Agilent 2100 Bioanalyzer and Agilent DNA 1000 Assay (Agilent Technologies) were used to measure the concentration and size distribution of each sample. Based on these results, each sample was normalized to 26 pM and pooled. Emulsion PCR and template-positive ion sphere particle (ISP) enrichment were performed on the Ion OneTouch™ 2 Instrument and the Ion OneTouch™ ES (Life Technologies) using the Ion PGM™ Template OT2 400 Kit (Life Technologies). Finally, the enriched ISPs were loaded onto an Ion 318™ Chip Kit v2 (Life Technologies) and whole mtGenome sequencing was performed on the Ion Torrent PGM using the Ion PGM™ Sequencing 400 Kit (Life Technologies).

Sequence Data and Statistical Analysis

Whole mtGenome sequence data processing, management and distribution were directed by the Ion Torrent™ Server (Life Technologies). Data files (.fastq) were uploaded into NextGENe® 2nd Generation Sequence Analysis Software (SoftGenetics, LLC, State College, PA) for file conversion (.fastq to .fasta), sequence alignment and variant calling. A FASTA file is a text-based format which includes nucleotide sequence data using-single letter codes.[259] The FASTQ file format is very similar to the FASTA but includes quality score information for each

nucleotide in addition to sequence data.[260] NextGENe® uses a modified Burrows–Wheeler transform (BWT) alignment method.[261] However, the mtGenome is circular and NextGENe® only aligns to linear DNA references. Therefore, the rCRS was modified to ensure proper mapping of reads that span and flank the origin; bases 15,569 to 16,569 were duplicated at the beginning of the reference and bases 1 to 1,000 were duplicated at the end of the reference (rCRS+2,000 positions 15,569-16,569; 1-16,569; 1-1,000). One thousand bases were duplicated on each end of the reference to accommodate any events of an exceptionally large sequence read aligned with greater than a 1,000 bp gap on the duplicated side of the origin. Using NextGENe Viewer (Softgenetics, LLC) data were presented in both a tabular summary of the haplotype, as well as a visual graphic, which permitted close inspection of the sequence reads and alignments. NextGENe analysis parameters used in this study are given in Figure 21. Mutation reports were generated for each sample and uploaded into MitoSAVE (Institute of Applied Genetics, University of North Texas Health Science Center, Fort Worth, TX) for haplotype assignment.[262] MitoSAVE is an Excel-based workbook that converts data within the variant call format (VCF) file into mtDNA haplotypes using phylogenetically-established nomenclature as well as rule-based alignments. Each haplotype was then formatted for import into HaploGrep2 (University of Innsbruck, Innsbruck, Austria) using HSD Generator where data were analyzed for haplogroup assignment and global private mutations (GPMs).[237,238,262] GPMs were evaluated independently and characterized based on the nature of the genetic alteration, *e.g.*, synonymous or nonsynonymous. Also, each divergent base call from the rCRS was screened against MitoMap to identify novel or unreported mtDNA variants.[263] Non-

synonymous genetic variants were further investigated with the open-source bioinformatics tool PolyPhen-2 to predict pathogenicity.[242]

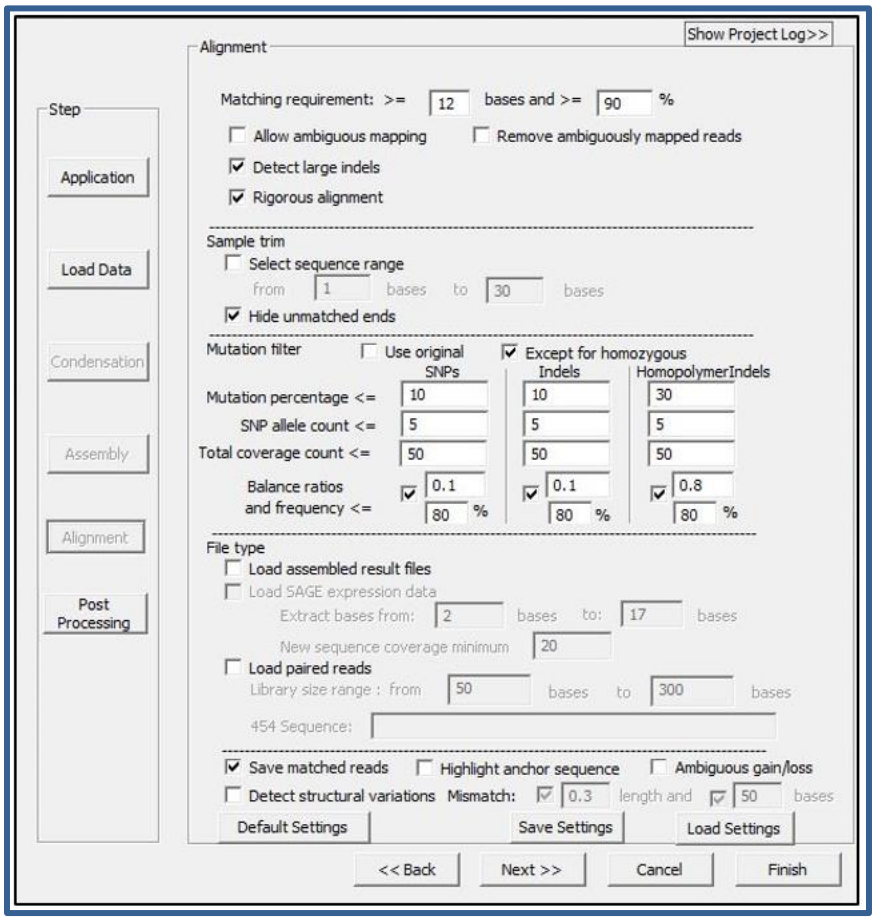


Figure 21: Screenshot from NextGENe® showing alignment parameters. NextGENe software settings and analysis parameters were adopted from Parson *et al*.[264]

Statistical analyses were performed using IBM© SPSS™ Statistics 21 (IBM Corporation) and Excel Data Analysis Tools (Microsoft Corporation). Data sets were tested for normality (Shapiro-Wilk test), equality of variances (Levene's test) and comparison of means. Group differences were tested using ANOVA or Kruskal-Wallis when the assumption of equal variances was not met. Pearson correlation was performed on sample variables to test for linear associations where linear regression analysis was used to determine predictor status. Tests were two-tailed and significance was assessed at an alpha level of 0.05.

Results

Whole mtGenome sequence data of 30 blood samples from men screened for PCa were generated using a next-generation sequencing protocol. A new front-end enrichment strategy that utilizes multiple displacement amplification (Qiagen's REPLI-g mitochondrial DNA kit) was used instead of the more traditional targeted PCR. Despite high quality run summary statistics (Figure 22), very few called bases aligned to the rCRS. Of the 30 samples tested, 28 generated sequence data and two did not. The 28 samples averaged 81,024 total reads but only 5.38% mapped to the rCRS (Table 12). The low percentage of aligned reads resulted in a very poor depth of coverage which averaged approximately 20X coverage (target was 50X coverage) per position. The two samples that did not generate sequence data is presumably due to poor mtDNA enrichment; no further testing was performed to elucidate the cause of drop-out for these two samples. However, an independent study was conducted to evaluate the enrichment success of MDA because of the poor performance of the 30 sample training set.

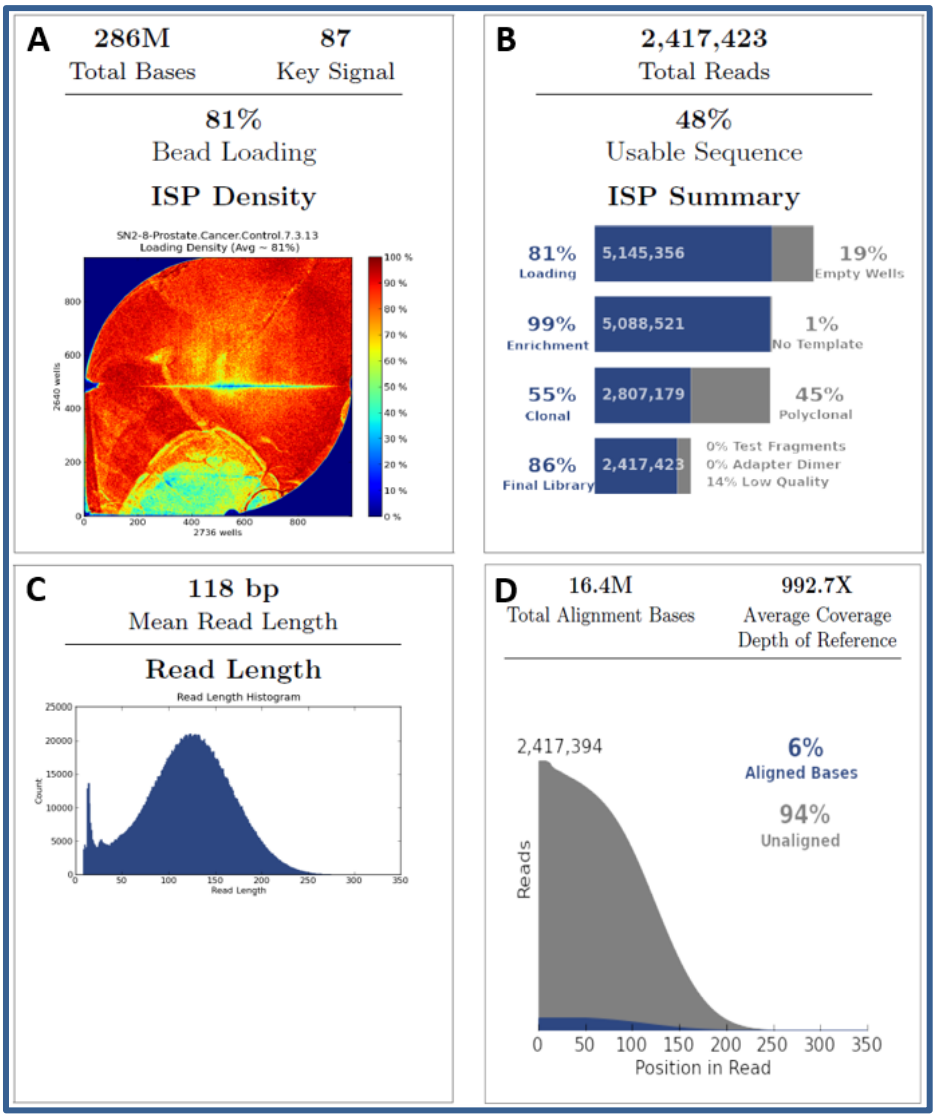


Figure 22: PGM run summary. A) ISP loading onto the Ion 318™ Chip was high (81%), indicated by the excellent coverage of the density map. B) Although ISP enrichment was high (99%), the amount of polyclonal ISPs (45%) suggests normalization of the mtDNA library was underestimated and too much was added to the emulsion PCR reaction. C) Size selection did not achieve the desired 400 bp target length, a mean read length of 118 bp was achieved. D) Of the 2.4 million reads, only six percent of the bases aligned to the rCRS.

		Reads Mapped to rCRS		Reads not Mapped	
Sample	Total Reads	Count	Percent Total	Count	Percent Total
023281	59,267	2,652	4.47%	56,615	95.53%
023299	40,829	5,741	14.06%	35,088	85.94%
030758	78,031	4,065	5.21%	73,966	94.79%
066797	73,790	4,353	5.90%	69,437	94.10%
066803	125,348	3,745	2.99%	121,603	97.01%
067047	197,289	3,971	2.01%	193,318	97.99%
067058	55,418	2,509	4.53%	52,909	95.47%
067098	64,344	1,704	2.65%	62,640	97.35%
067111	60,495	1,718	2.84%	58,777	97.16%
067141	83,704	3,512	4.20%	80,192	95.80%
067145	94,330	5,135	5.44%	89,195	94.56%
208074	74,599	4,074	5.46%	70,525	94.54%
352705	70,665	7,433	10.52%	63,232	89.48%
352805	110,391	4,536	4.11%	105,855	95.89%
352807	60,200	4,583	7.61%	55,617	92.39%
*352820	NA	NA	NA	NA	NA
366970	46,148	3,627	7.86%	42,521	92.14%
367007	72,143	5,964	8.27%	66,179	91.73%
380685	57,510	1,601	2.78%	55,909	97.22%
380709	69,738	3,199	4.59%	66,539	95.41%
380725	109,023	7,261	6.66%	101,762	93.34%
380745	76,796	4,359	5.68%	72,437	94.32%
380772	53,044	1,668	3.14%	51,376	96.86%
380775	168,677	9,598	5.69%	159,079	94.31%
380824	114,460	11,508	10.05%	102,952	89.95%
380826	64,875	1,353	2.09%	63,522	97.91%
380827	51,029	627	1.23%	50,402	98.77%
*461477	NA	NA	NA	NA	NA
853881	66,091	2,687	4.07%	63,404	95.93%
853885	70,449	4,513	6.41%	65,936	93.59%
Mean	81,024	4,203	5.38%	76,821	94.62%
SD	35,732	2,474	2.88%	34,595	2.88%

*No sequence data was obtained
Not Applicable (NA)

Table 12: rCRS alignment summary for 30 sample training set. Of the 30 samples tested, two did not generate sequence data likely due to poor mtDNA enrichment. The remaining 28 samples averaged 81,024 total reads but only 5.38% mapped to the rCRS.

A comparison of whole mtGenome sequencing success with and without mtDNA enrichment was performed using eight peripheral blood extracts. These eight samples were sequenced twice, once in which MDA was used for mtDNA enrichment and another where only crude genomic extract was used for library preparation. They were prepared for sequencing using the Nextera® XT DNA Sample Prep Kit (Illumia, San Diego, CA) and sequenced on the MiSeq® Benchtop Sequencer (Illumina). The average observed enrichment, as compared to direct sequencing of extracts, was 507.99% using MDA (Table 13). Due to the large magnitude of reads processed, the depth of coverage was proportionately large and averaged approximately 2,000X per position. However, only 27.24% of total reads generated by MDA enrichment mapped to the rCRS. Qiagen's REPLI-g mitochondrial DNA kit may be suitable for mtGenome enrichment with the MiSeq® platform but the Ion Torrent PGM's total sequencing capacity is much lower and requires a more robust enrichment strategy for adequate coverage. Therefore, a long-range PCR strategy was selected for the 134 SABOR cohort samples. The targeted PCR approach was used to amplify the entire mtGenome in two large overlapping amplicons. Five DNA extract dilutions (75, 37.5, 17, 8.5 and 5 ng) were evaluated to determine the optimal quantity of DNA for robust PCR product yield. Additionally, the performance of two primer pairs, mtG1F/R (Amp 1) and mtG2F/R (Amp 2), designed for whole mtGenome PCR enrichment were tested. All five dilutions successfully generated PCR product at the expected fragment sizes, 8.3 and 8.6 kb, for Amp 1 and Amp 2, respectively (Figure 23). Both 8.5 and 17 ng template amounts generated very distinct and clean bands on a one percent yield gel. Therefore, 10 ng of template DNA was added to the long-range PCR enrichment reactions.

	Total Reads		Reads Mapped to rCRS			
Sample	No Enrichment	Enrichment	No Enrichment		Enrichment	
1	515,389	1,828,746	317	0.06%	1,078,232	58.96%
2	1,382,767	4,043,332	1,024	0.07%	1,517,619	37.53%
3	1,392,769	878,957	12,080	0.87%	69,899	7.95%
4	500,067	1,510,918	544	0.11%	283,131	18.74%
5	1,108,920	2,619,132	327	0.03%	605,037	23.10%
6	311,087	3,484,952	173	0.06%	4,775	0.14%
7	633,863	3,680,492	215	0.03%	1,534,002	41.68%
8	141,687	2,632,417	103	0.07%	785,640	29.84%
Mean	748,319	2,584,868	1,848	0.16%	734,792	27.24%
SD	452,221	1,047,397	3,877	0.27%	566,561	17.76%

Table 13: Assessment of mtDNA enrichment using REPLI-g®. On average, only 27.24% of sequence reads mapped to the rCRS with mtDNA enrichment. When compared to no mtDNA enrichment, MDA averaged a 507.99% increase in sequence reads mapped to the rCRS.

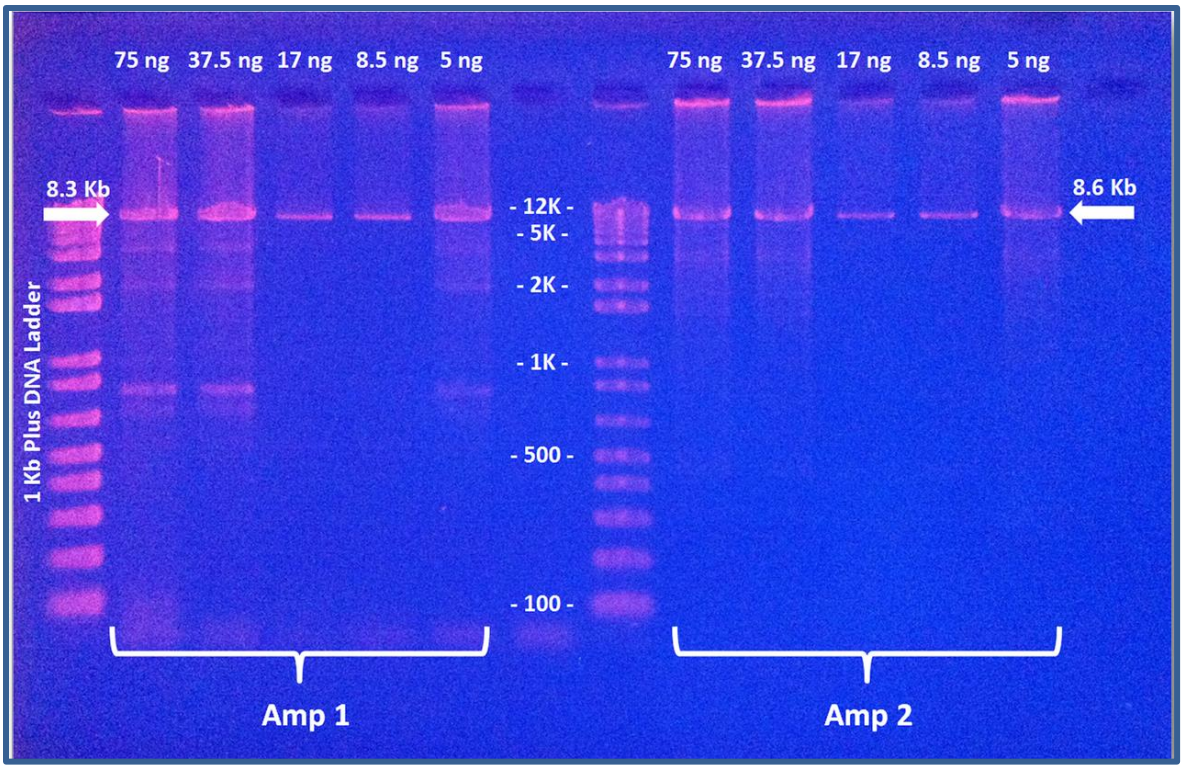


Figure 23: Yield gel of long-range PCR optimization. All five dilutions of mtDNA template generated detectable bands for both target amplicons at their expected sizes, 8.3kb and 8.6kb for Amp 1 and Amp 2, respectively, on a one percent agarose yield gel. Post-PCR product yield quantification revealed that 8.5 ng and 17 ng generated a single, distinct band for Amp 1 and Amp 2. The other template quantity amounts exhibited smearing and multiple bands which indicate non-specific or incomplete amplification. Therefore, 10 ng was chosen as an intermediate quantity for whole mtGenome enrichment.

Whole mtDNA sequence analysis was not performed for five samples due to either sample drop-out, *i.e.*, no data was obtained, or, a mixture was suspected. Despite two separate whole mtGenome enrichment attempts, PCR product yield for one NED sample was consistently low and resulted in no sequence data. It is possible that the concentration of this sample extract was more dilute than expected in which an insufficient quantity of template mtDNA would have been used in the whole mtGenome enrichment reactions. However, as no

other sample failed whole mtGenome enrichment, the event of a dilution error occurring during normalization for the DNA extract is unlikely. An alternative explanation is the presence of primer binding mutations which would hinder mtDNA amplification and result in poor PCR product yield. Whole mtGenome sequence data was also not obtained for one NC sample but not because of failed enrichment. On the contrary, whole mtGenome amplification for the NC sample was successful: 55.8 ng and 105 ng of PCR product were generated for Amp 1 and Amp 2, respectively, which is ample for downstream library preparation. The most likely explanation no sequence data was obtained for this sample is that a pipetting error could have occurred during library preparation (barcoding), normalization to 26 pM or pooling prior to emulsion PCR. In addition to sample drop-out, whole mtGenome sequence data of three NC samples were not analyzed because of suspected mixtures. One sample exhibited heteroplasmy at seven different positions which is exceeds the mean and three standard diviations observed of the entire study cohort (0.44 ± 0.89). The other two NC samples are suspected to be mixtures of each other; they exhibited nearly identical haplotypes (92%) and increased heteroplasmy.

Overall, whole mtGenome sequence data were obtained for 129 samples from the SABOR cohort; 42 in the NC group, 45 in the NED group and 42 in the BCR/MET group. One hundred percent coverage of the mtGenome was achieved for all 129 samples with an average read depth of 246X per base position. A total of 4,229 differences from the rCRS were observed in all 129 samples; 4,119 of these differences (97.40%) were ancestral/phylogenetic polymorphisms or common “hot spot” mutations. The remaining 109 mtDNA variants

(observed at 100 different positions) were designated as GPMs based on HaploGrep classification (Appendix B). These mtDNA variants are not present in Phylotree .[238] Seven GPMs were observed more than once in the SABOR cohort (Table 14). The most frequently observed GPMs were two deletions within the COI and ATP synthase F0 subunit 8 (ATP8) genes at positions 6,958 and 8,502, respectively. The deletion at position 6,958 was observed three times, once in all three groups, *i.e.*, NC, NED and BCR/MET. However, the 8,502 deletion was observed exclusively in PCa; once in the NED group and twice in the BCR/MET group. Due to the rarity of insertions and deletions (indels) within the coding region, the mapped read pileups were visually verified by manual inspection using NextGENe® Viewer (Figure 24).[265]

			Location			AA Change		
Variant	Frequency	Mutation	Region	Locus	AA	Before	After	Documented in MitoMap
149.1T	1.82%	Insertion	HV2	NA	NA	NA	NA	Yes
152d	1.82%	Deletion	HV2	NA	NA	NA	NA	Yes
573.1CCC	1.82%	Insertion	HV3	NA	NA	NA	NA	Yes
6,958d	2.73%	Frameshift	Coding	CO1	352 (2)	Gly	Ala	No
8,502d	2.73%	Frameshift	Coding	ATP8	46 (2)	Asn	Ile	No
8,764R	1.82%	Non-Syn	Coding	ATP6	80 (1)	Ala	Thr	Yes
8,994R	1.82%	Syn	Coding	ATP6	156 (3)	Leu	Leu	Yes

Not applicable (NA) and amino acid (AA)

Table 14: Most frequently observed mtDNA GPMs in SABOR cohort. Frequency was calculated by dividing the number of times the mtDNA variant was observed within the SABOR cohort by the total number of observed GPMs (N=109). The two most frequently observed mtDNA variants (observed three times) are also the only two of the seven that have not been previously reported in MitoMap. The other five GPMs were observed twice.

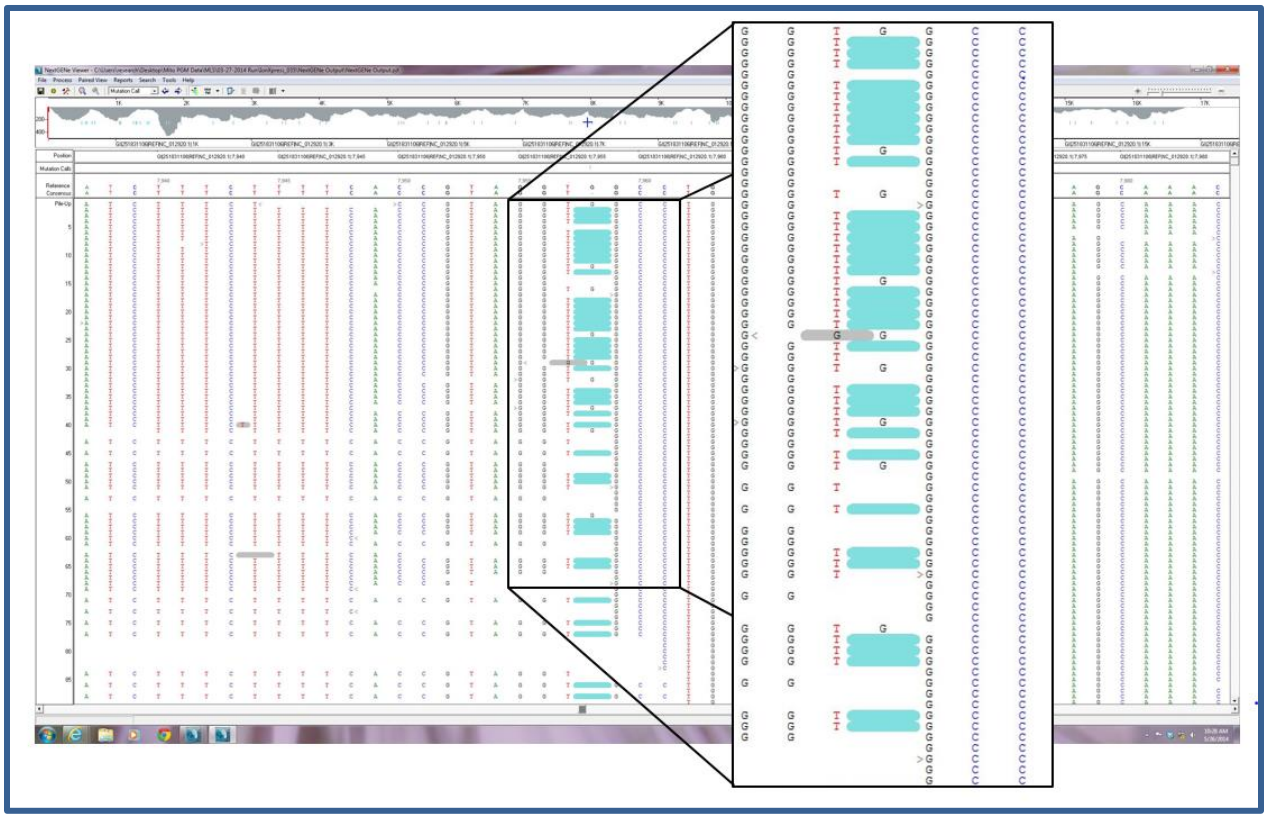


Figure 24: NextGENe® Viewer screen shot showing mtDNA deletion at position 6,958. Screenshot is of the mapped read pileup aligned to the rCRS in NextGENe® Viewer. Turquoise shading represents a called variant from the rCRS by the NextGENe® software. Cutaway is a zoomed-in section of the screenshot showing a closer view of the aligned reads at position 6,958 and the called mtDNA variant, a deletion. In this example, as well as the other two observed deletions at this position, the GPM is heteroplasmic. The minor species present at this position is the same as the rCRS, *i.e.*, guanine (G), whereas the major species is the deletion.

MitoMap is one of the largest and most comprehensive collections of reported polymorphisms and mutations in the human mtGenome. Therefore, all 100 GPMs were screened against the MitoMap database to identify novel or unreported mtDNA variants. In addition, each GPM was fully characterized and assessed for its pathological potential using PolyPhen-2 (Appendix C). Six unique GPMs were observed in the NC group which was the least

amount among all three groups: eight and seven were detected in the NED and BCR/MET groups, respectively. When stratified by the mtGenome locus, the highest number of GPMs occurred in the mitochondrially encoded cytochrome b (CYB) gene where a total of ten variants (nine previously reported and one unique) were detected (Figure 25). The second highest number of GPMs observed was shared by two genes: ND5 and COI. Each gene had nine GPMs detected where four observed in COI were classified as unique but ND5 had only one unique variant.

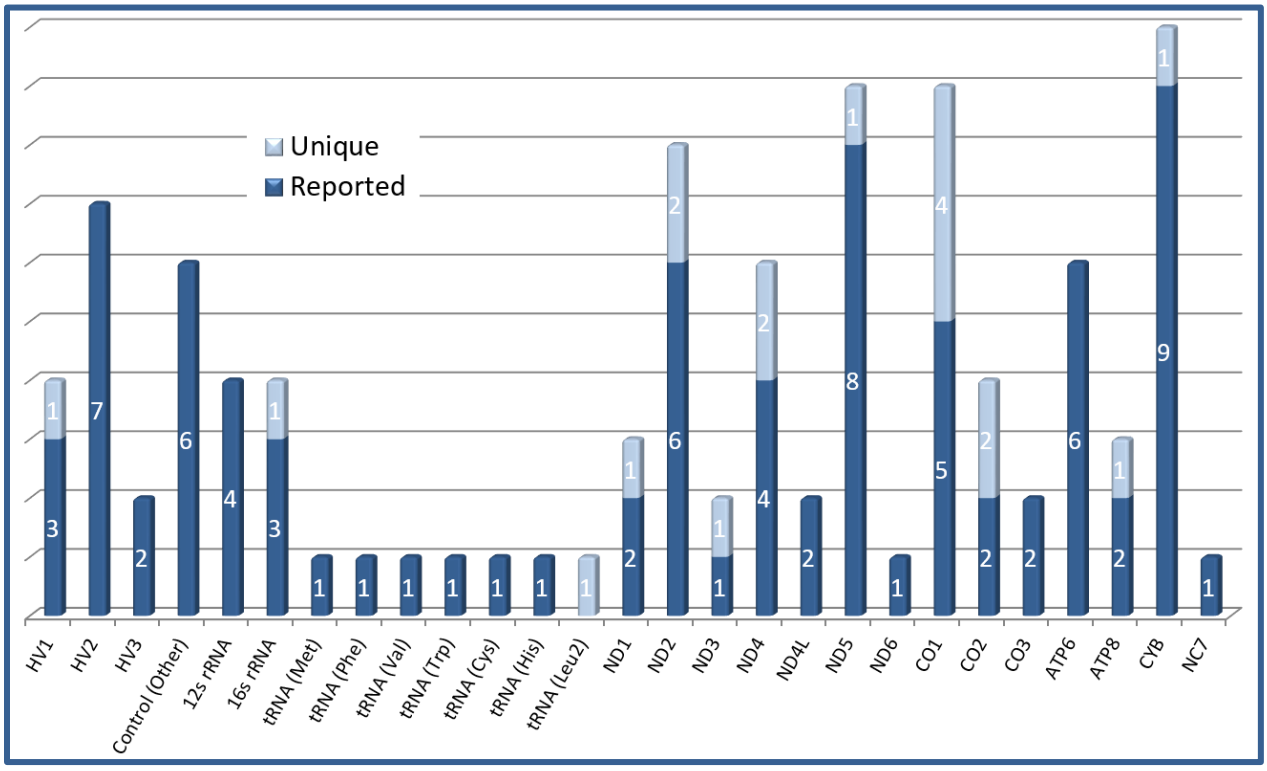


Figure 25: Number of observed GPMs by mtGenome locus. Dark blue bar represents the number of GPMs detected that have been previously reported in MitoMap whereas the light blue bar represents the number of unique GPMs, *i.e.*, those that have not yet been reported by MitoMap. Dark and light blue bars are stacked to represent the overall number of GPMs observed by mtGenome locus.

When more broadly stratified by functional group, Complex I had the highest number of GPMs (Figure 26). Moreover, Complex I is the largest enzyme of the ETC.[266] Seven of the 13 protein encoding genes by mtDNA are for Complex I which constitutes 6,356 bp (38.36%) of the mtGenome. Therefore, a ratio of the number of GPMs to the size (in bp) of each functional group was calculated to determine, proportionally, which group had the highest number of mtDNA variants detected. The effect of normalizing the number of GPMs to the size of each functional group was drastic. Prior to normalization, Complex I had the highest number of GPMs detected but proportionally had nearly the lowest number of GPMs (Figure 26 and 27). The control region, which constitutes 1,122 bp (6.77%) of the mtGenome, had a total of 22 mtDNA variants and, proportionally, had the highest number of GPMs detected among all the functional groups (Figure 27). As the largest non-coding region of the mtGenome, an excess of GPMs within the control region can be attributed to weaker functional constraints.[172,255] Additionally, an A→G transition was detected at position 8,280 which is located in a non-coding segment (NC7) of the mtGenome. Altogether, there are 10 non-coding segments (NC1-10) which account for only 58 total bases (0.35%) of the mtGenome. Therefore, the non-coding "functional" group proportionally had the second highest number of GPMs (Figure 27). ANOVA (p= 0.008) and Tukey's HSD post-hoc analysis revealed significant differences between the functional groups based on the ratio of their size and the number of GPMs observed. Specifically, the non-coding region was significantly different from three other functional groups: Complex I (p<0.05), Complex IV (p=0.003) and rRNA (p=0.026) (Figure 27).

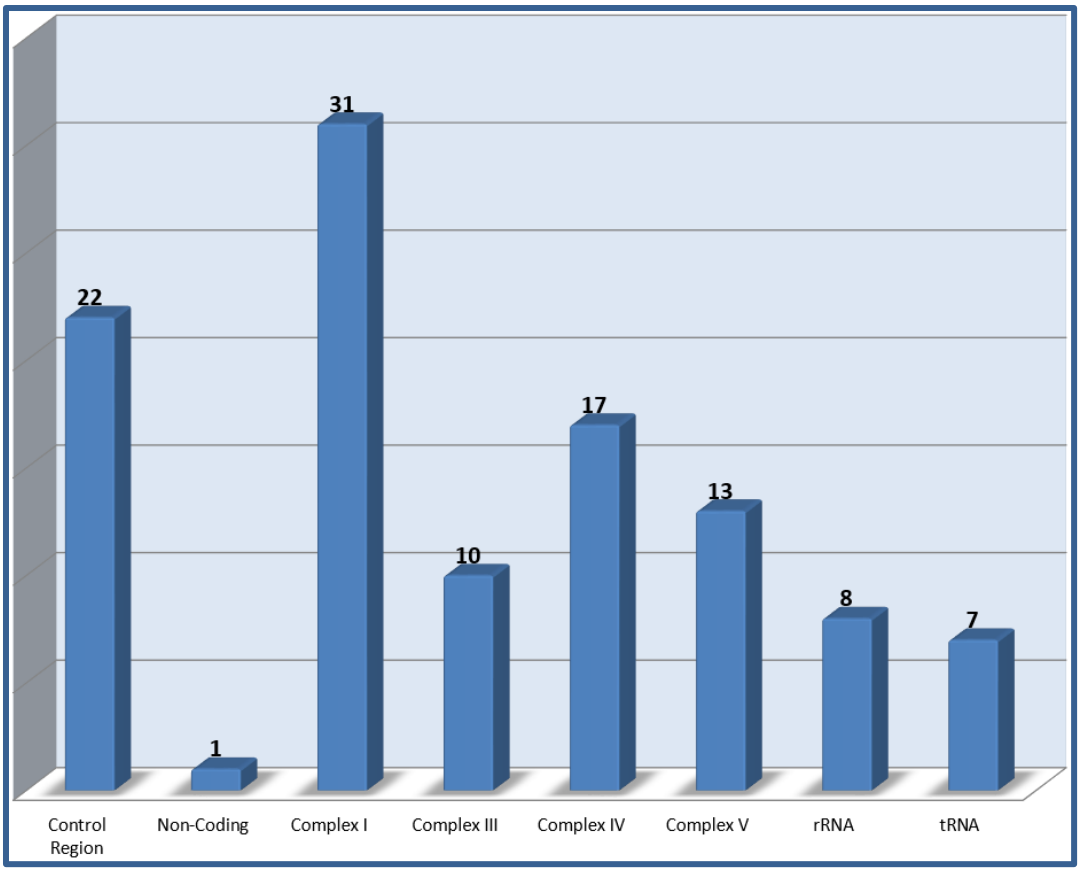


Figure 26: Total number of mtDNA GPMs observed in the SABOR cohort by functional group. Combined, HV1, HV2, HV3 and "Control Other" comprise the mtGenome control region. Complex I had the highest number of GPMs and the non-coding functional group had the least number of GPMs. The relative amount of GPMs observed in Complex I and the non-coding group are proportional to their size, 6,356 bp and 58 bp, respectively.

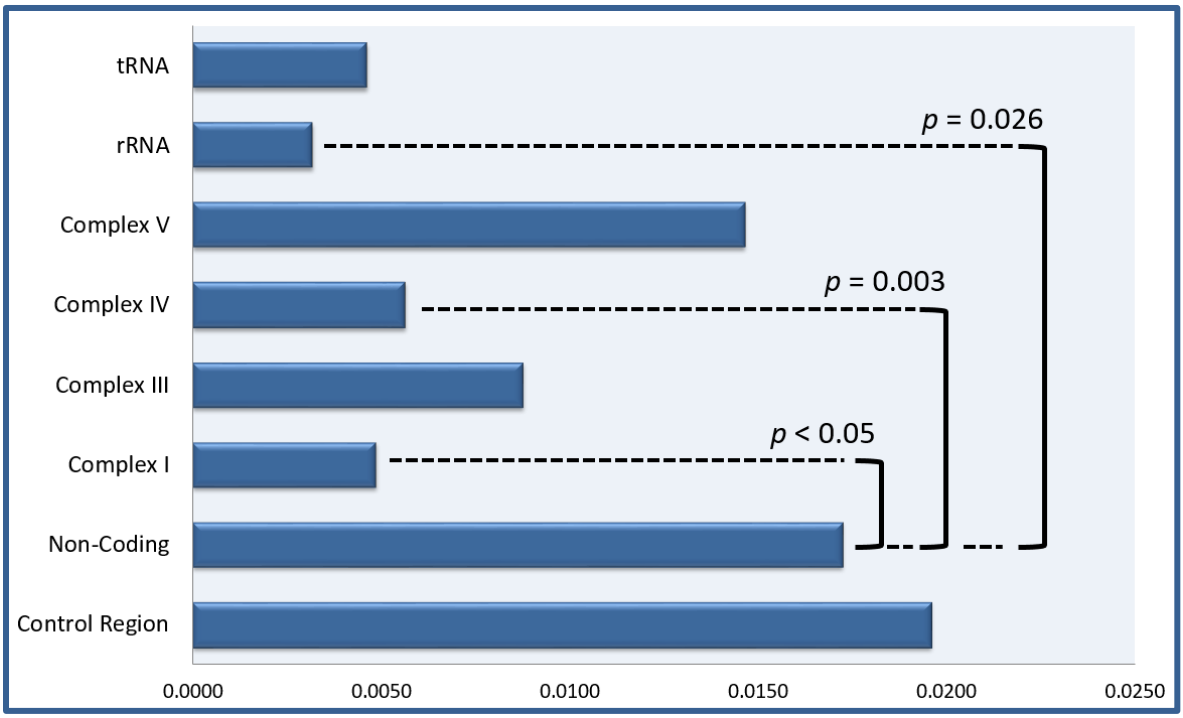


Figure 27: Ratio of the number of GPMs to functional group size (in bp). Proportionally, the control region had the highest number of detected GPMs but no significant difference was observed between it and any other functional group. Although only one GPM was observed in the non-coding region, proportionally it had the second highest number of GPMs detected and was significantly different from three other functional groups.

A total of 33, 33 and 43 GPMs were detected in the NC, NED and BCR/MET groups, respectively. Despite a 30.30% increase in the number of detected GPMs in the BCR/MET group no significant difference (p=0.3889) was observed between the three groups. However, when stratified based on functional group, there was a significant difference (p=0.029) observed within groups, *i.e.*, between NC, NED and BCR/MET, both with and without functional group size (bp) normalization. Post-hoc analysis identified a significant difference between disease status groups in coenzyme Q-cytochrome c reductase (Complex III); the BCR/MET group has a

significantly greater number of GPMs than NC and NED (Figure 28). Furthermore, a Pearson test showed a significant positive (r=0.216, p=0.014) correlation between disease status and the number of GPMs in Complex III genes.

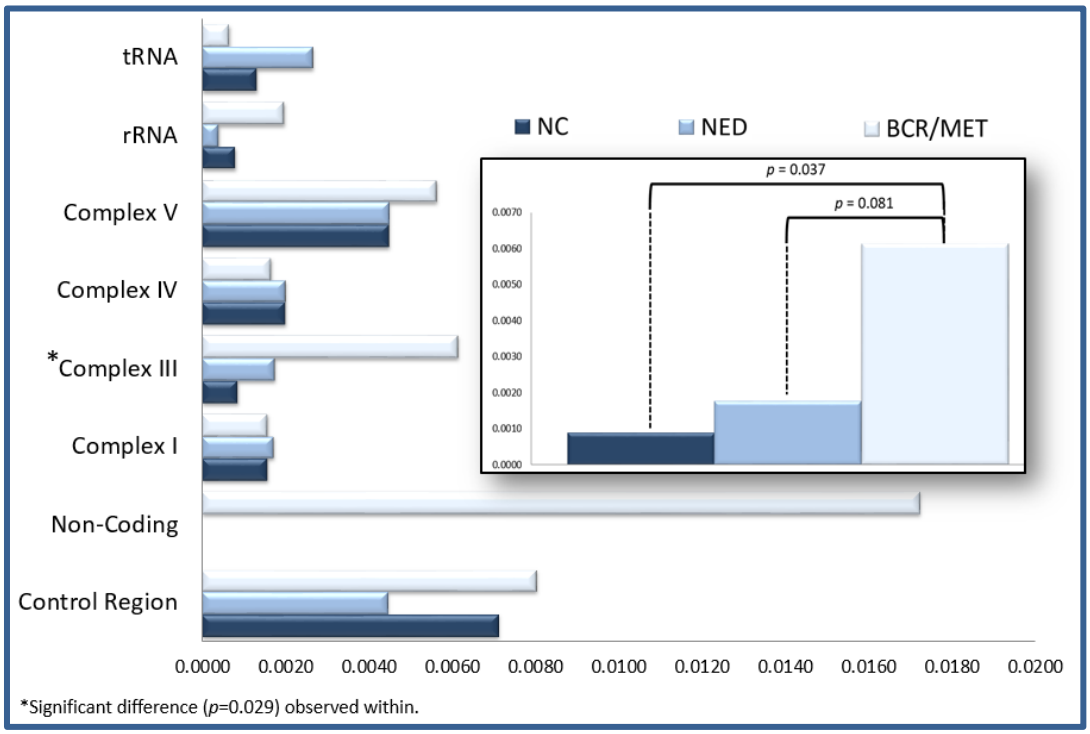


Figure 28: Number of GPMs by functional group and disease status. Horizontal bars represent the ratio of GPMs to the functional group size (in bp). Inset shows the difference in GPM ratio observed in Complex III between test groups: the BCR/MET group is significantly greater than the NC and NED groups based on Tukey's HSD post-hoc analysis.

An exhaustive characterization was performed on all 100 GPMs to assess their potential of having an effect on the respective translated protein. A PolyPhen-2 score was obtained for 28 missense mutations of which 16 were determined to be benign, three possibly damaging, and eight probably damaging. Although PolyPhen-2 cannot assess the pathological potential of

deletions, an indel in the middle of an open reading frame (ORF) will cause a frameshift and translation of an abnormal protein. As such, seven coding region deletions were determined likely to be damaging in addition to the 12 missense mutations identified by PolyPhen-2 to be possibly or probably damaging. One additional GPM was not evaluated by PolyPhen-2 because of the special nature of its mutational effect. A heteroplasmic transition was observed within the COI gene, at position 6,573, which causes a premature translational termination.[267] The observed GPM results in an “AGA” codon which causes a ribosomal -1 frameshift during translation. As a consequence, the upstream frameshift by the mitoribosome encounters an "AGG" codon which again, does not code for an amino acid but instead, causes a -1 frameshift. Presumably, the two subsequent -1 frameshifts cause the mitoribosome to stall and fall off the mRNA transcript. Nevertheless, a total of four, seven and eight GPMs likely to have a functional effect were observed in NC, NED, and BCR/MET, respectively. Comparatively, the number of GPMs in the PCa groups was higher than the NC group. However, neither a significant difference of their means (p=0.507) or correlation (r=0.102, p=0.249) was observed (Figure 29).

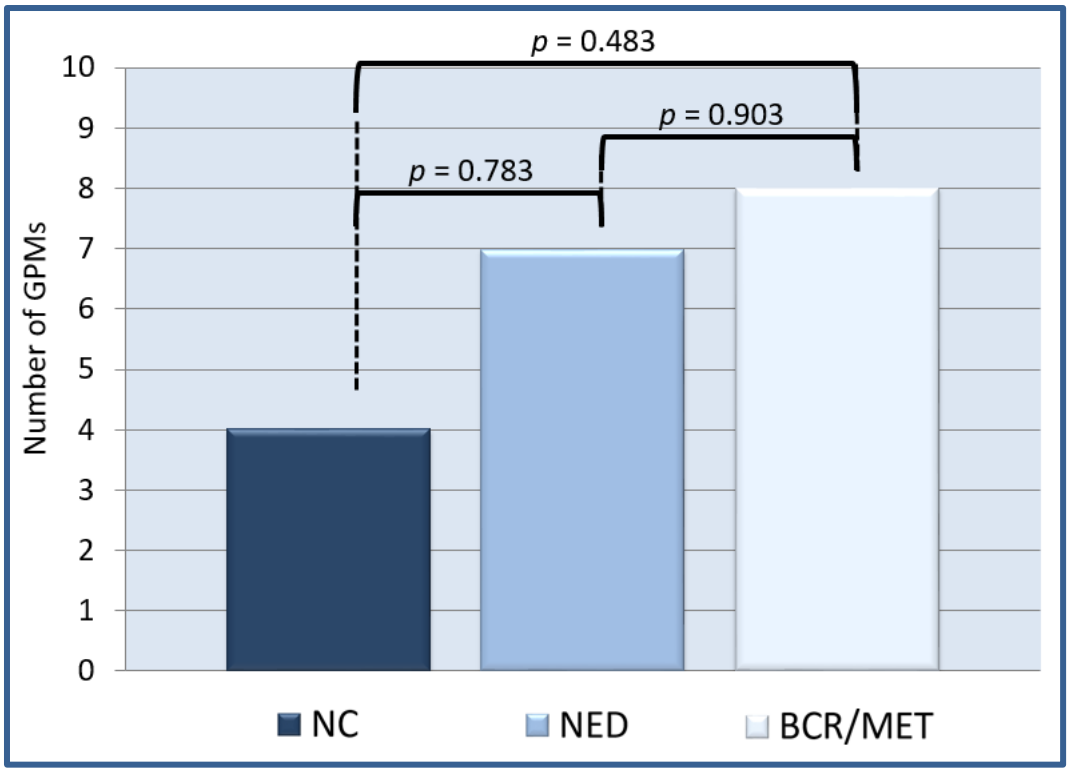


Figure 29: Number of GPMs predicted to have a functional effect in each SABOR group. The BCR/MET group exhibited the highest number of GPMs likely to have a functional affect whereas the NC group had the least number. However, no significant difference was observed between groups.

Conclusion and Discussion

Mitochondrial involvement in tumorigenesis has long been hypothesized when Warburg observed that cancer cells undergo a bioenergetic switch from OXPHOS to glycolysis for ATP production despite sufficient amounts of oxygen.[195] Additionally, its role in programmed cell death and cell proliferation further implicate mitochondria in cancer development.[136,268,269] Even before DNA sequencing technology was available, abnormal circular dimer and catenated forms of mtDNA were observed in leukemic myeloid cells using electron microscopy.[270,271] The

earliest genetic evidence that mitochondrial dysfunction is involved with cancer development came from the observation that mutations in the tricarboxylic acid cycle enzymes succinate dehydrogenase (SDH) and fumarate hydratase (FH) promote tumorigenesis by stabilizing the hypoxia-inducible factor 1-alpha (HIF1α) under normoxic conditions in heredity forms of paragangliomas (HPGL) and leiomyomatosis and renal cell cancer (HLRCC).[272–274] Subsequently, mutations in mtDNA have been reported in various types of cancer, *e.g.*, breast, endometrial, ovarian, leukemia, renal, colon, hepatocellular, glioblastoma, thyroid and prostate.[183,275] However, the role of these mtDNA mutations in tumorigenesis remains largely unclear. The most common type of somatic mtDNA mutations are benign.[130,276,277] It has been estimated that 75% of cancer-related mtDNA mutations that occur in the coding region are synonymous and have no effect on tumorigenesis.[277] Results of this study agree with that estimate. Indeed, 17.43% (19/109) of the detected mtDNA variants were characterized as having a functional effect and the corresponding 82.57% were characterized as synonymous, non-consequential mtDNA variants.

Complex I plays a central role in OXPHOS as it is the point of ingress for most electrons into the ETC.[278] It is also the main sources of ROS during normal metabolism.[279–282] Reactive oxygen species are generated at the flavin mononucleotide group (FMN) which is located on the matrix side of the inner membrane.[283] Production of superoxide (O_2^-) by Complex I occurs by two mechanisms: when the NADH/NAD^+ ratio is high, FMN becomes fully reduced and a backup of electrons react with molecular oxygen (O_2).[284,285] The second mechanism involves

reverse electron transport where ubiquinol (reduced form of coenzyme Q) transfers electrons back to Complex I.[283,286–288] However, the superoxide producing site involved in the second mechanism is unclear. It is possible that it occurs at the FMN site but some suggest it may occur at the ubiquinone binding site itself.[289] Nevertheless, inhibition of the respiratory chain by damaged or mutated protein subunits can lead to superoxide production. At first inspection, the whole mtGenome sequencing results presented here suggest a higher prevalence of GPMs occur in Complex I genes. Indeed, 31 GPMs were detected in the SABOR cohort of which 16.12% (five out of 31 GPMs) are predicted to have a functional effect. With regards to overall mutational load, these findings present a compelling case as to the role of Complex I contributing to a respiratory deficiency that may be related to a pathological state. However, the number of GPMs detected in the NC, NED and BCR/MET groups is nearly identical and no correlation was observed (r=0.000, p=1.000). Similarly, 17 GPMs were detected in Complex IV making it the functional group with the second highest mutational load. Also, 52.94% (nine out of 17 GPMs) are determined to likely have a functional effect which is the most of any functional group in this study. However, the distribution of GPMs among the disease status groups was nearly the same and again, no correlation was observed (r=-0.027, p=0.764).

Most compelling was the significant difference in the number of GPMs observed in Complex III. Proportionally, the CYB gene occupies very little real estate within the mtGenome as compared to the other ETC subunits. As such, the probability of random somatic mutational events occurring in Complex III genes is lower than all the other functional groups, *e.g.*,

Complex I, Complex IV, tRNAs, and rRNAs. The only exception is ATP synthase which constitutes only 888 bp (5.36%) of the mtGenome. Whether or not the higher frequency of GPMs in Complex III is a function of PCa progression requires further research. In part, the Complex III sequencing results is consistent with the hypothesis, *i.e.*, a higher number of mtDNA variants are present in the PCa groups than the NC group. Furthermore, the significant positive correlation suggests a linear relationship exists between PCa disease progression and mutational load. Further strengthening the role of mitochondrial dysfunction associated with PCa progression is the fact that the only Complex III GPMs predicted to have a functional effect where detected exclusively in the BCR/MET group. However, to gain further insight into the functional significance of these mtDNA mutations, further analysis is needed. Bioinformatic protein modeling is a powerful tool that can be used to make informed predictions regarding the functional significance of missense mutations. Chapter V describes the use of *in silico* 3D modeling in order to compare conformational energy states between wild-type (WT) and mutant proteins based on the mtGenome sequencing results presented here.

Chapter IV Results at a Glance

1) Homoplasmy between prostate tissue and bodily fluids of PCa patients supports the use of peripheral blood to screen for mtDNA GPMs associated with PCa.
2) Men with PCa exhibit a higher mutational load and a higher number of pathogenic mtDNA variants.
3) Unique variants, not currently indexed in Mitomap, were observed in this study; more unique variants were observed in the PCa group.
4) A significant difference (p=0.029) and a positive correlation (r=0.216, p=0.014) were observed between the number of GPMs in Complex III genes and disease status.

Chapter V

Protein Modeling and Discussion

The 3D structure of a protein is determined by its primary structure where non-covalent interactions are the prevailing forces that direct secondary and tertiary structure formation. Knowing a protein's 3D structure can infer its molecular function.[290] Although structural biology techniques can identify a protein's 3D structure, they are labor intensive, challenging and costly, especially for membrane bound proteins. Further, predicting a protein's 3D structure from its linear amino acid sequence is challenging and requires large computational resources to perform highly sophisticated probabilistic calculations. A less challenging and more accurate technique for predicting a protein's 3D structure uses comparative modeling. Building a protein by homology consists of using a known 3D structure of a protein as a template. This technique is based on the premise that the protein's tertiary structure is better conserved than its primary structure whereby evolutionarily distant but related sequences can adopt very similar structures.[291] Thus, proteins that have dissimilar amino acid sequences but have similar biochemical function and belong to the same protein family, will have similar structural properties and overall conformation.

Comparative modeling can provide a useful structural framework for generating hypotheses about a protein's function. In this study, *in silico* 3D modeling of ND5 was performed to compare the conformational energy between a WT protein and a mutant protein. The purpose of this experiment was to determine if a single missense mutation negatively affects the structural energy state of the respective protein. A software program was used to manually build the ND5 protein by homology. An atomic-resolution model of the human ND5 protein was constructed from its amino acid sequence and an experimental 3D structure based on a related homologous protein. The completed WT protein was then mutated at the specific amino acid residue based on its predicted pathogenicity from a previous PCa study (discussed in Chapter IV).

Materials and Methods

Template Selection and Sequence Alignment

A standard Protein BLAST (http://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Proteins) query identified *T. thermophilus* Complex I, subunit 12, chain L (606 amino acid residues) as the closest homolog, based on an identity score of 42%, of ND5 (603 amino acid residues) with a known 3D crystallographic structure (Appendix D). This structure was used as a template to build a comparative model of the human ND5 protein.[292] The Protein Data Bank (PDB) file format provides a standard representation for macromolecular structure data and is commonly used by many bioinformatic programs.[293] The PDB file (PDB ID: 4HE8) contains the

crystallographic structure of subunit 12, chain L and was downloaded from the Research Collaboratory for Structural Bioinformatics (RCSB) Protein Data Bank at www.rcsb.org.[294]

The bioinformatics tool EMBOSS Needle (http://www.ebi.ac.uk/Tools/psa/emboss_needle/) was used to perform a global, pairwise sequence alignment between the homologous proteins NADH-quinone oxidoreductase subunit 12, chain L (UniProtKB/Swiss-Prot: Q56227.1) and ND5 (GenBank: ADB78261.1) (Appendix E).[239,295–297] The EMBOSS Needle tool uses the Needleman-Wunsch alignment algorithm to optimally align two sequences along their entire length and then writes the optimal global sequence alignment to a file. Protein sequence alignment was performed online using the European Molecular Biology Laboratory-European Bioinformatics Institute (EMBL-EBI) web server (http://www.ebi.ac.uk/Tools/psa/emboss_needle/).

Importing and Cleaning Protein Crystal Structure

The comparative model of ND5 was constructed using the SCiGRESS Explorer v.7.7.0.47 software package (Fujitsu Ltd., Tokyo, Japan). The crystal structure of 4HE8 contains seven subunits and 14 chains (Figure 30.A). Subunits 7, 8, 10, 11, 12, 13 and 14 contain chains A and B; C and H; D and J; E and K; F and L; G and M; and I and N, respectively. Chain L of the 4HE8 template was isolated by deleting all other chains as well as a single molecule of undecyl-maltoside (UMQ) (Figure 30.B). UMQ is a non-ionic detergent used for extracting and solubilizing membrane-bound proteins while maintaining its native conformation. After

isolating chain L, its valence and hydrogen bonding were subsequently checked. Hydrogen atoms, electrons, and hybridization were added to the protein and water molecules to achieve the correct ionization and tautomeric states of the amino acid residues. Further, hydrogen atoms in hydroxyl groups and water molecules were rotated to maximize hydrogen bonding. An optimization experiment was performed using the augmented molecular mechanics force field geometry parameter (known as MM3) in order to optimize the energy of each molecule to its lowest stable energy state. This energy minimization process was performed until the energy change was less than 0.001 kcal/mol or the molecules had been updated 300 times.

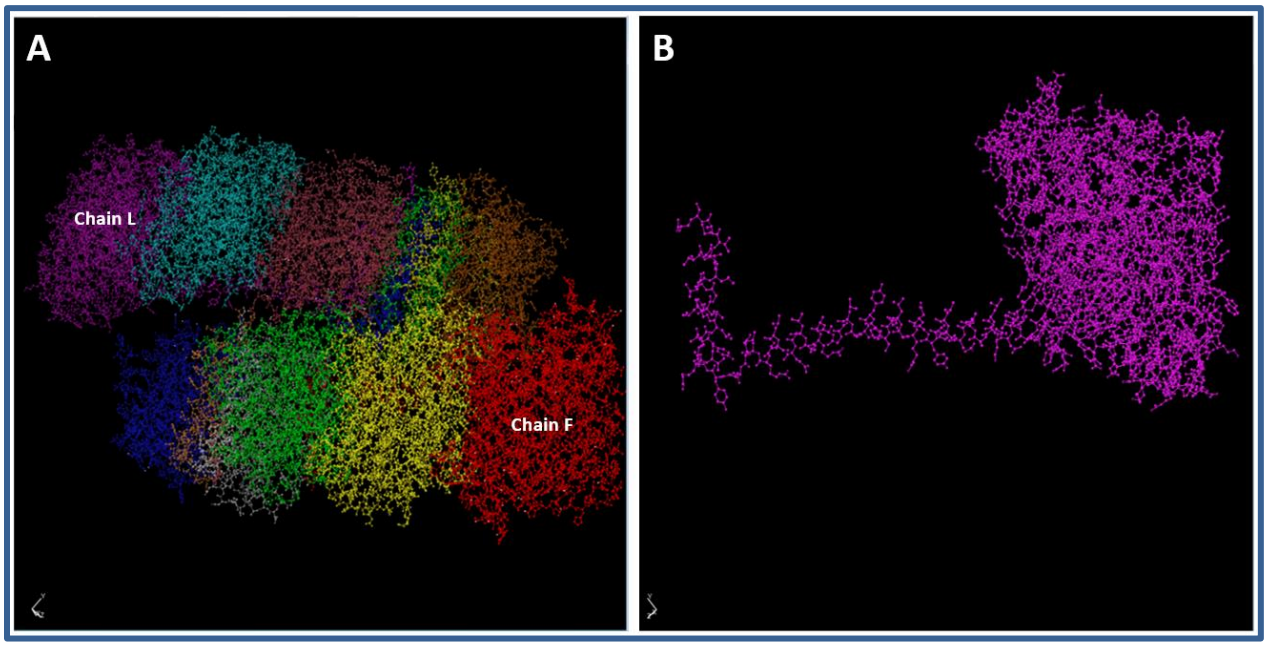


Figure 30: Isolation and clean-up of NADH-quinone oxidoreductase subunit 12, chain L. A) 3D wireframe model of 4HE8 before chain L isolation and cleaning. All 14 chains are present and differentiated by color; subunit 12 consists of chains F and L. B) 3D wireframe model of the isolated and cleaned subunit 12, chain L of 4HE8, rotated 180 degrees. Both 3D models are viewed in SCiGRESS.

Building by Homology

Starting from the 4HE8 chain L (4HE8L) amino acid sequence, polypeptides were deleted or inserted based on the EMBOSS Needle alignment using SCiGRESS Sequence View. Then, in an N-terminus to C-terminus direction, each divergent amino acid of 4HE8L was changed to that of the human ND5 primary structure. After all 603 residues were verified, their corresponding PDB numbers were resequenced, to correct for the disorder caused by deleting and inserting amino acid residues, and secondary structure was reanalyzed. Two optimization experiments were performed with the MM3 geometry parameter (using steepest descent and conjugate gradient options) in order to optimize the energy of each molecule to its lowest stable energy state. The preliminary ND5 3D model was refined by performing a sequential series of paired molecular dynamics and geometry optimization experiments after unlocking grouped peptides belonging to the secondary structures. After the WT ND5 ($ND5_{WT}$) protein was completed, a single amino acid residue was mutated to create the mutant protein ($ND5_{M}$). Based on the PMut results discussed in Chapter IV, amino acid residue number 126 was changed from isoleucine to serine (p. Ile126Ser) (Figure 31). Subsequent geometry optimization experiments were conducted afterwards to achieve the lowest, most stable energy state of the $ND5_{M}$ protein. Finally, structural conformation experiments were conducted to assess the energy status of the $ND5_{WT}$ and $ND5_{M}$ 3D models for stability comparison.

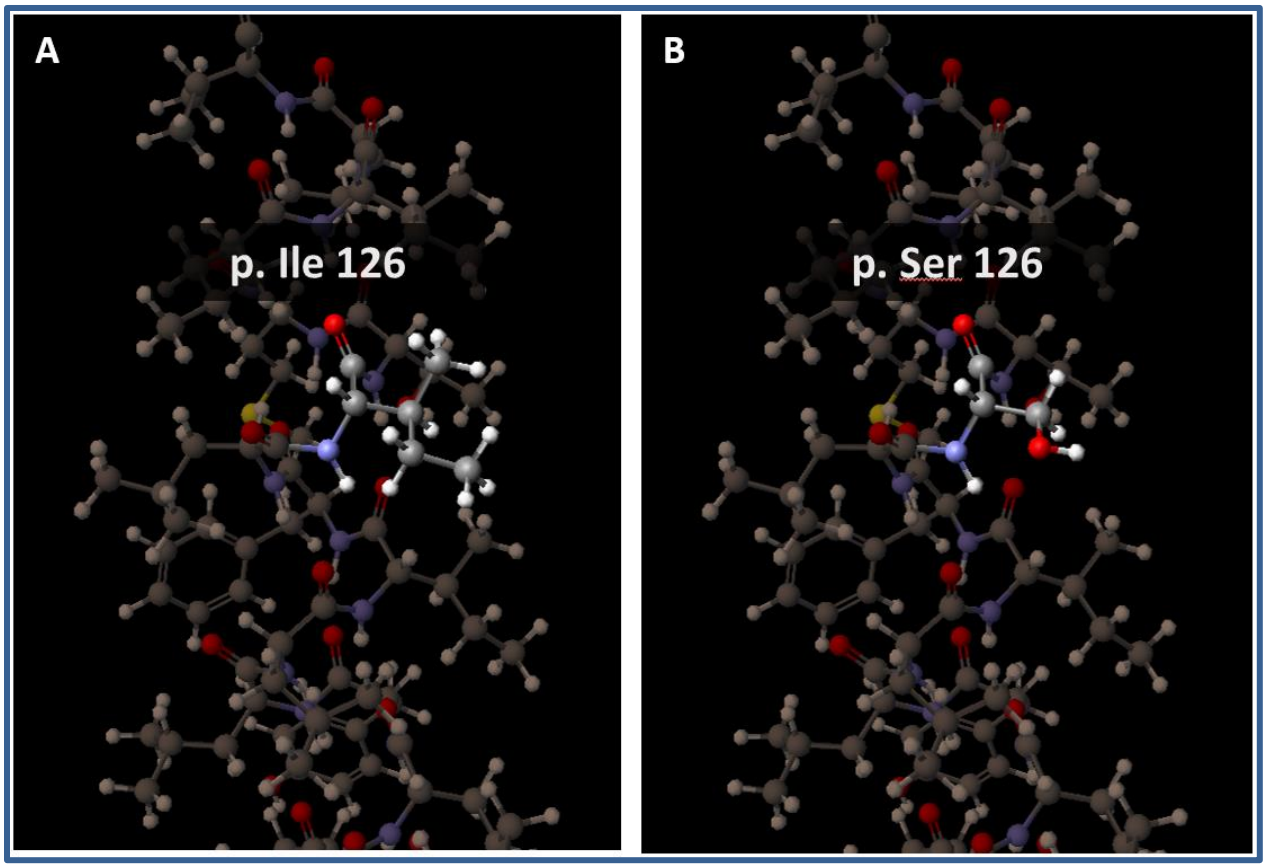


Figure 31: Close up view of a 3D ball and stick model of ND5. Amino acid residue number 126 of ND5 was changed from isoleucine to serine (p. Ile126Ser) to assess its energy state effect. Spatial orientation of both amino acid residues are identical save for their respective functional groups. Molecules of the highlighted amino acids are color-coded as follows: carbon (gray), nitrogen (purple), oxygen (red), and hydrogen (white). A) $ND5_{WT}$ with an isoleucine residue at position 126 has a bulky and strongly hydrophobic functional group. B) $ND5_{M}$ with a serine residue at position 126 has a less bulky and smaller, hydrophilic functional group.

Results

EMBOSS Needle alignment revealed that 4HE8L and ND5 have 31.7% identity and 45.6% similarity. This required insertion of three polypeptides, deletion of four polypeptides and modification of 434 (71.97%) amino acid residues to achieve the primary structure of $ND5_{WT}$. Refining the precursory 3D structure of $ND5_{WT}$ required a total of 20 sequential paired molecular dynamics and geometry optimization experiments. Prior to refinement, the energy state of $ND5_{WT}$ was calculated to be 273,238,338.5368 kcal/mol. After more than 12 days of accumulated computational time, molecular dynamic and geometry optimization reduced the $ND5_{WT}$ 3D model energy state by more than a 40 thousand fold decrease to 6,355.2993 kcal/mol. The final 3D structure for $ND5_{WT}$ is shown in Figure 32.

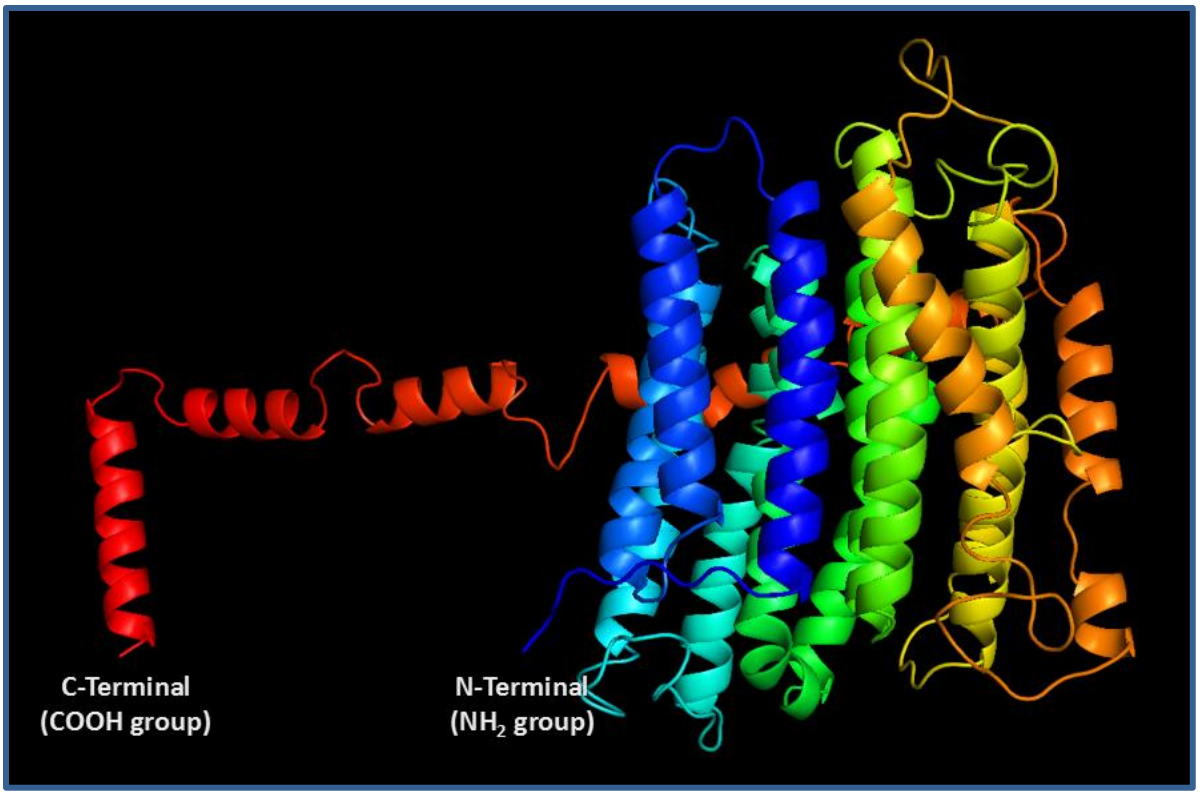


Figure 32: Finalized 3D structure of $ND5_{WT}$. A 3D ribbon model of ND5 based on a homology build using SCiGRESS and viewed in PyMOL (PyMOL Molecular Graphics System, v.1.7.4, Schrödinger, LLC). The $ND5_{WT}$ protein consists of 16 transmembrane helices, shown here in a color gradient, and two key domains: group of helices near the N-terminus and the L-shaped helices near the C-terminus.

Physics-based energy calculations were used to assess the stability of the homology models $ND5_{WT}$ and $ND5_{M}$. This technique calculates the interatomic forces that physically maintain a protein's stability such as van der Waals and electrostatic interactions. Table 15 lists the ten different energies which total the overall structural energy state of the proteins. An energy state difference of 0.56% was calculated between $ND5_{WT}$.and $ND5_{M}$. The largest contributors in the observed difference were from two steric energies: bend bend and

electrostatics. Bend bend energy is from bond angles that are "bent" differently than is optimal. The bend bend energy of $ND5_M$ is greater and therefore less stable than $ND5_{WT}$. Similarly, the electrostatics energy is higher in $ND5_M$ thereby suggesting less symmetrical interaction between bond dipoles in its final conformation. In addition, the $ND5_M$ protein exhibits less stable energy for the stretch bend, dihedral, torsion stretch and hydrogen bond energies. Overall, six of the ten calculated energies for $ND5_M$ are less stable than $ND5_{WT}$ which suggests that the Ile126Ser mutation has a negative steric energy effect. Therefore, the structural conformation of $ND5_M$ is less stable than $ND5_{WT}$ and may have a great impact on the protein's function.

Energies	$ND5_{WT}$	$ND5_M$	Difference
stretch	697.743	677.470	2.91%
stretch bend	57.219	57.576	0.62%
improp torsion	106.401	104.955	1.36%
bend bend	46.840	53.026	11.67%
electrostatics	1217.784	1266.390	3.84%
angle	2276.014	2255.436	0.90%
dihedral	1668.993	1684.819	0.94%
torsion stretch	-105.723	-103.698	1.92%
van der Waals	2427.356	2400.875	1.09%
hydrogen bond	-2037.327	-2006.053	1.54%
Structural Energy	**6355.299**	**6390.795**	**0.56%**

Table 15: Differences in calculated energies between $ND5_{WT}$ and $ND5_M$. Molecular mechanics calculations, using MM3 geometry parameters, show that the overall structural energy of $ND5_{WT}$ is lower (more stable) than $ND5_M$ by 0.56%. The largest contributors to this difference were bend bend and electrostatics energies.

Conclusion and Discussion

The most recent proposed mechanism of coupling electron transfer with proton translocation suggests ND5 is a critical component of Complex I's respiratory function.[292] The ND5 protein is located within the membrane domain of Complex I and exhibits two key structural components which play a critical role in its function. First, 14 transmembrane helices form a single proton antiporter. A number of charged residues, including some in helix number five, within this domain are integral for proton exchange during translocation from the mitochondrial matrix to the intermembrane space. Additionally, a long amphipathic helix, which extends along the C-terminus of ND5, interacts with several other subunits in the proton transporter region.[298] The function of the long C-terminal helix mechanically links the subunits responsible for proton translocation. Specifically, it coordinates a cascade of conformational changes within the transmembrane helices that form the half channels of the proton transporters. This mechanism facilitates a global conformational switch that is instrumental for proton translocation.[299,300] Taken together, these structural characteristics of ND5 demonstrates its critical role in the respiratory function of Complex I and, globally, the entire ETC. Therefore, any perturbation of ND5's structural conformation in sufficient quantity will have a profound effect and will cause a respiratory deficiency.

These results demonstrate that a single missense mutation in the mitochondrial ND5 gene will have a measurable negative effect on the structural energy state of the translated ND5 protein. The consequence of a 35.5 kcal/mol energy difference is dependent upon the

affected amino acid residue relative to the protein's secondary structure, its tertiary structure as an isolated subunit and its quaternary structure as a component in Complex I (personal communication with Dr. Laslo Prokai, Fall 2014). In context, having higher bend bend and electrostatics energies will affect the proper folding of the protein into its native structure in the mitochondrial matrix. Similarly, as a subunit of a very large protein complex, a single residue change negatively affects its interactions with adjacent subunits in Complex I. This is especially true for ND5 considering its structural relationship with the entire membrane domain of Complex I. The p. Ile126Ser mutation is located in the fifth transmembrane helix of ND5 (Figure 19). As an integral part of the half-channel responsible for proton translocation, this mutation interferes with the proper exchange of protons between helix five and the other polar and charged residues which constitute the antiporter channel. Further, the coupling mechanism of Complex I is dependent upon a continuous hydrophilic axis that spans the entire length of the membrane domain. Polar and charged residues of helix five are also part of this axis.[292] Therefore, the structural change caused by a p. Ile126Ser mutation will disrupt the continuity of this axis and have a global effect on the coupling mechanism of Complex I. As a consequence, the respiratory function of ND5 will be impaired and cause an energetic deficiency. Moreover, the reduced functional efficiency of ND5 causes increased oxidative stress due to reverse electron transport from ubiquinol.[283,286–288] However, analysis of the structural energies between ND5 and its interacting subunits can provide an even more detailed characterization of the potential effect of an amino acid substitution on Complex I.

In silico Protein–protein interaction, protein-protein docking and protein-ligand docking experiments should be performed to better understand the functional consequence(s) of mtDNA missense mutations. For example, a recent study was performed to investigate the functional effect of mtDNA mutations in Complexes III and IV associated with glioblastoma using 3D structural mapping.[301] This approach revealed new mechanistic insights of the functional consequences resulting from single point mutations. Specifically, a 15,500G>A mutation causes a reduction in Complex III's ion transfer rate which compromises the enzymes respiratory activity. Using the same 3D structural analysis to further investigate GPMs is a future direction that should be explored.

A manual approach was used to construct a 3D model of ND5 by homology. While the manual approach to protein modeling is still in use today, more automated procedures are available. These procedures apply different methods. The fragment assembly method constructs a complete protein model from conserved fragments of known crystallographic structures beginning with a conserved core. The segment matching method divides the target protein's primary structure and aligns it in segments, as opposed to the entire protein. Nevertheless, online open-source bioinformatic tools such as SWISS-MODEL are fully automated protein structure modeling tools that offer tremendous advantages over manual modeling software.[302] These homology-modeling servers use MM3 and other molecular mechanic parameters for geometry optimization. Among the added benefits of automated modeling are increased accuracy and time reduction. Online server based systems can achieve

increased accuracy of the resultant models through continuous surveillance by systems such as CAMEO (http://www.cameo3d.org/).[303] The manual modeling approach required more than 12 days of computational time whereas an automated approach would have been more expedient. A future direction of this study should include protein modeling for each mtDNA variant predicted to be pathogenic. Using automated modeling tools such as SWISS-MODEL to perform structural modeling on a larger scale is much more feasible than the manual approach. Furthermore, the structural energy data of multiple pathogenic mutations would provide much more evidence to support the role of mitochondrial dysfunction in PCa.

Chapter V Results at a Glance

1) Two 3D models of the human ND5 protein were constructed using a comparative modeling technique: $ND5_{WT}$ and $ND5_{M}$.
2) Energy calculations show that a single mtDNA missense mutation in the ND5 gene has a negative effect on the protein's structural integrity.
3) An increase in ND5's structural energy from a single missense mutation is predicted to cause a respiratory deficiency and increase oxidative stress.
4) To fully understand the pathogenic effect of GPMs protein structural analysis is recommended for genetic studies.

Summary

The core objective of this project was to prove that differences in mitochondrial genetics between different disease states of PCa are measurable in blood as well as establish a basis of using mtDNA as a molecular marker for disease progression. The overall analysis pipeline used to achieve these goals was a comprehensive strategy that included multiple aspects of the mtGenome. Part of this strategy involved whole mtGenome sequencing coupled with 3D structural modeling. The benefit of appending genetics with proteomics is the ability to simulate the natural manifestation of genetic abnormalities to their predicted molecular pathogenesis. This approach has not been performed to the extent presented here pertaining to mitochondrial dysfunction associated with PCa progression. Although the protein modeling study involved only one protein subunit and one amino acid change, the combined results of the pathological prediction from PMut (Chapter IV) and the observed structural energy analysis using SCiGRESS support the premise that a single mtDNA mutation can contribute to mitochondrial dysfunction.

The overarching hypothesis aimed at demonstrating that an observable difference in measures of mitochondrial dysfunction can delineate PCa progression. Within a biological context, results from the qPCR study discussed in Chapter III, the whole mtGenome sequencing in Chapter IV, and the bioinformatic tools used to predict and measure the functional consequence of observed GPMs form a narrative that adds to the limited body of knowledge regarding mitochondrial dysfunction associated with PCa progression. The translational impact

of the work presented here is only incremental as the level of significance between the study groups was not quite powerful enough to predict PCa progression. The primary goal then of any future studies is to expand on the work presented here in a manner that achieves more discernable differences between the NC, NED and BCR/MET groups. Doing so has the potential of establishing mtDNA as a tool for modeling disease progression and the ability to predict PCa cases likely to advance to an aggressive stage.

Figure 33 is a shematic representation of the cumulative findings of this project. In line with the projects goals, differences in the type and amount of mtDNA changes among men with different developmental stages of PCa were observed. Specifically, men with advanced stages of PCa exhibited an increase in $mtDNA_{CN}$, a decrease in $mtDNA_{DR}$ and a higher mutational load. Further, these differences were measured in blood and therefore establish a basis for for using mtDNA as a molecular marker for disease prognosis. Central to the proposed mechanism of mitochondrial dysfunction and resultant mtDNA markers found in blood associated with PCa progression is mitochondrial respiration (Figure 33).

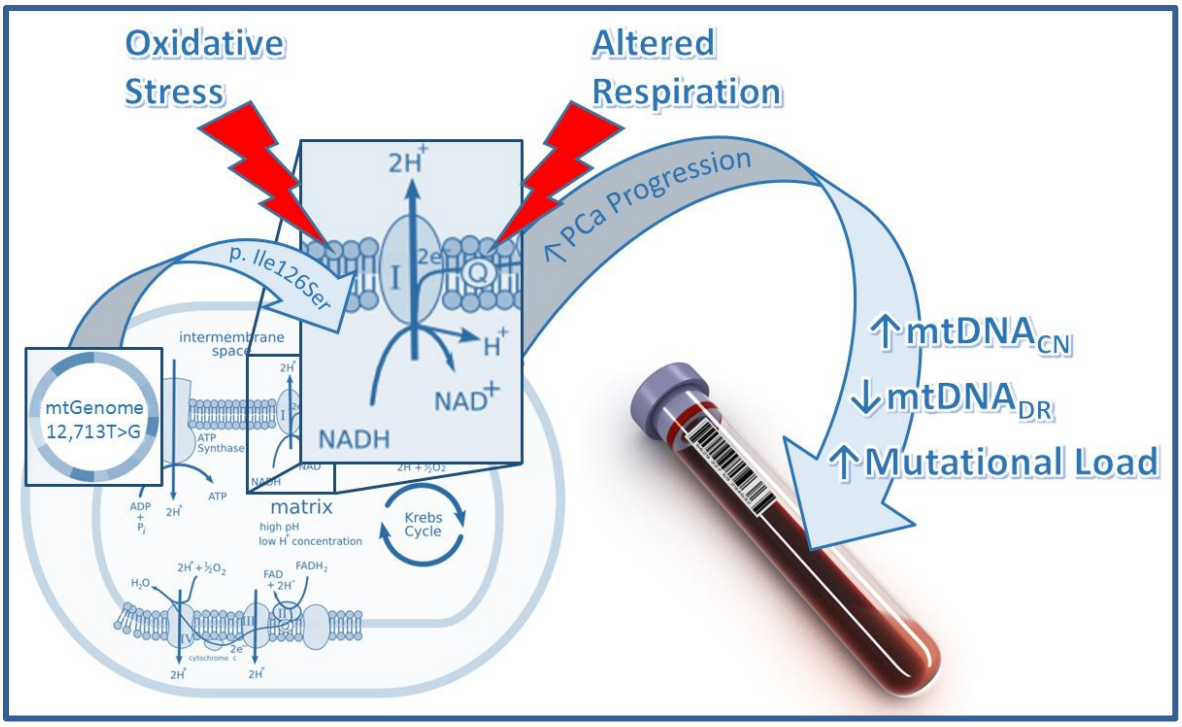


Figure 33. Proposed mechanism of mitochondrial dysfunction and resultant mtDNA markers found in blood associated with PCa progression. In this graphical representation, a single missense mutation (p. Ile126Ser) has a negative effect on the conformational energy state of ND5 which is a critical subunit in Complex I. As a consequence, the respiratory activity of Complex I is compromised which increases oxidative stress and potentiates an energetic deficiency. The resultant oxidative stress contributes to the adapted metabolic function of mitochondria in tumor cells and promotes disease progression. Changes to mtDNA caused by mitochondrial dysfunction associated with PCa accumulate overtime and increase with advancing stages of the disease. Measures of these mtDNA changes are detectable in the blood of PCa patients and can be used as molecular markers to assess disease progression.

As a function of PCa pathogenesis, mitochondrial bioenergetics are altered to accommodate the metabolic demand of proliferating tumor cells. Two well-characterized components of this metabolic shift are increased oxidative stress and decreased aerobic respiration. First, a ROS mediated axis is central to cancer development; ROS retrograde signaling promotes cellular adaptation of tumor cells which includes metabolic reprogramming,

cell proliferation and decreased apoptosis. Increased ROS in the mitochondria is also very damaging to mtDNA which, unrepaired, accumulates as genetic mutations with time as well as with PCa progression. Also, the consequence of metabolic reporgramming is two-fold. The high metabolic demand of proliferating tumor cells increases mitochondrial biogenesis, and therefore mtDNA replication, as a compensatory mechanism for an impaired ETC. In addition, the deficient mtDNA replication and mtDNA damage repair mechanisms which generate truncated mtDNAs under normal conditions is exacerbated by increased oxidative stress. Therefore, increased mtDNA replication in an exceedingly ROS-rich environment results in a reduction in the $mtDNA_{DR}$, *i.e.*, a higher number of truncated mtDNAs. Taken together, the body of evidence from this project supports the posit that changes to mtDNA, caused by mitochondrial dysfunction associated with PCa, accumulate over time and increase with advancing stages of the disease. Additional studies are expected to show that measures of these mtDNA changes are detectable in the blood of PCa patients and can be used as molecular markers to assess disease progression. Specifically, increased $mtDNA_{CN}$, decreased $mtDNA_{DR}$ and increased mutational load are indicators of advanced PCa and can be used to differentiate indolent from aggressive disease.

Appendix A: Haplotypes of 12 prostatectomy patients with PCa or BPH. The base position of every mtDNA variant observed among 12 prostatectomy patients is listed across the top of three panels: (A) HVI variants, (B) HVII variants, and (C) coding region variants. Listed below the base position is the corresponding revised Cambridge Reference Sequence (rCRS) nucleotide (A, T, G, or C). Each mtDNA variant is indexed as the divergent nucleotide in the column of its corresponding base position and row of the sample it was observed in. Not applicable (NA) is used in place of a nucleotide designator for insertions. Deletion = del.

A

HV1

	Position	16069	16092	16126	16140	16163	16183	16189	16192	16193.1	16218	16249	16256	16265	16267	16270	16287	16294	16296	16304	16311	16325	16327	16362
	rCRS*	C	T	T	T	A	A	T	C	NA	C	T	C	A	C	C	C	C	C	T	T	T	C	T
BPH	1350t										T													
BPH	1365t		C				C	del																
BPH	1391t			C							T						T	T	T	C				
PCa	1727t	T		C		G																		
PCa	1727b	T		C		G																		
PCa	1838t		C		C															C				
PCa	1838b		C		C															C				
PCa	2114t																							C
PCa	2114b																							C
PCa	2160t														T									
PCa	2160b														T									
PCa	2164t																				C	T	C	
PCa	2164b																				C	T	C	
PCa	2167t												G						C					
PCa	2167b												G						C					
PCa	2178t								T			T			T									C
PCa	2178b								T			T			T									C
PCa	2208t						C	C		C														
PCa	2208b						C	C		C														
PCa	2238t								T			C	T			T					C			
PCa	2238b								T			C	T			T					C			

B

HV2

	Position	73	92	143	146	185	188	195	228	239	247	249	257	263	290	291	295	309.1	309.3	315.1	458	462	477	489	493	499	523
	rCRS*	A	G	G	T	G	A	T	G	T	G	A	A	A	A	A	C	NA	NA	NA	C	C	T	T	A	G	A
BPH	1350t													G						C							
BPH	1365t													G				C		C						A	
BPH	1391t	G			C									G				C		C	T						
PCa	1727t	G				A	G		A					G			T	C		C		T		C			
PCa	1727b	G				A	G		A					G			T	C		C		T		C			
PCa	1838t							C						G				C		C							
PCa	1838b							C						G				C		C							
PCa	2114t									C				G				C		C							
PCa	2114b									C				G				C		C							
PCa	2160t													G						C							
PCa	2160b													G						C							
PCa	2164t	G		A				C				del		G	de	de		C		C				C	G		del
PCa	2164b	G		A				C				del		G	de	de		C		C				C	G		del
PCa	2167t							C			A			G						C							
PCa	2167b							C			A			G						C							
PCa	2178t	G												G						C							
PCa	2178b	G												G						C							
PCa	2208t												G	G				C	CCC	C			C				
PCa	2208b												G	G				C	CCC	C			C				
PCa	2238t													G						C							
PCa	2238b													G						C							

C

	Position	12,308	12,372	12,612	12,631	12,705	12,713	13,263	13,368	13,434	13,759	14,016	14,145	14,221	14,233
	Coding														
	rCRS*	A	G	A	T	C	T	A	G	A	G	G	A	T	A
BPH	1350t														
	1365t						G			G					
	1391t								A			A			G
PCa	1727t			G											
	1727b			G											
	1838t														
	1838b										A				
	2114t														
	2114b														
	2160t														
	2160b														
	2164t				G	T		G							
	2164b				G	T		G							
	2167t	G	A												
	2167b	G	A												
	2178t	G	A												
	2178b	G	A												
	2208t														
	2208b														
	2238t	G	A										G	C	
	2238b	G	A										G	C	

Appendix B: SABOR samples and their associated GPMs. (A) 42 samples in the normal (no cancer) control group; (B) 45 samples in the indolent PCa (NED) group; and (C) 42 samples in the advanced PCa (BCR/MET) group.

A

SampleID	Disease Status	GPMs		
110187	Normal	2558G	5972T	
110189	Normal	13224C		
110204	Normal	152Y	4435R	
120026	Normal	5C		
130117	Normal	150Y	8485R	
140066	Normal	149.1T	152d	
140094	Normal	12292Y		
140214	Normal			
140223	Normal	573.1CCC		
150049	Normal			
150178	Normal	8764R		
150297	Normal	7647Y		
150352	Normal			
150370	Normal	6908C	11992Y	
150427	Normal			
150606	Normal			
150671	Normal			
150726	Normal	15776G		
150987	Normal	10739G		
151181	Normal	10089R		
151225	Normal			
151305	Normal	14224R		
151307	Normal			
151396	Normal	2664Y	5484G	
151427	Normal			
151436	Normal	5069G		
160048	Normal	5033G		
190032	Normal	6307G		
190061	Normal			
190098	Normal	3918R	16182d	16183d
190101	Normal	7628A		
190275	Normal			
190353	Normal			
190357	Normal	8624T		
190389	Normal	6958d		
190405	Normal			
190493	Normal	8435G		
200256	Normal			
300060	Normal	10888T		
300084	Normal			
300102	Normal	390G		
300105	Normal			

B

SampleID	Disease Status	GPMs			
800008	NED	1619T	4890G	11453R	
800075	NED				
800098	NED				
800130	NED	292C			
800151	NED				
800154	NED				
800175	NED	12399S			
800178	NED	9660G	16485A		
800232	NED				
800233	NED				
800245	NED				
800264	NED				
810071	NED	4541R	6958d		
810087	NED				
810097	NED	13512R			
810109	NED				
810115	NED	6573R	15028T		
810132	NED				
810164	NED	5807G			
810167	NED	15007T			
810199	NED				
810250	NED	204Y			
820004	NED				
820024	NED				
820038	NED				
820059	NED				
820061	NED				
820072	NED	5098Y	13353G		
820087	NED				
820098	NED	5528Y	9591A		
820100	NED	16524C			
820104	NED				
820114	NED				
820146	NED	2222C			
820178	NED	4248Y	8042T	8842R	
820202	NED				
820203	NED				
820205	NED	5471R	8502d	8764R	9080R
820206	NED	5427G			
820215	NED				
820229	NED				
820230	NED	16189.1A	7079T	12501R	
820260	NED				
820309	NED	596C			
830057	NED				

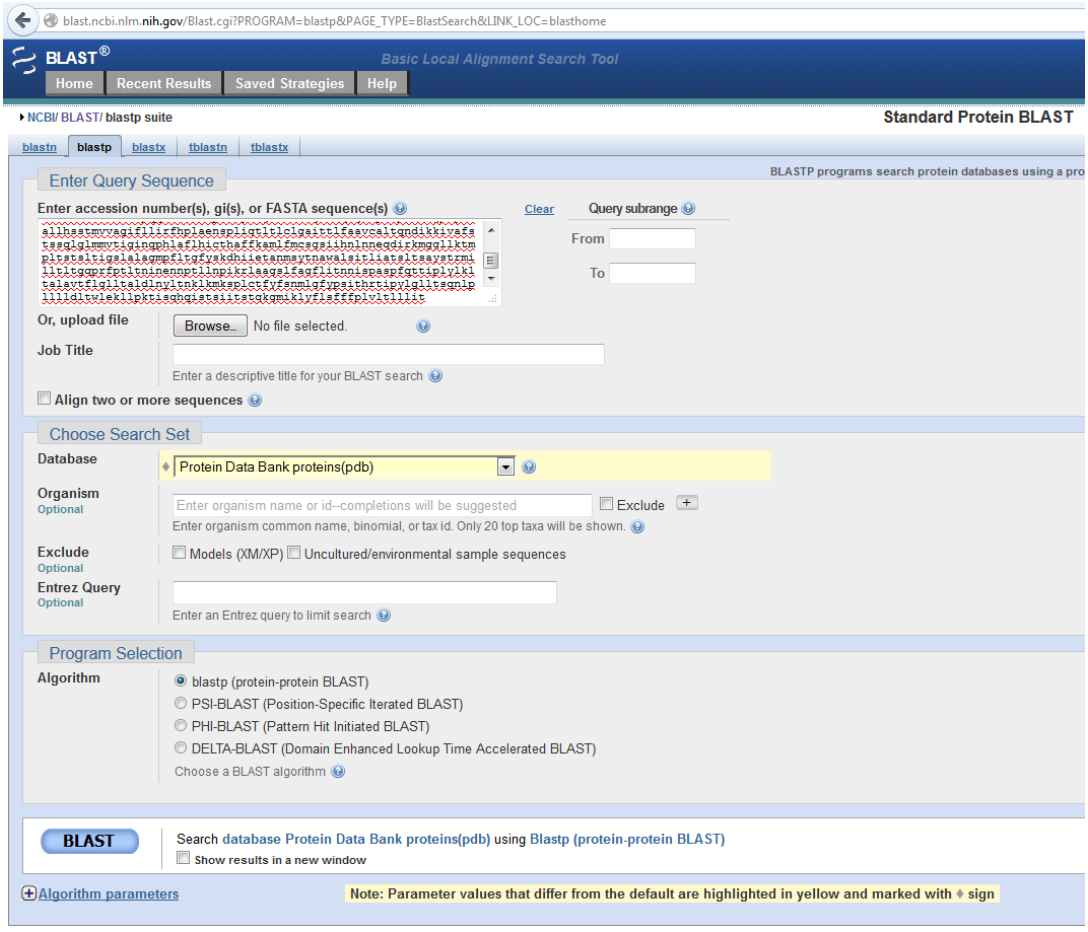
C

SampleID	Disease Status	GPMs				
800066	BCR/MET					
800138	BCR/MET	10712T				
800139	BCR/MET	5966G				
800148	BCR/MET	15758R				
800149	BCR/MET	1117G	6340T	6380G		
800155	BCR/MET	8994R	15064G			
800230	BCR/MET					
800256	BCR/MET					
800265	BCR/MET					
800270	BCR/MET	896R	13434R			
810006	BCR/MET	12172R				
810063	BCR/MET	8280G				
810134	BCR/MET					
810153	BCR/MET	15482C				
810172	BCR/MET					
810192	BCR/MET	9137C				
820037	BCR/MET					
820071	BCR/MET	573.1C	2356G	8994R		
820075	BCR/MET	149.1T	152d			
820168	BCR/MET	6958d				
820192	BCR/MET	573.1CCC				
820219	BCR/MET					
820224	BCR/MET	15731R				
820240	BCR/MET					
820243	BCR/MET	248d	16204A			
820263	BCR/MET	10290R				
820265	BCR/MET	3317G	13710T	15513T		
820285	BCR/MET	8237d				
820288	BCR/MET	11923R				
820290	BCR/MET	16512C	16541G	129TY	8502d	11347G
800063	BCR/MET	16278Y				
800106	BCR/MET					
800129	BCR/MET					
800141	BCR/MET	675G				
800160	BCR/MET	10905Y				
800215	BCR/MET	13263R				
800222	BCR/MET	8502d				
800226	BCR/MET	14013G				
820158	BCR/MET					
820233	BCR/MET	15104T				
820237	BCR/MET	15745T				
820248	BCR/MET					

Appendix C: Details of GPMs observed in the SABOR cohort.

				Location				Mutation		AA Change			
Global Prvt Mut	Frequency	rCRS	Observed	Region	Locus	AA	Codon Pos	Type	Effect	Before	After	Documented in MitoMap	PolyPhen-2 Score (0.000 - 1.000)
5C	1	A	C	Control	Control (Other)	NA	NA	Transition	NA	NA	NA	Yes	NA
149.1T	2	T	TT	Control	HV2	NA	NA	Insertion	NA	NA	NA	Yes	NA
150Y	1	C	Y (T/C)	Control	HV2	NA	NA	Transition	NA	NA	NA	Yes	NA
152Y	1	T	T/C	Control	HV2	NA	NA	Transition	NA	NA	NA	Yes	NA
152d	2	T	Del	Control	HV2	NA	NA	Deletion	NA	NA	NA	Yes	NA
204Y	1	T	Y (T/C)	Control	HV2	NA	NA	Transition	NA	NA	NA	Yes	NA
248d	1	A	Del	Control	HV2	NA	NA	Deletion	NA	NA	NA	Yes	NA
292C	1	T	C	Control	HV2	NA	NA	Transition	NA	NA	NA	Yes	NA
390G	1	A	G	Control	Control (Other)	NA	NA	Transition	NA	NA	NA	Yes	NA
573.1C	1	C	CC	Control	HV3	NA	NA	Insertion	NA	NA	NA	Yes	NA
573.1CCC	2	C	CCCC	Control	HV3	NA	NA	Insertion	NA	NA	NA	Yes	NA
596C	1	T	C	Coding	tRNA (Phe)	NA	NA	Transition	NA	NA	NA	Yes	NA
675G	1	A	G	Coding	12S rRNA	NA	NA	Transition	NA	NA	NA	Yes	NA
896R	1	A	R (A/G)	Coding	12S rRNA	NA	NA	Transition	NA	NA	NA	Yes	NA
1117G	1	A	G	Coding	12S rRNA	NA	NA	Transition	NA	NA	NA	Yes	NA
1291Y	1	T	Y (T/C)	Coding	12S rRNA	NA	NA	Transition	NA	NA	NA	Yes	NA
1619T	1	C	T	Coding	tRNA (Val)	NA	NA	Transition	NA	NA	NA	Yes	NA
2222C	1	T	C	Coding	16s rRNA	NA	NA	Transition	NA	NA	NA	Yes	NA
2356G	1	A	G	Coding	16S rRNA	NA	NA	Transition	NA	NA	NA	Yes	NA
2558G	1	A	G	Coding	16s rRNA	NA	NA	Transition	NA	NA	NA	Yes	NA
2664Y	1	T	Y (T/C)	Coding	16s rRNA	NA	NA	Transition	NA	NA	NA	No	NA
3317G	1	C	G	Coding	ND1	4	2	Transversion	Non-Syn	Ala	Gly	No	0.178(Benign)/0.043(Benign)
3918R	1	G	R (A/G)	Coding	ND1	204	3	Transition	Syn	Glu	Glu	Yes	NA
4248Y	1	T	Y (T/C)	Coding	ND1	314	3	Transition	Syn	Ile	Ile	Yes	NA
4435R	1	A	R (A/G)	Coding	tRNA (Met)	NA	NA	Transition	NA	NA	NA	Yes	NA
4541R	1	G	G/A	Coding	ND2	24	3	Transition	Syn	Ser	Ser	Yes	NA
4890G	1	A	G	Coding	ND2	141	1	Transition	Non-Syn	Ile	Val	Yes	0.997(Probably Damaging)/0.994(Probably Damaging)
5033G	1	A	G	Coding	ND2	188	3	Transition	Syn	Gly	Gly	Yes	NA
5069G	1	A	G	Coding	ND2	200	3	Transition	Syn	Met	Met	No	NA
5098Y	1	T	Y (T/C)	Coding	ND2	210	2	Transition	Non-Syn	Ile	Thr	No	0.003(Benign)/0.018(Benign)
5427G	1	A	G	Coding	ND2	320	1	Transition	Non-Syn	Thr	Ala	Yes	0.292(Benign)/0.265(Benign)
5471R	1	G	G/A	Coding	ND2	334	3	Transition	Syn	Thr	Thr	Yes	NA
5484G	1	A	G	Coding	ND2	339	1	Transition	Non-Syn	Ile	Val	Yes	0.011(Benign)/0.002(Benign)
5528Y	1	T	Y (T/C)	Coding	tRNA (Trp)	NA	NA	Transition	NA	NA	NA	Yes	NA
5807G	1	A	G/A	Coding	tRNA (Cys)	NA	NA	Transition	NA	NA	NA	Yes	NA
5966G	1	A	G	Coding	CO1	21	3	Transition	Syn	Leu	Leu	Yes	NA
5972T	1	C	T	Coding	CO1	23	3	Transition	Syn	Gly	Gly	Yes	NA
6307G	1	A	G	Coding	CO1	135	2	Transition	Non-Syn	Asn	Ser	No	0.993 (Probably Damaging)/0.973 (Probably Damaging)
6340T	1	C	T	Coding	CO1	146	2	Transition	Non-Syn	Thr	Ile	Yes	0.025(Benign)/0.111(Benign)
6380G	1	A	G	Coding	CO1	159	3	Transition	Syn	Leu	Leu	Yes	NA
6573R	1	G	R (G/A)	Coding	CO1	224	1	Transition	Non-Syn	Gly	AGA	No	Causes a -1 Frameshift
6908C	1	T	C	Coding	CO1	335	3	Transition	Syn	Ser	Ser	Yes	NA
6958d	3	G	Del	Coding	CO1	352	2	Deletion	Frameshift	Gly	Ala	No	NA
7079T	1	C	T	Coding	CO1	392	3	Transition	Syn	Gly	Gly	No	NA
7628A	1	C	A	Coding	CO2	15	1	Transition	Non-Syn	Pro	Thr	Yes	1.000(Probably Damaging)/0.996(Probably Damaging)
7647Y	1	T	Y (T/C)	Coding	CO2	21	2	Transition	Non-Syn	Ile	Thr	Yes	0.054(Benign)/0.063(Benign)
8042T	1	A	T	Coding	CO2	153	1	Transversion	Non-Syn	Met	Leu	No	0.698(Possibly Damaging)/0.878(Possibly Damaging)
8237d	1	A	Del	Coding	CO2	218	1	Deletion	Frameshift	Ile	Ser	No	NA
8280G	1	A	G	Coding	NC7	NA	NA	Transition	NA	NA	NA	Yes	NA
8435G	1	A	G	Coding	ATP8	24	1	Transition	Non-Syn	Thr	Ala	Yes	0.069(Benign)/0.088(Benign)
8485R	1	G	R (A/G)	Coding	ATP8	40	3	Transition	Syn	Lys	Lys	Yes	NA
8502d	3	A	Del	Coding	ATP8	46	2	Deletion	Frameshift	Asn	Ile	No	NA
8624T	1	C	T	Coding	ATP6	33	2	Transition	Non-Syn	Thr	Ile	Yes	0.272 (Benign)/0.124(Benign)
8764R	2	G	G/A	Coding	ATP6	80	1	Transition	Non-Syn	Ala	Thr	Yes	0.001 (Benign)/0.007 (Benign)
8842R	1	A	R (A/G)	Coding	ATP6	106	1	Transition	Non-Syn	Ile	Val	Yes	0.002(Benign)/0.003(Benign)
8994R	2	G	R (G/A)	Coding	ATP6	156	3	Transition	Syn	Leu	Leu	Yes	NA
9080R	1	A	A/G	Coding	ATP6	185	2	Transition	Non-Syn	Asn	Ser	Yes	0.002 (Benign)/0.010 (Benign)
9137C	1	T	C	Coding	ATP6	204	2	Transition	Non-Syn	Ile	Thr	Yes	.002 (benign)/.005 (benign)
9591A	1	G	A	Coding	CO3	129	1	Transition	Non-Syn	Val	Ile	Yes	0.770(Possibly Damaging)/0.899(Possibly Damaging)
9660G	1	A	G	Coding	CO3	152	1	Transition	Non-Syn	Met	Val	Yes	0.335(Benign)/0.367(Benign)
10089R	1	A	R (A/G)	Coding	ND3	11	1	Transition	Non-Syn	Thr	Ala	Yes	0.966 (Probably Damaging)/0.903 (Probably Damaging)
10290R	1	G	R (G/A)	Coding	ND3	78	1	Transition	Non-Syn	Ala	Thr	No	0.985 (Probably Damaging)/0.941 (Probably Damaging)
10712T	1	C	T	Coding	ND4L	81	3	Transition	Syn	Ile	Ile	Yes	NA
10739G	1	A	G	Coding	ND4L	90	3	Transition	Syn	Val	Val	Yes	NA
10888T	1	C	T	Coding	ND4	43	3	Transition	Syn	Asn	Asn	No	NA
10905Y	1	T	Y (T/C)	Coding	ND4	49	2	Transition	Non-Syn	Leu	Pro	No	0.999(Probably Damaging)/0.998(Probably Damaging)
11347G	1	A	G	Coding	ND4	196	3	Transition	Syn	Trp	Trp	Yes	NA
11453R	1	G	R (G/A)	Coding	ND4	232	1	Transition	Non-Syn	Ala	Thr	Yes	0.999(Probably Damaging)/0.995(Probably Damaging)
11923R	1	A	R (A/G)	Coding	ND4	388	3	Transition	Syn	Trp	Trp	Yes	NA
11992Y	1	T	Y (T/C)	Coding	ND4	411	3	Transition	Syn	Phe	Phe	Yes	NA
12172R	1	A	R (A/G)	Coding	tRNA (His)	NA	NA	Transition	NA	NA	Na	Yes	NA
12292Y	1	T	Y (T/C)	Coding	tRNA (Leu 2)	NA	NA	Transition	NA	NA	NA	No	NA
12399S	1	C	S (C/G)	Coding	ND5	21	3	Transversion	Syn	Thr	Thr	No	NA
12501R	1	G	G/A	Coding	ND5	55	3	Transition	Syn	Met	Met	Yes	NA
13224C	1	T	C	Coding	ND5	296	3	Transition	Syn	Asn	Asn	Yes	NA
13263R	1	A	R (A/G)	Coding	ND5	309	3	Transition	Syn	Gln	Gln	Yes	NA
13353G	1	A	G	Coding	ND5	339	3	Transition	Syn	Leu	Leu	Yes	NA
13434R	1	A	R (A/G)	Coding	ND5	366	3	Transition	Syn	Met	Met	Yes	NA
13512R	1	A	R (A/G)	Coding	ND5	392	3	Transition	Syn	Lys	Lys	Yes	NA
13710T	1	A	T	Coding	ND5	458	3	Transversion	Syn	Ala	Ala	Yes	NA
14013G	1	A	G	Coding	ND5	559	3	Transition	Syn	Glu	Glu	Yes	NA
14224R	1	A	R (A/G)	Coding	ND6	150	3	Transition	Syn	Arg	Arg	Yes	NA
15007T	1	C	T	Coding	CYB	87	3	Transition	Syn	Ala	Ala	Yes	NA
15028T	1	C	T	Coding	CYB	94	3	Transition	Syn	Lys	Lys	Yes	NA
15064G	1	A	G	Coding	CYB	106	3	Transition	Syn	Ser	Ser	Yes	NA
15104T	1	C	T	Coding	CYB	120	1	Transition	Syn	Leu	Leu	Yes	NA
15482C	1	T	C	Coding	CYB	246	1	Transition	Non-Syn	Ser	Pro	Yes	0.574(Possibly Damaging)/0.392(Benign)
15513T	1	A	T	Coding	CYB	256	2	Transversion	Non-Syn	Tyr	Phe	No	0.992(Probably Damaging)/0.986(Probably Damaging)
15731R	1	G	R (G/A)	Coding	CYB	329	1	Transition	Non-Syn	Ala	Thr	Yes	.001 (Benign)/.004 (Benign)
15745T	1	C	T	Coding	CYB	333	3	Transversion	Syn	Leu	Leu	Yes	NA
15758R	1	A	R (A/G)	Coding	CYB	338	1	Transition	Non-Syn	Ile	Val	Yes	0.000(Benign)/0.004(Benign)
15776G	1	A	G	Coding	CYB	344	1	Transition	Non-Syn	Ser	Gly	Yes	0.348 (Benign)/0.380 (Benign)
16182d	1	A	AA-Del	Control	HV1	NA	NA	Deletion	NA	NA	NA	Yes	NA
16189.1A	1	A	AA	Control	HV1	NA	NA	Insertion	NA	NA	NA	No	NA
16204A	1	G	A	Control	HV1	NA	NA	Transition	NA	NA	NA	Yes	NA
16278Y	1	C	Y (C/T)	Control	HV1	NA	NA	Transversion	NA	NA	NA	Yes	NA
16485A	1	G	A	Control	Control (Other)	NA	NA	Transition	NA	NA	NA	Yes	NA
16512C	1	T	C	Control	Control (Other)	NA	NA	Transition	NA	NA	NA	Yes	NA
16524C	1	A	C	Control	Control (Other)	NA	NA	Transversion	NA	NA	NA	Yes	NA
16541G	1	A	G	Control	Control (Other)	NA	NA	Transition	NA	NA	NA	Yes	NA

Appendix D: Protein BLAST (blastp) of ND5.



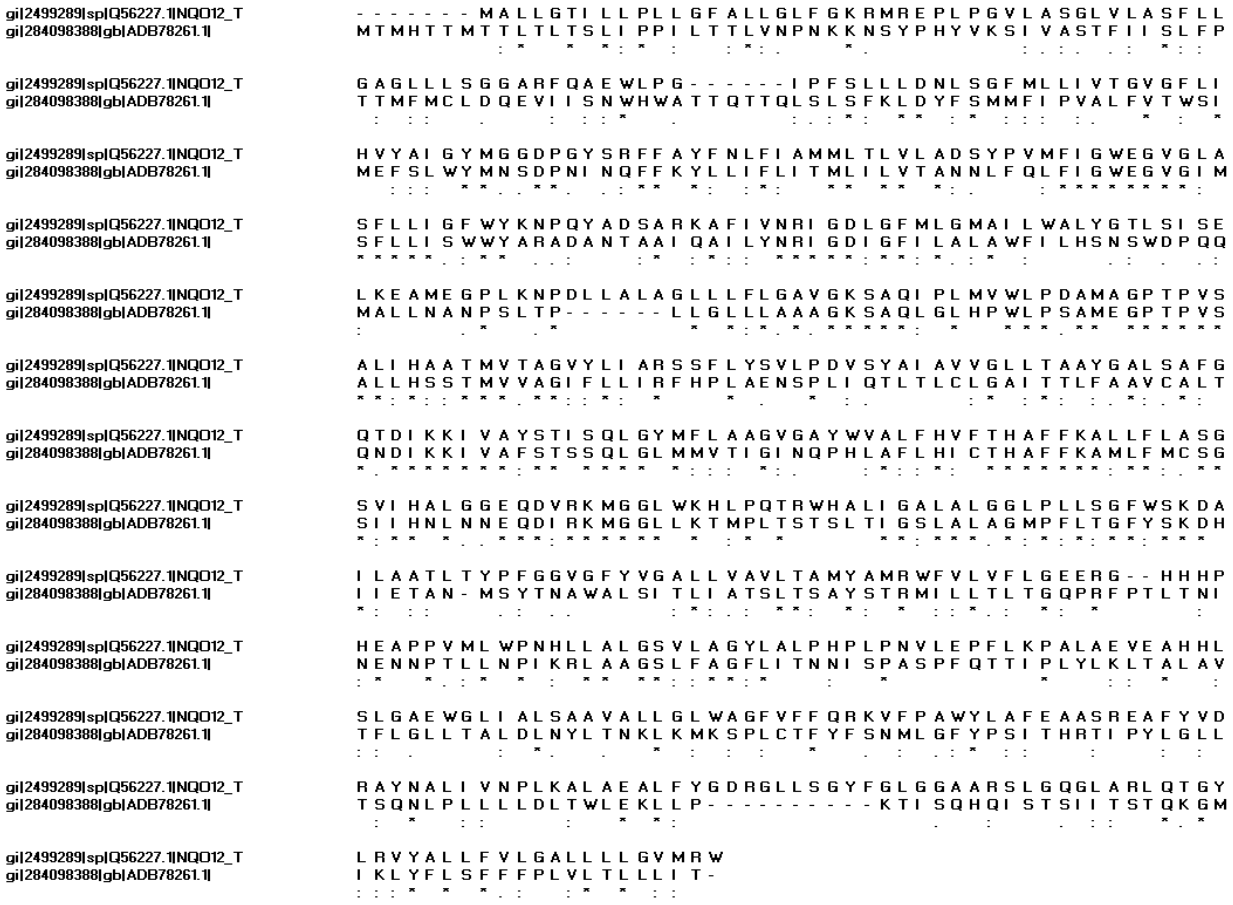
Sequences producing significant alignments:

Select: All None Selected:0

Alignments Download GenPept Graphics Distance tree of results Multiple alignment

	Max score	Total score	Query cover	E value	Ident	Accession
Chain L, Crystal Structure Of The Membrane Domain Of Respiratory Complex I From E. Coli At 3.0 Angstrom	259	259	71%	5e-77	36%	3RKO_L
Chain L, Crystal Structure Of The Membrane Domain Of Respiratory Complex I From Thermus Thermophilusal S	249	249	64%	2e-73	42%	4HE8_L
Chain N, Crystal Structure Of The Membrane Domain Of Respiratory Complex I From E. Coli At 3.0 Angstrom	78.6	78.6	60%	4e-15	21%	3RKO_N
Chain M, Crystal Structure Of The Membrane Domain Of Respiratory Complex I From E. Coli At 3.0 Angstrom	75.1	75.1	53%	5e-14	27%	3RKO_M
Chain N, Crystal Structure Of The Membrane Domain Of Respiratory Complex I From Thermus Thermophilusal S	53.1	53.1	35%	4e-07	24%	4HE8_N
Chain M, Crystal Structure Of The Membrane Domain Of Respiratory Complex I From Thermus Thermophilussta	40.8	40.8	34%	0.003	25%	4HE8_M

Appendix E: EMBOSS Needle alignment. Pairwise global alignment between NADH-quinone oxidoreductase subunit 12 (UniProtKB/Swiss-Prot: Q56227.1) and ND5 (GenBank: ADB78261.1).


gi|2499289|sp|Q56227.1|NQO12_T
gi|284098388|gb|ADB78261.1|

BIBLIOGRAPHY

1. Siegel, R. L., Miller, K. D. & Jemal, A. Cancer statistics, 2015. *CA. Cancer J. Clin.* **65,** 5–29 (2015).

2. Jemal, A. *et al.* Global cancer statistics. *CA. Cancer J. Clin.* **61,** 69–90

3. Kiciński, M., Vangronsveld, J. & Nawrot, T. S. An epidemiological reappraisal of the familial aggregation of prostate cancer: a meta-analysis. *PLoS One* **6,** e27130 (2011).

4. Zeegers, M. P. A., Jellema, A. & Ostrer, H. Empiric risk of prostate carcinoma for relatives of patients with prostate carcinoma: a meta-analysis. *Cancer* **97,** 1894–903 (2003).

5. Carter, B. S., Beaty, T. H., Steinberg, G. D., Childs, B. & Walsh, P. C. Mendelian inheritance of familial prostate cancer. *Proc. Natl. Acad. Sci. U. S. A.* **89,** 3367–71 (1992).

6. PROSTATE.COM – What is Prostate Cancer? at <http://www.prostate.com/patient/ProstateCancerBasics/WhatIsProstateCancer.aspx>

7. McNeal, J. E. The zonal anatomy of the prostate. *Prostate* **2,** 35–49 (1981).

8. Wadhera, P. An introduction to acinar pressures in BPH and prostate cancer. *Nat. Rev. Urol.* **10,** 358–66 (2013).

9. McNeal, J. E., Redwine, E. A., Freiha, F. S. & Stamey, T. A. Zonal distribution of prostatic adenocarcinoma. Correlation with histologic pattern and direction of spread. *Am. J. Surg. Pathol.* **12,** 897–906 (1988).

10. Cohen, R. J. *et al.* Central zone carcinoma of the prostate gland: a distinct tumor type with poor prognostic features. *J. Urol.* **179,** 1762–7; discussion 1767 (2008).

11. McNeal, J. E. Origin and evolution of benign prostatic enlargement. *Invest. Urol.* **15,** 340–5 (1978).

12. McNeal, J. E. Anatomy of the prostate and morphogenesis of BPH. *Prog. Clin. Biol. Res.*

145, 27–53 (1984).

13. Augustin, H. *et al.* Biochemical recurrence following radical prostatectomy: a comparison between prostate cancers located in different anatomical zones. *Prostate* **55,** 48–54 (2003).

14. Shen, M. M. & Abate-Shen, C. Molecular genetics of prostate cancer: new prospects for old challenges. *Genes Dev.* **24,** 1967–2000 (2010).

15. Chung, L. W. K., Isaacs, W. B. & Simons, J. W. *Prostate Cancer: Biology, Genetics, and the New Therapeutics*. **10,** (Springer Science & Business Media, 2007).

16. Liu, A. Y. *et al.* Cell-cell interaction in prostate gene regulation and cytodifferentiation. *Proc. Natl. Acad. Sci. U. S. A.* **94,** 10705–10 (1997).

17. Okada, H. *et al.* Keratin profiles in normal/hyperplastic prostates and prostate carcinoma. *Virchows Arch. A. Pathol. Anat. Histopathol.* **421,** 157–61 (1992).

18. Sherwood, E. R. *et al.* Differential cytokeratin expression in normal, hyperplastic and malignant epithelial cells from human prostate. *J. Urol.* **143,** 167–71 (1990).

19. Sherwood, E. R. *et al.* Differential expression of specific cytokeratin polypeptides in the basal and luminal epithelia of the human prostate. *Prostate* **18,** 303–14 (1991).

20. Webpathology.com: A Collection of Surgical Pathology Images. at <http://www.webpathology.com/image.asp?case=14&n=1>

21. Abrahamsson, P. A. Neuroendocrine differentiation and hormone-refractory prostate cancer. *Prostate. Suppl.* **6,** 3–8 (1996).

22. Krijnen, J. L. *et al.* Do neuroendocrine cells in human prostate cancer express androgen receptor? *Histochemistry* **100,** 393–8 (1993).

23. Xue, Y. *et al.* Prostatic neuroendocrine cells have a unique keratin expression pattern and do not express Bcl-2: cell kinetic features of neuroendocrine cells in the human prostate. *Am. J. Pathol.* **151,** 1759–65 (1997).

24. di Sant'Agnese, P. A. Neuroendocrine differentiation in prostatic carcinoma: an update. *Prostate. Suppl.* **8,** 74–9 (1998).

25. Sciarra, A. *et al.* Neuroendocrine differentiation in human prostate tissue: is it detectable and treatable? *BJU Int.* **91,** 438–45 (2003).

26. Nelson, W. G., De Marzo, A. M. & Isaacs, W. B. Prostate cancer. *N. Engl. J. Med.* **349,** 366–81 (2003).

27. Gleason, D. F. Classification of prostatic carcinomas. *Cancer Chemother. Rep.* **50,** 125–8 (1966).

28. Korsten, H., Ziel-van der Made, A., Ma, X., van der Kwast, T. & Trapman, J. Accumulating progenitor cells in the luminal epithelial cell layer are candidate tumor initiating cells in a Pten knockout mouse prostate cancer model. *PLoS One* **4,** e5662 (2009).

29. Wang, X. *et al.* A luminal epithelial stem cell that is a cell of origin for prostate cancer. *Nature* **461,** 495–500 (2009).

30. Grisanzio, C. & Signoretti, S. p63 in prostate biology and pathology. *J. Cell. Biochem.* **103,** 1354–68 (2008).

31. De Marzo, A. M., Nelson, W. G., Meeker, A. K. & Coffey, D. S. Stem cell features of benign and malignant prostate epithelial cells. *J. Urol.* **160,** 2381–92 (1998).

32. Goldstein, A. S., Huang, J., Guo, C., Garraway, I. P. & Witte, O. N. Identification of a cell of origin for human prostate cancer. *Science* **329,** 568–71 (2010).

33. Lawson, D. A. *et al.* Basal epithelial stem cells are efficient targets for prostate cancer initiation. *Proc. Natl. Acad. Sci. U. S. A.* **107,** 2610–5 (2010).

34. Barker, N. *et al.* Crypt stem cells as the cells-of-origin of intestinal cancer. *Nature* **457,** 608–11 (2009).

35. Huntly, B. J. P. *et al.* MOZ-TIF2, but not BCR-ABL, confers properties of leukemic stem

cells to committed murine hematopoietic progenitors. *Cancer Cell* **6,** 587–96 (2004).

36. Cunha, G. R. *et al.* The endocrinology and developmental biology of the prostate. *Endocr. Rev.* **8,** 338–62 (1987).

37. Lilja, H. A kallikrein-like serine protease in prostatic fluid cleaves the predominant seminal vesicle protein. *J. Clin. Invest.* **76,** 1899–903 (1985).

38. Lilja, H., Oldbring, J., Rannevik, G. & Laurell, C. B. Seminal vesicle-secreted proteins and their reactions during gelation and liquefaction of human semen. *J. Clin. Invest.* **80,** 281–5 (1987).

39. Lee, C., Tsai, Y., Sensibar, J., Oliver, L. & Grayhack, J. T. Two-dimensional characterization of prostatic acid phosphatase, prostatic specific antigen and prostate binding protein in expressed prostatic fluid. *Prostate* **9,** 135–46 (1986).

40. Coffey, D. S. & Pienta, K. J. New concepts in studying the control of normal and cancer growth of the prostate. *Prog. Clin. Biol. Res.* **239,** 1–73 (1987).

41. Fair, W. R. & Wehner, N. The prostatic antibacterial factor: identity and significance. *Prog. Clin. Biol. Res.* **6,** 383–403 (1976).

42. Kavanagh, J. P. Sodium, potassium, calcium, magnesium, zinc, citrate and chloride content of human prostatic and seminal fluid. *J. Reprod. Fertil.* **75,** 35–41 (1985).

43. Korkmaz, K. S. *et al.* Molecular cloning and characterization of STAMP1, a highly prostate-specific six transmembrane protein that is overexpressed in prostate cancer. *J. Biol. Chem.* **277,** 36689–96 (2002).

44. deGroat, W. C. & Booth, A. M. Physiology of male sexual function. *Ann. Intern. Med.* **92,** 329–31 (1980).

45. Priest, R., Garzotto, M. & Kaufman, J. Benign prostatic hyperplasia: a brief overview of pathogenesis, diagnosis, and therapy. *Tech. Vasc. Interv. Radiol.* **15,** 261–4 (2012).

46. Enlarged-Prostate-2.jpg (JPEG Image, 400 × 251 pixels). at

<http://hybridrastamama.com/wp-content/uploads/2014/05/Enlarged-Prostate-2.jpg>

47. Matsumoto, A. M. Andropause: clinical implications of the decline in serum testosterone levels with aging in men. *J. Gerontol. A. Biol. Sci. Med. Sci.* **57,** M76–99 (2002).

48. Vermeulen, A., Kaufman, J. M., Goemaere, S. & van Pottelberg, I. Estradiol in elderly men. *Aging Male* **5,** 98–102 (2002).

49. Peehl, D. M. Primary cell cultures as models of prostate cancer development. *Endocr. Relat. Cancer* **12,** 19–47 (2005).

50. Tamburrino, L. *et al.* Androgen receptor (AR) expression in prostate cancer and progression of the tumor: Lessons from cell lines, animal models and human specimens. *Steroids* **77,** 996–1001 (2012).

51. Foster, C. S. *et al.* Cellular and molecular pathology of prostate cancer precursors. *Scand. J. Urol. Nephrol. Suppl.* 19–43 (2000). at <http://www.ncbi.nlm.nih.gov/pubmed/11144897>

52. Hanahan, D. & Weinberg, R. A. The hallmarks of cancer. *Cell* **100,** 57–70 (2000).

53. Mathew, R., Karantza-Wadsworth, V. & White, E. Role of autophagy in cancer. *Nat. Rev. Cancer* **7,** 961–7 (2007).

54. Barbieri, C. E. *et al.* The mutational landscape of prostate cancer. *Eur. Urol.* **64,** 567–76 (2013).

55. Barbieri, C. E. *et al.* Exome sequencing identifies recurrent SPOP, FOXA1 and MED12 mutations in prostate cancer. *Nat. Genet.* **44,** 685–9 (2012).

56. Grasso, C. S. *et al.* The mutational landscape of lethal castration-resistant prostate cancer. *Nature* **487,** 239–43 (2012).

57. Taylor, B. S. *et al.* Integrative genomic profiling of human prostate cancer. *Cancer Cell* **18,** 11–22 (2010).

58. Wilson, J. D. The critical role of androgens in prostate development. *Endocrinol. Metab. Clin. North Am.* **40,** 577–90, ix (2011).

59. Huggins, C. ENDOCRINE CONTROL OF PROSTATIC CANCER. *Science* **97,** 541–4 (1943).

60. Hakimi, J. M., Rondinelli, R. H., Schoenberg, M. P. & Barrack, E. R. Androgen-receptor gene structure and function in prostate cancer. *World J. Urol.* **14,** 329–37 (1996).

61. Huggins, C. Endocrine-induced regression of cancers. *Cancer Res.* **27,** 1925–30 (1967).

62. Huggins, C. & Hodges, C. V. Studies on prostatic cancer. I. The effect of castration, of estrogen and androgen injection on serum phosphatases in metastatic carcinoma of the prostate. *CA. Cancer J. Clin.* **22,** 232–40

63. De Marzo, A. M. *et al.* Pathological and molecular mechanisms of prostate carcinogenesis: implications for diagnosis, detection, prevention, and treatment. *J. Cell. Biochem.* **91,** 459–77 (2004).

64. Ramsay, A. K. & Leung, H. Y. Signalling pathways in prostate carcinogenesis: potentials for molecular-targeted therapy. *Clin. Sci. (Lond).* **117,** 209–28 (2009).

65. Denmeade, S. R., Lin, X. S. & Isaacs, J. T. Role of programmed (apoptotic) cell death during the progression and therapy for prostate cancer. *Prostate* **28,** 251–65 (1996).

66. Pienta, K. J. & Bradley, D. Mechanisms underlying the development of androgen-independent prostate cancer. *Clin. Cancer Res.* **12,** 1665–71 (2006).

67. Feldman, B. J. & Feldman, D. The development of androgen-independent prostate cancer. *Nat. Rev. Cancer* **1,** 34–45 (2001).

68. Grossmann, M. E., Huang, H. & Tindall, D. J. Androgen receptor signaling in androgen-refractory prostate cancer. *J. Natl. Cancer Inst.* **93,** 1687–97 (2001).

69. Heinlein, C. A. & Chang, C. Androgen receptor in prostate cancer. *Endocr. Rev.* **25,** 276–308 (2004).

70. Litvinov, I. V, De Marzo, A. M. & Isaacs, J. T. Is the Achilles' heel for prostate cancer therapy a gain of function in androgen receptor signaling? *J. Clin. Endocrinol. Metab.* **88,** 2972–82 (2003).

71. Buchanan, G., Irvine, R. A., Coetzee, G. A. & Tilley, W. D. Contribution of the androgen receptor to prostate cancer predisposition and progression. *Cancer Metastasis Rev.* **20,** 207–23 (2001).

72. Eder, I. E. *et al.* Inhibition of LncaP prostate cancer cells by means of androgen receptor antisense oligonucleotides. *Cancer Gene Ther.* **7,** 997–1007 (2000).

73. Hååg, P., Bektic, J., Bartsch, G., Klocker, H. & Eder, I. E. Androgen receptor down regulation by small interference RNA induces cell growth inhibition in androgen sensitive as well as in androgen independent prostate cancer cells. *J. Steroid Biochem. Mol. Biol.* **96,** 251–8 (2005).

74. Liao, X., Tang, S., Thrasher, J. B., Griebling, T. L. & Li, B. Small-interfering RNA-induced androgen receptor silencing leads to apoptotic cell death in prostate cancer. *Mol. Cancer Ther.* **4,** 505–15 (2005).

75. Zegarra-Moro, O. L., Schmidt, L. J., Huang, H. & Tindall, D. J. Disruption of androgen receptor function inhibits proliferation of androgen-refractory prostate cancer cells. *Cancer Res.* **62,** 1008–13 (2002).

76. Migliaccio, A. *et al.* Steroid-induced androgen receptor-oestradiol receptor beta-Src complex triggers prostate cancer cell proliferation. *EMBO J.* **19,** 5406–17 (2000).

77. Cher, M. L. *et al.* Cancer interaction with the bone microenvironment: a workshop of the National Institutes of Health Tumor Microenvironment Study Section. *Am. J. Pathol.* **168,** 1405–12 (2006).

78. Chung, L. W. K., Baseman, A., Assikis, V. & Zhau, H. E. Molecular insights into prostate cancer progression: the missing link of tumor microenvironment. *J. Urol.* **173,** 10–20 (2005).

79. Karlou, M., Tzelepi, V. & Efstathiou, E. Therapeutic targeting of the prostate cancer microenvironment. *Nat. Rev. Urol.* **7,** 494–509 (2010).

80. Morrissey, C. & Vessella, R. L. The role of tumor microenvironment in prostate cancer bone metastasis. *J. Cell. Biochem.* **101,** 873–86 (2007).

81. Bussard, K. M., Gay, C. V & Mastro, A. M. The bone microenvironment in metastasis; what is special about bone? *Cancer Metastasis Rev.* **27,** 41–55 (2008).

82. Bubendorf, L. *et al.* Metastatic patterns of prostate cancer: an autopsy study of 1,589 patients. *Hum. Pathol.* **31,** 578–83 (2000).

83. Paget, S. The distribution of secondary growths in cancer of the breast. 1889. *Cancer Metastasis Rev.* **8,** 98–101 (1989).

84. Hill, R., Song, Y., Cardiff, R. D. & Van Dyke, T. Selective evolution of stromal mesenchyme with p53 loss in response to epithelial tumorigenesis. *Cell* **123,** 1001–11 (2005).

85. Tuxhorn, J. A., Ayala, G. E. & Rowley, D. R. Reactive stroma in prostate cancer progression. *J. Urol.* **166,** 2472–83 (2001).

86. Thalmann, G. N. *et al.* Androgen-independent cancer progression and bone metastasis in the LNCaP model of human prostate cancer. *Cancer Res.* **54,** 2577–81 (1994).

87. Fidler, I. J. The organ microenvironment and cancer metastasis. *Differentiation.* **70,** 498–505 (2002).

88. Horak, C. E. & Steeg, P. S. Metastasis gets site specific. *Cancer Cell* **8,** 93–5 (2005).

89. Minn, A. J. *et al.* Distinct organ-specific metastatic potential of individual breast cancer cells and primary tumors. *J. Clin. Invest.* **115,** 44–55 (2005).

90. Olsson, A. Y., Bjartell, A., Lilja, H. & Lundwall, A. Expression of prostate-specific antigen (PSA) and human glandular kallikrein 2 (hK2) in ileum and other extraprostatic tissues. *Int. J. Cancer* **113,** 290–7 (2005).

91. Kuriyama, M. *et al.* Quantitation of prostate-specific antigen in serum by a sensitive enzyme immunoassay. *Cancer Res.* **40,** 4658–62 (1980).

92. Papsidero, L. D., Wang, M. C., Valenzuela, L. A., Murphy, G. P. & Chu, T. M. A prostate antigen in sera of prostatic cancer patients. *Cancer Res.* **40,** 2428–32 (1980).

93. Sävblom, C. *et al.* Blood levels of free-PSA but not complex-PSA significantly correlates to prostate release of PSA in semen in young men, while blood levels of complex-PSA, but not free-PSA increase with age. *Prostate* **65,** 66–72 (2005).

94. Thompson, I. M. *et al.* Prevalence of prostate cancer among men with a prostate-specific antigen level < or =4.0 ng per milliliter. *N. Engl. J. Med.* **350,** 2239–46 (2004).

95. Aus, G. *et al.* Individualized screening interval for prostate cancer based on prostate-specific antigen level: results of a prospective, randomized, population-based study. *Arch. Intern. Med.* **165,** 1857–61 (2005).

96. Živković, S. Correlation between prostate-specific antigen and histopathological difference of prostate carcinoma. *Arch. Oncol.* **12,** 148–151

97. Grossklaus, D. J. *et al.* The free/total prostate-specific antigen ratio (%fPSA) is the best predictor of tumor involvement in the radical prostatectomy specimen among men with an elevated PSA. *Urol. Oncol.* **7,** 195–8

98. Pinsky, P. F. *et al.* Prostate-specific antigen velocity and prostate cancer gleason grade and stage. *Cancer* **109,** 1689–95 (2007).

99. Stamey, T. A. *et al.* Prostate-specific antigen as a serum marker for adenocarcinoma of the prostate. *N. Engl. J. Med.* **317,** 909–16 (1987).

100. Cooner, W. H. *et al.* Prostate cancer detection in a clinical urological practice by ultrasonography, digital rectal examination and prostate specific antigen. *J. Urol.* **143,** 1146–52; discussion 1152–4 (1990).

101. Morgan, T. O. *et al.* Age-specific reference ranges for prostate-specific antigen in black men. *N. Engl. J. Med.* **335,** 304–10 (1996).

102. Oesterling, J. E. Prostate specific antigen: a critical assessment of the most useful tumor marker for adenocarcinoma of the prostate. *J. Urol.* **145,** 907–23 (1991).

103. Ankerst, D. P. & Thompson, I. M. Sensitivity and specificity of prostate-specific antigen for prostate cancer detection with high rates of biopsy verification. *Arch. Ital. Urol. Androl.* **78,** 125–9 (2006).

104. Andriole, G. L. *et al.* Prostate Cancer Screening in the Prostate, Lung, Colorectal and Ovarian (PLCO) Cancer Screening Trial: findings from the initial screening round of a randomized trial. *J. Natl. Cancer Inst.* **97,** 433–8 (2005).

105. Hugosson, J., Aus, G., Lilja, H., Lodding, P. & Pihl, C.-G. Results of a randomized, population-based study of biennial screening using serum prostate-specific antigen measurement to detect prostate carcinoma. *Cancer* **100,** 1397–405 (2004).

106. Crawford, E. D. *et al.* Serum prostate-specific antigen and digital rectal examination for early detection of prostate cancer in a national community-based program. The Prostate Cancer Education Council. *Urology* **47,** 863–9 (1996).

107. Luboldt, H.-J., Schindler, J. F. & Rübben, H. Age-Specific Reference Ranges for Prostate-Specific Antigen as a Marker for Prostate Cancer. *EAU-EBU Updat. Ser.* **5,** 38–48 (2007).

108. Barry, M. J. Clinical practice. Prostate-specific-antigen testing for early diagnosis of prostate cancer. *N. Engl. J. Med.* **344,** 1373–7 (2001).

109. Hoffman, R. M., Stone, S. N., Espey, D. & Potosky, A. L. Differences between men with screening-detected versus clinically diagnosed prostate cancers in the USA. *BMC Cancer* **5,** 27 (2005).

110. Friedrich, M. J. Debate continues on use of PSA testing for early detection of prostate cancer. *JAMA* **305,** 2273–5 (2011).

111. Sirovich, B. E., Schwartz, L. M. & Woloshin, S. Screening men for prostate and colorectal cancer in the United States: does practice reflect the evidence? *JAMA* **289,** 1414–20 (2003).

112. Altekruse SF, Kosary CL, Krapcho M, Neyman N, Aminou R, Waldron W, Ruhl J, Howlader N, Tatalovich Z, Cho H, Mariotto A, Eisner MP, Lewis DR, Cronin K, Chen HS, Feuer EJ, Stinchcomb DG, E. B. *SEER Cancer Statistics Review, 1975-2007*. (2010). at <http://seer.cancer.gov/csr/1975_2007/>

113. Seidman, H., Mushinski, M. H., Gelb, S. K. & Silverberg, E. Probabilities of eventually developing or dying of cancer--United States, 1985. *CA. Cancer J. Clin.* **35,** 36–56

114. Ung, J. O., Richie, J. P., Chen, M.-H., Renshaw, A. A. & D'Amico, A. V. Evolution of the presentation and pathologic and biochemical outcomes after radical prostatectomy for patients with clinically localized prostate cancer diagnosed during the PSA era. *Urology* **60,** 458–63 (2002).

115. Andriole, G. L. *et al.* Mortality results from a randomized prostate-cancer screening trial. *N. Engl. J. Med.* **360,** 1310–9 (2009).

116. Schröder, F. H. *et al.* Prostate-cancer mortality at 11 years of follow-up. *N. Engl. J. Med.* **366,** 981–90 (2012).

117. Schröder, F. H. *et al.* Screening and prostate-cancer mortality in a randomized European study. *N. Engl. J. Med.* **360,** 1320–8 (2009).

118. Moyer, V. A. Screening for prostate cancer: U.S. Preventive Services Task Force recommendation statement. *Ann. Intern. Med.* **157,** 120–34 (2012).

119. Miller, A. B. New data on prostate-cancer mortality after PSA screening. *N. Engl. J. Med.* **366,** 1047–8 (2012).

120. Rosario, D. J. *et al.* Short term outcomes of prostate biopsy in men tested for cancer by prostate specific antigen: prospective evaluation within ProtecT study. *BMJ* **344,** d7894 (2012).

121. Klein, T. *et al.* The impact of prostate biopsy and periprostatic nerve block on erectile and voiding function: a prospective study. *J. Urol.* **184,** 1447–52 (2010).

122. Coley, C. M., Barry, M. J., Fleming, C. & Mulley, A. G. Early detection of prostate cancer. Part I: Prior probability and effectiveness of tests. The American College of Physicians. *Ann. Intern. Med.* **126,** 394–406 (1997).

123. Catalona, W. J., Smith, D. S., Ratliff, T. L. & Basler, J. W. Detection of organ-confined prostate cancer is increased through prostate-specific antigen-based screening. *JAMA* **270,** 948–54 (1993).

124. Shao, Y.-H. *et al.* Risk profiles and treatment patterns among men diagnosed as having prostate cancer and a prostate-specific antigen level below 4.0 ng/ml. *Arch. Intern. Med.* **170,** 1256–61 (2010).

125. Wilt, T. J. *et al.* Systematic review: comparative effectiveness and harms of treatments for clinically localized prostate cancer. *Ann. Intern. Med.* **148,** 435–48 (2008).

126. Wilt, T. J. *et al.* Radical prostatectomy versus observation for localized prostate cancer. *N. Engl. J. Med.* **367,** 203–13 (2012).

127. Bill-Axelson, A. *et al.* Radical prostatectomy versus watchful waiting in localized prostate cancer: the Scandinavian prostate cancer group-4 randomized trial. *J. Natl. Cancer Inst.* **100,** 1144–54 (2008).

128. Vickers, A. J. *et al.* Empirical estimates of prostate cancer overdiagnosis by age and prostate-specific antigen. *BMC Med.* **12,** 26 (2014).

129. Check, E. Proteomics and cancer: running before we can walk? *Nature* **429,** 496–7 (2004).

130. Lu, J., Sharma, L. K. & Bai, Y. Implications of mitochondrial DNA mutations and mitochondrial dysfunction in tumorigenesis. *Cell Res.* **19,** 802–15 (2009).

131. Babcock, D. F. & Hille, B. Mitochondrial oversight of cellular Ca2+ signaling. *Curr. Opin. Neurobiol.* **8,** 398–404 (1998).

132. Rizzuto, R., Bernardi, P. & Pozzan, T. Mitochondria as all-round players of the calcium game. *J. Physiol.* **529 Pt 1,** 37–47 (2000).

133. Ajioka, R. S., Phillips, J. D. & Kushner, J. P. Biosynthesis of heme in mammals. *Biochim. Biophys. Acta* **1763,** 723–36 (2006).

134. Kroemer, G. & Reed, J. C. Mitochondrial control of cell death. *Nat. Med.* **6,** 513–9 (2000).

135. Lill, R. & Mühlenhoff, U. Maturation of iron-sulfur proteins in eukaryotes: mechanisms, connected processes, and diseases. *Annu. Rev. Biochem.* **77,** 669–700 (2008).

136. Rustin, P. Mitochondria, from cell death to proliferation. *Nat. Genet.* **30,** 352–3 (2002).

137. Wang, X. The expanding role of mitochondria in apoptosis. *Genes Dev.* **15,** 2922–33 (2001).

138. Anderson, S. *et al.* Sequence and organization of the human mitochondrial genome. *Nature* **290,** 457–65 (1981).

139. Andrews, R. M. *et al.* Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nat. Genet.* **23,** 147 (1999).

140. Robin, E. D. & Wong, R. Mitochondrial DNA molecules and virtual number of mitochondria per cell in mammalian cells. *J. Cell. Physiol.* **136,** 507–13 (1988).

141. Satoh, M. & Kuroiwa, T. Organization of multiple nucleoids and DNA molecules in mitochondria of a human cell. *Exp. Cell Res.* **196,** 137–40 (1991).

142. Wai, T. *et al.* The role of mitochondrial DNA copy number in mammalian fertility. *Biol. Reprod.* **83,** 52–62 (2010).

143. Díez-Sánchez, C. *et al.* Mitochondrial DNA content of human spermatozoa. *Biol. Reprod.* **68,** 180–5 (2003).

144. Clay Montier, L. L., Deng, J. J. & Bai, Y. Number matters: control of mammalian mitochondrial DNA copy number. *J. Genet. Genomics* **36,** 125–31 (2009).

145. Morita, M. *et al.* mTORC1 controls mitochondrial activity and biogenesis through 4E-BP-dependent translational regulation. *Cell Metab.* **18,** 698–711 (2013).

146. El-Hattab, A. W. & Scaglia, F. Mitochondrial DNA depletion syndromes: review and updates of genetic basis, manifestations, and therapeutic options. *Neurotherapeutics* **10,** 186–98 (2013).

147. Wu, C.-W. *et al.* Mitochondrial DNA mutations and mitochondrial DNA depletion in gastric cancer. *Genes. Chromosomes Cancer* **44,** 19–28 (2005).

148. Selvanayagam, P. & Rajaraman, S. Detection of mitochondrial genome depletion by a novel cDNA in renal cell carcinoma. *Lab. Invest.* **74,** 592–9 (1996).

149. Xing, J. *et al.* Mitochondrial DNA content: its genetic heritability and association with renal cell carcinoma. *J. Natl. Cancer Inst.* **100,** 1104–12 (2008).

150. Yin, P. H. *et al.* Alteration of the copy number and deletion of mitochondrial DNA in human hepatocellular carcinoma. *Br. J. Cancer* **90,** 2390–6 (2004).

151. Mizumachi, T. *et al.* Increased distributional variance of mitochondrial DNA content associated with prostate cancer cells as compared with normal prostate cells. *Prostate* **68,** 408–17 (2008).

152. Higuchi, M. Regulation of mitochondrial DNA content and cancer. *Mitochondrion* **7,** 53–7 (2007).

153. Melov, S., Shoffner, J. M., Kaufman, A. & Wallace, D. C. Marked increase in the number and variety of mitochondrial DNA rearrangements in aging human skeletal muscle. *Nucleic Acids Res.* **23,** 4122–6 (1995).

154. Bender, A. *et al.* High levels of mitochondrial DNA deletions in substantia nigra neurons in aging and Parkinson disease. *Nat. Genet.* **38,** 515–7 (2006).

155. Kraytsberg, Y. *et al.* Mitochondrial DNA deletions are abundant and cause functional impairment in aged human substantia nigra neurons. *Nat. Genet.* **38,** 518–20 (2006).

156. Sciacco, M., Bonilla, E., Schon, E. A., DiMauro, S. & Moraes, C. T. Distribution of wild-type and common deletion forms of mtDNA in normal and respiration-deficient muscle fibers from patients with mitochondrial myopathy. *Hum. Mol. Genet.* **3,** 13–9 (1994).

157. Bua, E. *et al.* Mitochondrial DNA-deletion mutations accumulate intracellularly to detrimental levels in aged human skeletal muscle fibers. *Am. J. Hum. Genet.* **79,** 469–80 (2006).

158. Moraes, C. T. *et al.* Mitochondrial DNA deletions in progressive external ophthalmoplegia

and Kearns-Sayre syndrome. *N. Engl. J. Med.* **320,** 1293–9 (1989).

159. Lee, H. C., Pang, C. Y., Hsu, H. S. & Wei, Y. H. Differential accumulations of 4,977 bp deletion in mitochondrial DNA of various tissues in human ageing. *Biochim. Biophys. Acta* **1226,** 37–43 (1994).

160. Krishnan, K. J. *et al.* What causes mitochondrial DNA deletions in human cells? *Nat. Genet.* **40,** 275–9 (2008).

161. Yasukawa, T. *et al.* Replication of vertebrate mitochondrial DNA entails transient ribonucleotide incorporation throughout the lagging strand. *EMBO J.* **25,** 5358–71 (2006).

162. Robberson, D. L. & Clayton, D. A. Replication of mitochondrial DNA in mouse L cells and their thymidine kinase - derivatives: displacement replication on a covalently-closed circular template. *Proc. Natl. Acad. Sci. U. S. A.* **69,** 3810–4 (1972).

163. Holt, I. J., Lorimer, H. E. & Jacobs, H. T. Coupled leading- and lagging-strand synthesis of mammalian mitochondrial DNA. *Cell* **100,** 515–24 (2000).

164. Taylor, R. W. & Turnbull, D. M. Mitochondrial DNA mutations in human disease. *Nat. Rev. Genet.* **6,** 389–402 (2005).

165. Boveris, A. Mitochondrial production of superoxide radical and hydrogen peroxide. *Adv. Exp. Med. Biol.* **78,** 67–82 (1977).

166. Halliwell, B. & Gutteridge, J. M. Free radicals, lipid peroxidation, and cell damage. *Lancet (London, England)* **2,** 1095 (1984).

167. Brand, M. D. The sites and topology of mitochondrial superoxide production. *Exp. Gerontol.* **45,** 466–72 (2010).

168. Roede, J. R. & Jones, D. P. Reactive species and mitochondrial dysfunction: mechanistic significance of 4-hydroxynonenal. *Environ. Mol. Mutagen.* **51,** 380–90 (2010).

169. Cadet, J., Douki, T. & Ravanat, J.-L. Oxidatively generated base damage to cellular DNA. *Free Radic. Biol. Med.* **49,** 9–21 (2010).

170. De Bont, R. & van Larebeke, N. Endogenous DNA damage in humans: a review of quantitative data. *Mutagenesis* **19,** 169–85 (2004).

171. Liu, P. & Demple, B. DNA repair in mammalian mitochondria: Much more than we thought? *Environ. Mol. Mutagen.* **51,** 417–26 (2010).

172. Wallace, D. C. Mitochondrial DNA sequence variation in human evolution and disease. *Proc. Natl. Acad. Sci. U. S. A.* **91,** 8739–46 (1994).

173. HARMAN, D. Aging: a theory based on free radical and radiation chemistry. *J. Gerontol.* **11,** 298–300 (1956).

174. Jackson, A. L. & Loeb, L. A. The contribution of endogenous sources of DNA damage to the multiple mutations in cancer. *Mutat. Res.* **477,** 7–21 (2001).

175. Benhar, M., Engelberg, D. & Levitzki, A. ROS, stress-activated kinases and stress signaling in cancer. *EMBO Rep.* **3,** 420–5 (2002).

176. Kong, Q., Beel, J. A. & Lillehei, K. O. A threshold concept for cancer therapy. *Med. Hypotheses* **55,** 29–35 (2000).

177. Burdon, R. H. Superoxide and hydrogen peroxide in relation to mammalian cell proliferation. *Free Radic. Biol. Med.* **18,** 775–94 (1995).

178. Laurent, A. *et al.* Controlling tumor growth by modulating endogenous production of reactive oxygen species. *Cancer Res.* **65,** 948–56 (2005).

179. Sauer, H., Wartenberg, M. & Hescheler, J. Reactive oxygen species as intracellular messengers during cell growth and differentiation. *Cell. Physiol. Biochem.* **11,** 173–86 (2001).

180. Naka, K., Muraguchi, T., Hoshii, T. & Hirao, A. Regulation of Reactive Oxygen Species and Genomic Stability in Hematopoietic Stem Cells. *Antioxid. Redox Signal.* **10,** 1883–1894 (2008).

181. Lambeth, J. D. Nox enzymes, ROS, and chronic disease: an example of antagonistic pleiotropy. *Free Radic. Biol. Med.* **43,** 332–47 (2007).

182. Khandrika, L., Kumar, B., Koul, S., Maroni, P. & Koul, H. K. Oxidative stress in prostate cancer. *Cancer Lett.* **282,** 125–36 (2009).

183. Brandon, M., Baldi, P. & Wallace, D. C. Mitochondrial mutations in cancer. *Oncogene* **25,** 4647–62 (2006).

184. Guerra, F. *et al.* Placing mitochondrial DNA mutations within the progression model of type I endometrial carcinoma. *Hum. Mol. Genet.* **20,** 2394–405 (2011).

185. Lim, S. W. *et al.* High-frequency minisatellite instability of the mitochondrial genome in colorectal cancer tissue associated with clinicopathological values. *Int. J. Cancer* **131,** 1332–41 (2012).

186. Tseng, L.-M. *et al.* Somatic mutations of the mitochondrial genome in human breast cancers. *Genes. Chromosomes Cancer* **50,** 800–11 (2011).

187. Wallace, D. C. Mitochondria and cancer. *Nat. Rev. Cancer* **12,** 685–98 (2012).

188. Chen, J. Z., Gokden, N., Greene, G. F., Mukunyadzi, P. & Kadlubar, F. F. Extensive somatic mitochondrial mutations in primary prostate cancer using laser capture microdissection. *Cancer Res.* **62,** 6470–4 (2002).

189. Parr, R. L. *et al.* Somatic mitochondrial DNA mutations in prostate cancer and normal appearing adjacent glands in comparison to age-matched prostate samples without malignant histology. *J. Mol. Diagn.* **8,** 312–9 (2006).

190. Petros, J. A. *et al.* mtDNA mutations increase tumorigenicity in prostate cancer. *Proc. Natl. Acad. Sci. U. S. A.* **102,** 719–24 (2005).

191. Arbiser, J. L. Molecular regulation of angiogenesis and tumorigenesis by signal transduction pathways: evidence of predictable and reproducible patterns of synergy in diverse neoplasms. *Semin. Cancer Biol.* **14,** 81–91 (2004).

192. Sotgia, F., Martinez-Outschoorn, U. E. & Lisanti, M. P. Mitochondrial oxidative stress drives tumor progression and metastasis: should we use antioxidants as a key component of cancer treatment and prevention? *BMC Med.* **9,** 62 (2011).

193. Pavlides, S. *et al.* The autophagic tumor stroma model of cancer: Role of oxidative stress and ketone production in fueling tumor cell metabolism. *Cell Cycle* **9,** 3485–505 (2010).

194. Warburg, O., Wind, F. & Negelein, E. THE METABOLISM OF TUMORS IN THE BODY. *J. Gen. Physiol.* **8,** 519–30 (1927).

195. WARBURG, O. On the origin of cancer cells. *Science* **123,** 309–14 (1956).

196. Kavlick, M. F. *et al.* Quantification of human mitochondrial DNA using synthesized DNA standards. *J. Forensic Sci.* **56,** 1457–63 (2011).

197. Arya, M. *et al.* Basic principles of real-time quantitative PCR. *Expert Rev. Mol. Diagn.* **5,** 209–19 (2005).

198. Meissner, C. *et al.* Quantification of mitochondrial DNA in human blood cells using an automated detection system. *Forensic Sci. Int.* **113,** 109–12 (2000).

199. Timken, M. D., Swango, K. L., Orrego, C. & Buoncristiani, M. R. A duplex real-time qPCR assay for the quantification of human nuclear and mitochondrial DNA in forensic samples: implications for quantifying DNA in degraded samples. *J. Forensic Sci.* **50,** 1044–60 (2005).

200. Andréasson, H., Nilsson, M., Budowle, B., Lundberg, H. & Allen, M. Nuclear and mitochondrial DNA quantification of various forensic materials. *Forensic Sci. Int.* **164,** 56–64 (2006).

201. Niederstätter, H. *et al.* A modular real-time PCR concept for determining the quantity and quality of human nuclear and mitochondrial DNA. *Forensic Sci. Int. Genet.* **1,** 29–34 (2007).

202. FBI — Standards and Guidelines - Forensic Science Communications - October 2008. at <https://www.fbi.gov/about-us/lab/forensic-science-communications/fsc/oct2008/standards/2008_10_standards01.htm/2008_10_standards

01b.htm>

203. Sprouse, M. L., Phillips, N. R., Kavlick, M. F. & Roby, R. K. Internal Validation of Human Mitochondrial DNA Quantification Using Real-Time PCR. *J. Forensic Sci.* **59,** 1049–1056 (2014).

204. Maniatis, T., Sambrook, J. & Fritsch, E. F. *Molecular cloning: A laboratory manual*. (Cold Spring Harbor Laboratory, 1982).

205. Green, R. L., Roinestad, I. C., Boland, C. & Hennessy, L. K. Developmental validation of the quantifiler real-time PCR kits for the quantification of human nuclear DNA samples. *J. Forensic Sci.* **50,** 809–25 (2005).

206. Schmittgen, T. D. *et al.* Quantitative reverse transcription-polymerase chain reaction to study mRNA decay: comparison of endpoint and real-time methods. *Anal. Biochem.* **285,** 194–204 (2000).

207. Mutter, G. L. & Boynton, K. A. PCR bias in amplification of androgen receptor alleles, a trinucleotide repeat marker used in clonality studies. *Nucleic Acids Res.* **23,** 1411–8 (1995).

208. Polz, M. F. & Cavanaugh, C. M. Bias in template-to-product ratios in multitemplate PCR. *Appl. Environ. Microbiol.* **64,** 3724–30 (1998).

209. Walsh, P. S., Erlich, H. A. & Higuchi, R. Preferential PCR amplification of alleles: mechanisms and solutions. *PCR Methods Appl.* **1,** 241–50 (1992).

210. Opel, K. L., Chung, D. & McCord, B. R. A study of PCR inhibition mechanisms using real time PCR. *J. Forensic Sci.* **55,** 25–33 (2010).

211. He, Y.-H. *et al.* Mitochondrial DNA content contributes to healthy aging in Chinese: a study from nonagenarians and centenarians. *Neurobiol. Aging* **35,** 1779.e1–4 (2014).

212. Mengel-From, J. *et al.* Mitochondrial DNA copy number in peripheral blood cells declines with age and is associated with general health among elderly. *Hum. Genet.* **133,** 1149–59 (2014).

213. Zhou, W. *et al.* Peripheral blood mitochondrial DNA copy number is associated with prostate cancer risk and tumor burden. *PLoS One* **9,** e109470 (2014).

214. Simonnet, H. *et al.* Low mitochondrial respiratory chain content correlates with tumor aggressiveness in renal cell carcinoma. *Carcinogenesis* **23,** 759–68 (2002).

215. Yu, M. *et al.* Reduced mitochondrial DNA copy number is correlated with tumor progression and prognosis in Chinese breast cancer patients. *IUBMB Life* **59,** 450–7 (2007).

216. Moro, L., Arbini, A. A., Marra, E. & Greco, M. Mitochondrial DNA depletion reduces PARP-1 levels and promotes progression of the neoplastic phenotype in prostate carcinoma. *Cell. Oncol.* **30,** 307–22 (2008).

217. Moro, L. *et al.* Mitochondrial DNA depletion in prostate epithelial cells promotes anoikis resistance and invasion through activation of PI3K/Akt2. *Cell Death Differ.* **16,** 571–83 (2009).

218. Dimberg, J. *et al.* Common 4977 bp deletion and novel alterations in mitochondrial DNA in Vietnamese patients with breast cancer. *Springerplus* **4,** 58 (2015).

219. Rogounovitch, T. I. *et al.* Large deletions in mitochondrial DNA in radiation-associated human thyroid tumors. *Cancer Res.* **62,** 7031–41 (2002).

220. Abnet, C. C. *et al.* Control region mutations and the 'common deletion' are frequent in the mitochondrial DNA of patients with esophageal squamous cell carcinoma. *BMC Cancer* **4,** 30 (2004).

221. Chen, T. *et al.* The mitochondrial DNA 4,977-bp deletion and its implication in copy number alteration in colorectal cancer. *BMC Med. Genet.* **12,** 8 (2011).

222. Ye, C. *et al.* Quantitative analysis of mitochondrial DNA 4977-bp deletion in sporadic breast cancer and benign breast diseases. *Breast Cancer Res. Treat.* **108,** 427–34 (2008).

223. Jessie, B. C. *et al.* Accumulation of mitochondrial DNA deletions in the malignant prostate of patients of different ages. *Exp. Gerontol.* **37,** 169–74 (2001).

224. Maki, J. *et al.* Mitochondrial genome deletion aids in the identification of false- and true-negative prostate needle core biopsy specimens. *Am. J. Clin. Pathol.* **129,** 57–66 (2008).

225. Robinson, K. *et al.* Accurate prediction of repeat prostate biopsy outcomes by a mitochondrial DNA deletion assay. *Prostate Cancer Prostatic Dis.* **13,** 126–31 (2010).

226. Parson, W. & Dür, A. EMPOP--a forensic mtDNA database. *Forensic Sci. Int. Genet.* **1,** 88–92 (2007).

227. Malik, A. N., Shahni, R., Rodriguez-de-Ledesma, A., Laftah, A. & Cunningham, P. Mitochondrial DNA as a non-invasive biomarker: accurate quantification using real time quantitative PCR without co-amplification of pseudogenes and dilution bias. *Biochem. Biophys. Res. Commun.* **412,** 1–7 (2011).

228. Wolf, A. M. D. *et al.* American Cancer Society guideline for the early detection of prostate cancer: update 2010. *CA. Cancer J. Clin.* **60,** 70–98

229. Chinnery, P. F., Samuels, D. C., Elson, J. & Turnbull, D. M. Accumulation of mitochondrial DNA mutations in ageing, cancer, and mitochondrial disease: is there a common mechanism? *Lancet (London, England)* **360,** 1323–5 (2002).

230. Copeland, W. C., Wachsman, J. T., Johnson, F. M. & Penta, J. S. Mitochondrial DNA alterations in cancer. *Cancer Invest.* **20,** 557–69 (2002).

231. Gasparre, G. *et al.* Clonal expansion of mutated mitochondrial DNA is associated with tumor formation and complex I deficiency in the benign renal oncocytoma. *Hum. Mol. Genet.* **17,** 986–95 (2008).

232. Bartoletti-Stella, A. Mitochondrial DNA Mutations in Oncocytic Adnexal Lacrimal Glands of the Conjunctiva. *Arch. Ophthalmol.* **129,** 664 (2011).

233. Pereira, L., Soares, P., Máximo, V. & Samuels, D. C. Somatic mitochondrial DNA mutations in cancer escape purifying selection and high pathogenicity mutations lead to the oncocytic phenotype: pathogenicity analysis of reported somatic mtDNA mutations in tumors. *BMC Cancer* **12,** 53 (2012).

234. Rozen, S. & Skaletsky, H. Primer3 on the WWW for general users and for biologist

programmers. *Methods Mol. Biol.* **132,** 365–86 (2000).

235. Kumar, M., Tanwar, M., Saxena, R., Sharma, P. & Dada, R. Identification of novel mitochondrial mutations in Leber's hereditary optic neuropathy. *Mol. Vis.* **16,** 782–92 (2010).

236. Roby, R. K. *et al.* Mitochondrial genome Interrogation for forensic casework and research studies. *Curr. Protoc. Hum. Genet.* (2014).

237. Kloss-Brandstätter, A. *et al.* HaploGrep: a fast and reliable algorithm for automatic classification of mitochondrial DNA haplogroups. *Hum. Mutat.* **32,** 25–32 (2011).

238. van Oven, M. & Kayser, M. Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Hum. Mutat.* **30,** E386–94 (2009).

239. Goujon, M. *et al.* A new bioinformatics analysis tools framework at EMBL-EBI. *Nucleic Acids Res.* **38,** W695–W699 (2010).

240. Kogelnik, A. M., Lott, M. T., Brown, M. D., Navathe, S. B. & Wallace, D. C. MITOMAP: a human mitochondrial genome database. *Nucleic Acids Res.* **24,** 177–9 (1996).

241. Lott, M. T. *et al.* mtDNA Variation and Analysis Using MITOMAP and MITOMASTER. *Curr. Protoc. Bioinformatics* **1,** 1.23.1–1.23.26 (2013).

242. Adzhubei, I. A. *et al.* A method and server for predicting damaging missense mutations. *Nat. Methods* **7,** 248–9 (2010).

243. Ferrer-Costa, C. *et al.* PMUT: a web-based tool for the annotation of pathological mutations on proteins. *Bioinformatics* **21,** 3176–8 (2005).

244. Combet, C., Jambon, M., Deléage, G. & Geourjon, C. Geno3D: automatic comparative molecular modelling of protein. *Bioinformatics* **18,** 213–4 (2002).

245. Sayle, R. A. & Milner-White, E. J. RASMOL: biomolecular graphics for all. *Trends Biochem. Sci.* **20,** 374 (1995).

246. Biffi, A. *et al.* Principal-component analysis for assessment of population stratification in mitochondrial medical genetics. *Am. J. Hum. Genet.* **86,** 904–17 (2010).

247. Ziegler, A., König, I. R. & Thompson, J. R. Biostatistical Aspects of Genome-Wide Association Studies. *Biometrical J.* **50,** 8–28 (2008).

248. Frousios, K., Iliopoulos, C. S., Schlitt, T. & Simpson, M. A. Predicting the functional consequences of non-synonymous DNA sequence variants--evaluation of bioinformatics tools and development of a consensus strategy. *Genomics* **102,** 223–8 (2013).

249. Williams, S. *Analysis of in silico tools for evaluating missense variants*. (2012). at <http://www.ngrl.org.uk/Manchester/sites/default/files/publications/Informatics/ToolAnalyses/Missense_Prediction_Tool_Report.pdf>

250. Pace, C. N. & Scholtz, J. M. A helix propensity scale based on experimental studies of peptides and proteins. *Biophys. J.* **75,** 422–7 (1998).

251. Jerónimo, C. *et al.* Mitochondrial mutations in early stage prostate cancer and bodily fluids. *Oncogene* **20,** 5195–8 (2001).

252. Gómez-Zaera, M. *et al.* Identification of somatic and germline mitochondrial DNA sequence variants in prostate cancer patients. *Mutat. Res.* **595,** 42–51 (2006).

253. Paez, J. G. *et al.* Genome coverage and sequence fidelity of phi29 polymerase-based multiple strand displacement whole genome amplification. *Nucleic Acids Res.* **32,** e71 (2004).

254. REPLI-g Mitochondrial DNA Kit Handbook. (2013).

255. Gunnarsdóttir, E. D., Li, M., Bauchet, M., Finstermeier, K. & Stoneking, M. High-throughput sequencing of complete human mtDNA genomes from the Philippines. *Genome Res.* **21,** 1–11 (2011).

256. PrimeSTAR GXL DNA Polymerase Product Manual.

257. Qubit dsDNA BR Assay Kits. (2011).

258. Qubit 2.0 Fluorometer User Manual. (2010).

259. Pearson, W. R. & Lipman, D. J. Improved tools for biological sequence comparison. *Proc. Natl. Acad. Sci. U. S. A.* **85,** 2444–8 (1988).

260. Li, H., Ruan, J. & Durbin, R. Mapping short DNA sequencing reads and calling variants using mapping quality scores. *Genome Res.* **18,** 1851–8 (2008).

261. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25,** 1754–60 (2009).

262. King, J. L., Sajantila, A. & Budowle, B. mitoSAVE: mitochondrial sequence analysis of variants in Excel. *Forensic Sci. Int. Genet.* **12,** 122–5 (2014).

263. Ruiz-Pesini, E. *et al.* An enhanced MITOMAP with a global mtDNA mutational phylogeny. *Nucleic Acids Res.* **35,** D823–8 (2007).

264. Parson, W. *et al.* Evaluation of next generation mtGenome sequencing using the Ion Torrent Personal Genome Machine (PGM). *Forensic Sci. Int. Genet.* **7,** 543–549 (2013).

265. Diroma, M. A. *et al.* Extraction and annotation of human mitochondrial genomes from 1000 Genomes Whole Exome Sequencing data. *BMC Genomics* **15 Suppl 3,** S2 (2014).

266. Brandt, U. Energy converting NADH:quinone oxidoreductase (complex I). *Annu. Rev. Biochem.* **75,** 69–92 (2006).

267. Temperley, R., Richter, R., Dennerlein, S., Lightowlers, R. N. & Chrzanowska-Lightowlers, Z. M. Hungry codons promote frameshifting in human mitochondrial ribosomes. *Science* **327,** 301 (2010).

268. Arnould, T. *et al.* CREB activation induced by mitochondrial dysfunction is a new signaling pathway that impairs cell proliferation. *EMBO J.* **21,** 53–63 (2002).

269. Wallace, D. C. Mitochondria as chi. *Genetics* **179,** 727–35 (2008).

270. Clayton, D. A. & Vinograd, J. Circular dimer and catenate forms of mitochondrial DNA in human leukaemic leucocytes. *Nature* **216,** 652–7 (1967).

271. Clayton, D. A. & Vinograd, J. Complex mitochondrial DNA in leukemic and normal human myeloid cells. *Proc. Natl. Acad. Sci. U. S. A.* **62,** 1077–84 (1969).

272. Baysal, B. E. *et al.* Mutations in SDHD, a mitochondrial complex II gene, in hereditary paraganglioma. *Science* **287,** 848–51 (2000).

273. Astuti, D. *et al.* Gene Mutations in the Succinate Dehydrogenase Subunit SDHB Cause Susceptibility to Familial Pheochromocytoma and to Familial Paraganglioma. *Am. J. Hum. Genet.* **69,** 49–54 (2001).

274. Tomlinson, I. P. M. *et al.* Germline mutations in FH predispose to dominantly inherited uterine fibroids, skin leiomyomata and papillary renal cell cancer. *Nat. Genet.* **30,** 406–10 (2002).

275. Chatterjee, A., Mambo, E. & Sidransky, D. Mitochondrial DNA mutations in human cancer. *Oncogene* **25,** 4663–74 (2006).

276. Iommarini, L., Calvaruso, M. A., Kurelac, I., Gasparre, G. & Porcelli, A. M. Complex I impairment in mitochondrial diseases and cancer: parallel roads leading to different outcomes. *Int. J. Biochem. Cell Biol.* **45,** 47–63 (2013).

277. Yu, M. Somatic mitochondrial DNA mutations in human cancers. *Adv. Clin. Chem.* **57,** 99–138 (2012).

278. Sharma, L. K., Lu, J. & Bai, Y. Mitochondrial respiratory complex I: structure, function and implication in human diseases. *Curr. Med. Chem.* **16,** 1266–77 (2009).

279. Lesnefsky, E. J., Moghaddas, S., Tandler, B., Kerner, J. & Hoppel, C. L. Mitochondrial dysfunction in cardiac disease: ischemia--reperfusion, aging, and heart failure. *J. Mol. Cell. Cardiol.* **33,** 1065–89 (2001).

280. Wallace, D. C. Mitochondrial defects in cardiomyopathy and neuromuscular disease. *Am. Heart J.* **139,** S70–85 (2000).

281. Sugioka, K. *et al.* Mechanism of O2- generation in reduction and oxidation cycle of ubiquinones in a model of mitochondrial electron transport systems. *Biochim. Biophys. Acta* **936,** 377–85 (1988).

282. Turrens, J. F. & Boveris, A. Generation of superoxide anion by the NADH dehydrogenase of bovine heart mitochondria. *Biochem. J.* **191,** 421–7 (1980).

283. Liu, Y., Fiskum, G. & Schubert, D. Generation of reactive oxygen species by the mitochondrial electron transport chain. *J. Neurochem.* **80,** 780–7 (2002).

284. Kussmaul, L. & Hirst, J. The mechanism of superoxide production by NADH:ubiquinone oxidoreductase (complex I) from bovine heart mitochondria. *Proc. Natl. Acad. Sci. U. S. A.* **103,** 7607–12 (2006).

285. Hirst, J., King, M. S. & Pryde, K. R. The production of reactive oxygen species by complex I. *Biochem. Soc. Trans.* **36,** 976–80 (2008).

286. Adam-Vizi, V. & Chinopoulos, C. Bioenergetics and the formation of mitochondrial reactive oxygen species. *Trends Pharmacol. Sci.* **27,** 639–45 (2006).

287. Kudin, A. P., Bimpong-Buta, N. Y.-B., Vielhaber, S., Elger, C. E. & Kunz, W. S. Characterization of superoxide-producing sites in isolated brain mitochondria. *J. Biol. Chem.* **279,** 4127–35 (2004).

288. Votyakova, T. V & Reynolds, I. J. DeltaPsi(m)-Dependent and -independent production of reactive oxygen species by rat brain mitochondria. *J. Neurochem.* **79,** 266–77 (2001).

289. Treberg, J. R., Quinlan, C. L. & Brand, M. D. Evidence for two sites of superoxide production by mitochondrial NADH-ubiquinone oxidoreductase (complex I). *J. Biol. Chem.* **286,** 27103–10 (2011).

290. Lee, D., Redfern, O. & Orengo, C. Predicting protein function from sequence and structure. *Nat. Rev. Mol. Cell Biol.* **8,** 995–1005 (2007).

291. Martí-Renom, M. A. *et al.* Comparative protein structure modeling of genes and genomes. *Annu. Rev. Biophys. Biomol. Struct.* **29,** 291–325 (2000).

292. Baradaran, R., Berrisford, J. M., Minhas, G. S. & Sazanov, L. A. Crystal structure of the entire respiratory complex I. *Nature* **494,** 443–448 (2013).

293. Berman, H. M. The Protein Data Bank: a historical perspective. *Acta Crystallogr. A.* **64,** 88–95 (2008).

294. Berman, H. M. *et al.* The Protein Data Bank. *Nucleic Acids Res.* **28,** 235–42 (2000).

295. Li, W. *et al.* The EMBL-EBI bioinformatics web and programmatic tools framework. *Nucleic Acids Res.* **43,** W580–4 (2015).

296. Rice, P., Longden, I. & Bleasby, A. EMBOSS: The European Molecular Biology Open Software Suite. *Trends Genet.* **16,** 276–277 (2000).

297. McWilliam, H. *et al.* Analysis Tool Web Services from the EMBL-EBI. *Nucleic Acids Res.* **41,** W597–600 (2013).

298. Efremov, R. G., Baradaran, R. & Sazanov, L. A. The architecture of respiratory complex I. *Nature* **465,** 441–5 (2010).

299. Efremov, R. G. & Sazanov, L. A. Structure of the membrane domain of respiratory complex I. *Nature* **476,** 414–20 (2011).

300. Treberg, J. R. & Brand, M. D. A Model of the Proton Translocation Mechanism of Complex I. *J. Biol. Chem.* **286,** 17579–17584 (2011).

301. Lloyd, R. E. *et al.* Identification and functional prediction of mitochondrial complex III and IV mutations associated with glioblastoma. *Neuro. Oncol.* **17,** 942–52 (2015).

302. Biasini, M. *et al.* SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Res.* **42,** W252–8 (2014).

303. Haas, J. *et al.* The Protein Model Portal--a comprehensive resource for protein structure and model information. *Database (Oxford).* **2013,** bat031 (2013).