Title
Characterizing the Utility of Cell-free DNA in Prostate Cancer

Permalink
https://escholarship.org/uc/item/6jx911vq

Author
Chen, Emmalyn

Publication Date
2020

Peer reviewed|Thesis/dissertation

Characterizing the Utility of Cell-free DNA in Prostate Cancer

by
Emmalyn Chen

DISSERTATION
Submitted in partial satisfaction of the requirements for degree of
DOCTOR OF PHILOSOPHY

in

Pharmaceutical Sciences and Pharmacogenomics

in the

GRADUATE DIVISION
of the
UNIVERSITY OF CALIFORNIA, SAN FRANCISCO

Approved:

DocuSigned by:

John S. Witte
5BC721F504504AE...

John S. Witte

Chair

DocuSigned by:

Pamela Paris

Pamela Paris

DocuSigned by:

Felix Feng
A5CABA754AC449B...

Felix Feng

Committee Members

DEDICATION

To my mom, who taught me to be curious and independent. To my dad, who fostered my creativity and diligence. And to Aiden, whose constant companionship I could never repay, even with a lifetime of cat treats.

ACKNOWLEDGMENTS

You were always invested in my work, and provided thoughtful feedback that undoubtedly strengthened my results. Felix, you have always been my biggest supporter, cheering me on at every step. Your lab was a second home where I could run my experiments in the company of some of the smartest and most diligent people I've met, and even though your schedule was packed, you always made time and made sure I had what I needed to succeed. It always brightened my day when I ran into both of you in the hallways of Diller, and I will miss our impromptu chats greatly.

Of course, this project never would have happened without **Clint Cario**, who spearheaded the cfDNA effort in the lab and set up everything, and I mean *everything*. From the coordination of clinical sample collection, to identifying the optimal protocols for extraction and the best strategies for sequencing, all while thinking through the details in between. You are a superhero. Not only do you have the prowess that comes with years of training in computer science, biochemistry, and bioinformatics, your process is thoughtful and creative, and is guided by your intuition and ability to think both deeply and abstractly. We've been through late nights at the office and early mornings at the hospital. We've talked about the multiverse (peace among worlds), what it means to be alive, and how to figure out if your cat loves or hates you. I'm going to miss you, man. Thank you for everything.

Deanna Kroetz, you were our program mom and were always there for me when I needed you most. **Kathy Giacomini**, you gave me the flexibility to work with John during my first rotation, encouraging collaboration and clear communication. **Marina Sirota**, you showed me the

importance of understanding all the details of the data, and gave me the amazing opportunity to intern at Parker Institute for Cancer Immunotherapy.

David Breslauer and **Dan Widmaier**, without Bolt Threads and your support I may never have pursued a PhD at UCSF. The one-year internship you provided was first of its kind, with the specific goal of helping recent college graduates transition to graduate school. Your encouragement and faith propelled me forward, and for that I am grateful.

Pier Federico Gherardini, Danny Wells, and **Lacey Kitch**, thank you for showing me what it was like to be a part of a brilliant informatics team. Federico, I learned so much from you. You showed me the importance of being direct and making a decision, yet remaining open and flexible to potential solutions. Danny, you walked me through setting up and running pipelines on Google Cloud with Terra, definitely expediting my PhD by a few months. Lacey, you always brought up important technical considerations and probed the biological underpinnings of my results, helping me to think more deeply about my work.

Linda Kachuri and **Taylor Cavazos**, you two have been there for the ups and downs of daily life, keeping things light when they were heavy, and cheerful when they were dreary. Taylor, you are my favorite by the book person, and without you I would still be setting up virtual environments incorrectly. I'm going to miss our killer F45 workouts and dramatic succulent replanting. Linda, you are the most stylish and elegant of all postdocs, and without you I would still be writing cover letters without a fancy UCSF letterhead. I am so proud of you for so many reasons, including receiving a huge grant and pursuing significant collaborations – I can't wait to follow your career

and see where you're headed next! **Nima Emami**, you always believed in me. You always indulged me whenever I was stuck thinking through some concept, whether it was the specific details of sequencing by synthesis, or dissecting the math underlying statistical tests. You have the best memory, and the ability to draw people in and listen. I am so proud of you for starting a company, and am so excited to follow your work in this incredibly important space of immuno-oncology.

My PSPG family, without you classes and retreats would have been unbearably lonely. **Marjorie Imperial** and **Liz Levy**, you were my first friends at UCSF. Marjorie, it has been so amazing to watch you crush it at work and grow your family with David. Liz, our love for dogs and all animals will forever bond us. **Hayley Donnella,** you are my soul sister for all things career and design. **Kat Chua**, I can't wait to start cooking with you and doing code reviews together again. **Deanna Brackman**, you took me in my first year and showed me all things PSPG. **Nina Gonzaludo**, we conquered our fears and you are my forever scuba buddy.

To my best friends away from graduate school, I am *so sorry* for neglecting you during this time. Thank you for being patient with me. **Ginna Oates** and **Khoury Ibrahim**, you are my rocks. Ginna, I can always count on you to bring the brightest, most positive energy to the table, and to be nurturing in times of distress. You've shown me how to love more openly and with conviction. Khoury, it's been so amazing to see you grow with Julian, and so energizing to see you excel at every task you set your mind to. I will never forget our time in Japan. **Camille Ong**, we will forever be connected by our fascination with the human condition, and I miss our shenanigans that always end up with us laughing deliriously for no reason. **Richard Yuan**, you are the brother I never had.

Our love for cats and all things food is eternal, and without you I may have never gone on a backpacking trip or cooked trout caught in the alpine lakes. In times of extreme physical sadness, you taught me discipline and mental fortitude. **Supada Sritanyaratana**, I miss our long chats over homemade desserts, and will always remember our time in Thailand. **Meena Nagappan**, you were my first friend in college. I miss our porch life with La Val's. **Marcus Gomez**, you are the little brother I never had, and I am so proud of you for all the heart you put into your work. Thank you for letting me tag along to Arizona and fishing with me.

Mitchell Nahmias, you've been the most unexpected surprise of 2020, even with a world-altering pandemic. You bring an enthusiasm that is unmatched, and a predilection for the cerebral. Without you, life would be incredibly dull.

And finally, to my parents, **Lea-Ying Amy Lin** and **Lung Albert Chen**, without you I actually wouldn't be here. Thank you for feeding me and taking care of me while I finished this dissertation. I owe it all to you.

If I have seen further,

it is by standing on the shoulders of giants.

— Isaac Newton

ABSTRACT

Characterizing the Utility of Cell-free DNA in Prostate Cancer

Emmalyn Chen

Motivation: Prostate cancer remains the most commonly diagnosed neoplasm in American men, with existing biomarkers (i.e. PSA, nomograms, MRI) having varying levels of sensitivity and specificity in identifying more advanced and potentially aggressive disease. Tumor tissue biopsies remain the gold standard for confirming the presence of prostate cancer, as well as evaluating the genomic heterogeneity and clonal architecture that may be predictive of poor outcomes (i.e. recurrence and metastasis). However, tissue biopsies are limited in their ability to comprehensively assess tumors, and may lead to underestimation of disease grade and stage. These hurdles may be overcome with cell-free DNA (cfDNA), which allows for minimally invasive, repeated sampling through blood draws. This is particularly important when tumor tissue is unavailable during active surveillance or disease monitoring for the detection of residual disease or progression. Additionally, genomic interrogation via cfDNA sequencing typically requires prior knowledge of existing mutations from a patient's tumor. The work presented here leverages a number of methods to ensure broad, yet sensitive detection of cfDNA variants for patients with localized prostate cancer, including sequencing with a machine-learning guided 2.5Mb targeted panel. In this dissertation, I investigate the use of cfDNA concentration, fragment size, and sequencing to identify advanced prostate cancer, as well as detect somatic mutations present in patient-matched tumors.

Methods: The patient cohort included in these studies are composed of 268 individuals: 34 healthy individuals, 112 men with localized prostate cancer who underwent radical prostatectomy (RP), and 122 men with metastatic castration-resistant prostate cancer (mCRPC). Plasma cfDNA

concentration and fragment size were quantified with a Qubit fluorometer or Bioanalyzer utilizing a chip-based capillary electrophoresis method for nucleic acid analysis. Low-pass whole-genome and targeted sequencing were used to identify single nucleotide variants (SNVs), small insertions and deletions (indels), and copy number alterations (CNAs) for a subset of patients. Plasma cfDNA was barcoded with duplex Unique Molecular Identifiers (UMIs) to construct consensus reads and improve variant detection by leveraging duplicate reads and sequence complementarity of the two DNA strands. Extensive tissue sampling was used to capture tumor heterogeneity and provide a patient-specific gold standard for comparison of matched cfDNA.

Results and Conclusions: Patients with advanced mCRPC had higher cfDNA concentration than men with localized disease or healthy controls, and those with localized disease had shorter average fragment sizes than controls. Importantly, cfDNA concentration and fragment size remained independent predictors after adjusting for age and PSA. We found that targeted sequencing of cfDNA—without *a priori* patient-specific tumor mutation information—identified somatic alterations found in matched tumor tissue from multiple regions, potentially allowing for dynamic monitoring of emerging resistant subclones throughout the course of disease. Detection of these concordant variants was associated with seminal vesicle invasion and the number of somatic variants found in the tumor tissue samples, predicating its use for patients with poor prognostic factors in a localized setting. Similar to cfDNA concentration, plasma cfDNA mutational burden was also found to increase with disease severity. The results from our studies demonstrate the ability of cfDNA to identify somatic variants in patients with heterogeneous, localized prostate cancer.

Keywords: cell-free DNA, prostate cancer, tumor heterogeneity, prognostic biomarker, next-generation sequencing, duplex UMIs, clonal evolution, targeted panel, whole-genome sequencing

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter I

Introduction

The advent of non-invasive cell-free DNA (cfDNA) analysis has ushered in a new era of cancer management, including potential applications in screening, early detection, treatment stratification, and disease monitoring. In the past decade, testing based on plasma cfDNA has emerged as a promising alternative to invasive, "gold standard" tests such as tumor tissue biopsies, and is currently being incorporated into the clinical management of cancer. However, available testing focuses on specific genetic mutations and only benefits a subset of patients due to technological and analytical limitations, sampling noise, and the ambiguity surrounding cancer-associated mutations. In this chapter, I will delve into the importance and complexity of human genetics, review the biological and clinical implications of cancer evolution, and detail the origins and applications of cfDNA. Additionally, I will highlight the convergence of multiple advances in the field that motivated our lab to investigate the use of cfDNA concentration, fragment size, and sequencing to identify localized and advanced prostate cancer, as well as detect somatic mutations present in patient-matched tumors.

1.1 Human genetics and next-generation sequencing

The diploid human genome consists of nearly 6.4 billion genetic building blocks, forming a double helix DNA structure, and serves as the set of instructions for development, survival, and reproduction. The information contained within the DNA sequence is first transcribed into RNA, and subsequently translated into proteins, which carry out the majority of all cell functions. While

all cells in the body share the majority of the same DNA, differences in gene expression are what result in the diversity of specialized tissue and organs that create a multicellular organism, with genetic alterations resulting in potentially severe consequences on downstream biological processes[1]. Specifically, carcinogenesis can occur when somatic mutations accumulate as a result of errors in DNA replication, which along with epigenetic changes can modify molecular pathways that allow for immune evasion, sustained angiogenesis for neoplastic vascularization, as well as limitless replicative potential and insensitivity to anti-growth signals. Somatic tumor mutations arise later in development and are acquired in an individual's tissues, which are distinct from germline mutations that are inherited and transmitted to offspring.

The implications of the human genome are vast, and yet the genome itself is compressed into a microscopic package, which eluded scientists for centuries. In the mid-1800's, the results from Gregor Mendel's study of pea plant breeding later established our understanding of trait transmission and inheritance, which has had a lasting impact on field of study ranging from inherited diseases and ancestry to modern day agriculture. But it wasn't until 1953 when James D. Watson and Francis Crick accurately described the molecular structure of DNA through the use of X-ray diffraction, paving the road for Watson's help in establishing the Human Genome Project in the early 1990's, an international 13-year effort to sequence the entire human genome[2].

Since the completion of the Human Genome Project in 2003, the cost of next-generation sequencing (NGS) has decreased dramatically and outpaced Moore's law with the creation of massively parallel sequencing processes. The first human genome ever sequenced cost nearly $100M, but today the cost to fully sequence an individual's genome is approximately $1,000[3–5].

The reduction in cost has accelerated genetics research, making it possible to elucidate the role of genetic variation in human disease by analyzing large cohorts of patients and healthy individuals for association studies, as well as discover tumor alterations that can be used as clinical biomarkers to inform cancer diagnosis and treatment.

1.2 Genomic heterogeneity and cancer evolution

Approximately 40% of men and women will be diagnosed with cancer at some point during their lifetime[6]. The collection of diseases we know as cancer can be initiated by both environmental risk factors (i.e. smoking) and inherited germline genetic factors (i.e. *BRCA1* or *BRCA2* mutations) that result in genetic alterations that disrupt the cell cycle, triggering unrestrained growth and proliferation. While 10% of cancer cases are attributed to alterations in a single gene, the overwhelming majority of cancers are considered complex genetic traits arising from multiple genetic variants[7].

Intratumor heterogeneity, which is driven by tumor evolution, is a critical component determining the lethality of cancer due to its impact on therapeutic failure, drug resistance, and disease progression. Multiple tumor cell subpopulations, defined by a diverse set of genetic mutations, undergo Darwinian evolution and allow for the exploitation of escape mechanisms for tumor survival. This process is subsequently accelerated when tumors undergo under sudden selective pressure induced by treatments, and in particular, targeted therapy[8].

1.3 Prostate cancer

Prostate cancer, a disease that accounts for approximately 20% of all new cancer diagnoses in American men, is known to harbor extensive tumor heterogeneity that includes subclones with large genetic divergence at the time of diagnosis[9–11]. While the majority of prostate cancers are diagnosed when the disease is still localized and are successfully treated, an estimated 2.9% of patients develop bone metastases and 2.4% die of the disease within six years[12]. Genomic instability resulting from recurrent copy number alterations (CNAs) in genes such as *MYC, NKX3-1,* and *PTEN* has been found to be prognostic for biochemical recurrence following surgery or radiotherapy[13,14]. Prostate tumors are also known to harbor single nucleotide variants (SNVs) in both coding and noncoding regions of the genome, with the most common nonsynonymous mutations in *SPOP* occurring in only 10% of primary prostate cancers, making detection of this genomic heterogeneity challenging[15]. Importantly, patients with polyclonal tumors have also been found to be at a higher risk for relapse following treatment[16].

The current "gold standard" for cancer diagnosis and detection of heterogeneity is through tissue biopsy, an invasive procedure that removes a small amount of tissue for further examination. However, this is not always feasible when tumor tissue may be inaccessible. Additionally, a solid tissue biopsy may miss some of the tumor, leading to underestimation of disease grade and stage, and cannot be easily performed repeatedly for disease monitoring. Circulating blood biomarkers, including circulating tumor cells, cell-free DNA, cell-free RNA, and exosomes, are a promising alternative that may provide better insight into the evolution of tumor dynamics throughout the disease.

1.4 Cell-free DNA

Cell-free DNA (cfDNA) consists of short DNA fragments released into circulation as a result of cell lysis, apoptosis, necrosis, and active secretion, and can be found in blood, urine, stool, cerebrospinal fluid, and saliva[17–20]. Despite its discovery in 1948, circulating tumor DNA (ctDNA) in cancer patients was only found later in 1989[18]. In 1994, *KRAS* mutations were detected in the cfDNA of pancreatic cancer using PCR with allele-specific primers, which were also found in the patient's tumor and confirmed the ctDNA fragments were of tumor origin[21]. A decade later, fetal-derived cfDNA was detected in the circulation of pregnant women, and is now regularly used for noninvasive prenatal testing to screen for monogenic disorders, aneuploidies such as Down syndrome (trisomy 21), and sex determination. In recent years, there has been a resurgence of interest in the clinical utility of cfDNA in the context of cancer, with rapid development leading to its adoption by clinicians.

In healthy individuals, cfDNA is predominantly of hematopoietic origin. In patients with cancer, cfDNA includes DNA from hematopoietic cells, but also circulating tumor DNA derived from tumor cells[22]. The fraction of ctDNA in overall cfDNA can vary greatly depending on tumor type and disease burden that lead to differences in rates of cell death in individual tumors, and ranges from 0.1% to 90%[23]. Both quantification of total cfDNA and genomic interrogation of ctDNA fragments have been used to guide diagnosis, treatment selection, prognosis, and disease monitoring, with distinct advantages and disadvantages in cost, specificity, and turnaround time[24]. Elevated preoperative cfDNA concentrations in men with localized prostate cancer who underwent radical prostatectomy (RP) have been associated with PSA recurrence, and detection of *BRCA2*

reversion mutations in cfDNA has been associated with resistance to PARP inhibitors in patients with metastatic castration-resistant prostate cancer (mCRPC)[25,26].

Due to the short half-life and potentially low concentration of ctDNA fragments, specialized approaches are necessary for collection and analysis. Due to the potential for normal blood cell lysis contaminating cfDNA with germline DNA, blood must be drawn with large gauge needles to keep blood cells intact and collected in tubes that must be centrifuged and separated within 1 to 4 hours of collection or collected in tubes with a stabilizing agent, and centrifuged twice with an initial low-speed spin followed by a high-speed spin to remove white blood cells and cellular debris[27]. Ultrasensitive methods are required to detect mutations, copy-number changes, and other alterations present in cfDNA at low variant allele frequencies. Allele-specific techniques such as PCR analysis (i.e. BEAMing and ddPCR) can be used to identify variants present at 0.01% allele frequencies. While next-generation sequencing methods allow for broader detection of alterations in cfDNA, they are limited by the error rate of DNA polymerase and the sequence by synthesis reaction. To address this challenge, molecular barcoding methods that tag cfDNA fragments with unique nucleotide sequences are used when preparing the cfDNA library for sequencing.

The ability to analyze ctDNA from a routine blood draw allows for serial sample collection, which is particularly advantageous when tissue biopsy is difficult or unavailable. Clinical applications in early detection, identification of residual disease, molecular profiling for treatment guidance, treatment response monitoring, and identification of resistance mechanisms are being investigated. In June 2016, the U.S. Food and Drug Administration (FDA) approved the Roche Cobas EGFR Mutation Test v2 for use with plasma cfDNA, the first liquid biopsy companion diagnostic test to

be approved by this agency. This test detects mutations that confer mechanisms of resistance to EGFR-targeted therapies for non-small cell lung cancer (NSCLC) patients, which prompts a change in therapy to mitigate resistance[28]. Additionally, ongoing efforts by companies such as Guardant Health, Grail, Freenome, and Foundation Medicine, are currently underway, with potentially transformative applications in oncology and immuno-oncology.

In this dissertation, I aim to investigate the utility of cfDNA in both localized and advanced prostate cancer. First, I examine the prognostic potential of overall cfDNA concentration and fragment size. Then, I evaluate the ability of cfDNA sequencing—without *a priori* patient-specific tumor mutation information—to identify clonal heterogeneity found in matched tumor tissue from multiple regions in patients with localized prostate cancer. Specifically, cfDNA is tagged with unique barcodes and sequenced with a custom 2.5Mb targeted panel. Finally, I discuss the promises and pitfalls of cfDNA testing, and the importance of continued, rigorous studies to identify potential avenues that can transform the way we diagnose and manage cancer. Distinct challenges lie ahead, but with a sense of urgency that we owe to those who built the canon of knowledge before us, the future of guided cancer treatment remains promising.

1. Lewis, R. *Human Genetics: Concepts and Applictaions*. (2010).

2. Crick, F. & Watson, J. Molecular structure of nucleic acids. (1953).

3. Wetterstrand KA. DNA Sequencing Costs: Data from the NHGRI Genome Sequencing Program (GSP) available at www.genome.gov/sequencingcostsdata. *Natl. Hum. Genome Res. Inst.*

4. Metzker, M. L. Sequencing technologies the next generation. *Nat. Rev. Genet.* **11**, 31–46 (2010).

5. Mardis, E. R. A decade's perspective on DNA sequencing technology. *Nature* **470**, 198–203 (2011).

6. Howlader, N. & Cronin, K. SEER Cancer Statistics Review, 1975-2017. *Natl. Cancer Inst.* (2020).

7. Balmain, A. Cancer as a complex genetic trait: tumor susceptibility in humans and mouse models. *Cell* **108**, 145–152 (2002).

8. Amirouchene-Angelozzi, N., Swanton, C. & Bardelli, A. Tumor evolution as a therapeutic target. *Cancer Discov.* **7**, 805–817 (2017).

9. Boutros, P. C. *et al.* Spatial genomic heterogeneity within localized, multifocal prostate cancer. *Nat. Genet.* **47**, 736–745 (2015).

10. Cooper, C. S. *et al.* Analysis of the genetic phylogeny of multifocal prostate cancer identifies multiple independent clonal expansions in neoplastic and morphologically normal prostate tissue. *Nat. Genet.* **47**, 367–72 (2015).

11. Siegel, R. L., Miller, K. D. & Jemal, A. Cancer statistics, 2020. *Cancer J Clin* **70**, 7–30 (2020).

12. Cooperberg, M. R., Broering, J. M. & Carroll, P. R. Risk assessment for prostate cancer

metastasis and mortality at the time of diagnosis. *J. Natl. Cancer Inst.* **101**, 878–887 (2009).

13.   Locke, J. A. *et al.* NKX3.1 haploinsufficiency is prognostic for prostate cancer relapse following surgery or image-guided radiotherapy. *Clin. Cancer Res.* **18**, 308–316 (2012).

14.   Zafarana, G. *et al.* Copy number alterations of c-MYC and PTEN are prognostic factors for relapse after prostate cancer radiotherapy. *Cancer* **118**, 4053–4062 (2012).

15.   Fraser, M. *et al.* Genomic hallmarks of localized, non-indolent prostate cancer. (2017). doi:10.1038/nature20788

16.   Melijah, S. *et al.* The Evolutionary Landscape of Localized Prostate Cancers Drives Clinical Aggression. *Cell* **173**, (2018).

17.   Mandel, P. & Metais, P. Les acides nucleiques du plasma sanguin ches l'homme. *C R Seances Soc Biol Fil.* **142**, 241–243 (1948).

18.   Stroun, M. *et al.* Neoplastic characteristics of the DNA found in the plasma of cancer patients. *Oncology* **46**, 318–22 (1989).

19.   Jahr, S. *et al.* DNA fragments in the blood plasma of cancer patients: Quantitations and evidence for their origin from apoptotic and necrotic cells. *Cancer Res.* **61**, 1659–1665 (2001).

20.   Stroun, M., Lyautey, J., Lederrey, C., Olson-Sand, A. & Anker, P. About the possible origin and mechanism of circulating DNA: Apoptosis and active DNA release. *Clin. Chim. Acta* **313**, 139–142 (2001).

21.   Sorenson, G. D. *et al.* Soluble Normal and Mutated DNA Sequences from Single-Copy Genes in Human Blood. *Cancer Epidemiol. Biomarkers Prev.* **3**, 67–71 (1994).

22.   Lui, Y. Y. N. *et al.* Predominant hematopoietic origin of cell-free dna in plasma and serum

after sex-mismatched bone marrow transplantation. *Clin. Chem.* **48**, 421–427 (2002).

23. Bettegowda, C. *et al.* Detection of Circulating Tumor DNA in Early- and Late-Stage Human Malignancies. *Sci. Transl. Med.* **6**, 224ra24-224ra24 (2014).

24. Bronkhorst, A. J., Ungerer, V. & Holdenrieder, S. The emerging role of cell-free DNA as a molecular marker for cancer management. *Biomol. Detect. Quantif.* **17**, (2019).

25. Bastian, P. J. *et al.* Prognostic value of preoperative serum cell-free circulating DNA in men with prostate cancer undergoing radical prostatectomy. *Clin. Cancer Res.* **13**, 5361–5367 (2007).

26. Quigley, D. *et al.* Analysis of Circulating Cell-free DNA Identifies Multi-clonal Heterogeneity of BRCA2 Reversion Mutations Associated with Resistance to PARP Inhibitors. *Cancer Discov.* CD-17-0146 (2017). doi:10.1158/2159-8290.CD-17-0146

27. Greytak, S. R. *et al.* Harmonizing Cell-Free DNA Collection and Processing Practices through Evidence-Based Guidance. *Clin. Cancer Res.* **26**, 3104–3109 (2020).

28. Kwapisz, D. The first liquid biopsy test approved. Is it a new era of mutation testing for non-small cell lung cancer? *Ann. Transl. Med.* **5**, 1–7 (2017).

29. Siegel, R. L., Miller, K. D. & Jemal, A. Cancer statistics, 2019. *CA. Cancer J. Clin.* **69**, 7–34 (2019).

30. Lilja, H., Ulmert, D. & Vickers, A. J. Prostate-specific antigen and prostate cancer: Prediction, detection and monitoring. *Nat. Rev. Cancer* **8**, 268–278 (2008).

31. Carroll, P. H. & Mohler, J. L. NCCN guidelines updates: Prostate cancer and prostate cancer early detection. *JNCCN J. Natl. Compr. Cancer Netw.* **16**, 620–623 (2018).

32. Koffler, D., Agnello, V., Winchester, R. & Kunkel, H. G. The occurrence of single-stranded DNA in the serum of patients with systemic lupus erythematosus and other

diseases. *J. Clin. Invest.* **52**, 198–204 (1973).

33.   Tan, E. M., Schur, P. H., Carr, R. I. & Kunkel, H. G. Deoxybonucleic acid (DNA) and

         antibodies to DNA in the serum of patients with systemic lupus erythematosus. *J. Clin.*

         Invest. **45**, 1732–1740 (1966).

34.   Tissot, C. *et al.* Circulating free DNA concentration is an independent prognostic

         biomarker in lung cancer. *Eur Respir J* **46**, 1773–1780 (2015).

35.   Ellinger, J. *et al.* The role of cell-free circulating DNA in the diagnosis and prognosis of

         prostate cancer. *Urol. Oncol. Semin. Orig. Investig.* **29**, 124–129 (2011).

36.   Jung, K. *et al.* Increased cell-free DNA in plasma of patients with metastatic spread in

         prostate cancer. *Cancer Lett.* **205**, 173–80 (2004).

37.   Feng, J. *et al.* Plasma cell-free DNA and its DNA integrity as biomarker to distinguish

         prostate cancer from benign prostatic hyperplasia in patients with increased serum prostate-

         specific antigen. *Int Urol Nephrol* **45**, 1023–1028 (2013).

38.   Lapin, M. *et al.* Fragment size and level of cell-free DNA provide prognostic information

         in patients with advanced pancreatic cancer. *J. Transl. Med.* **16**, 1–10 (2018).

39.   Mouliere, F. *et al. Enhanced detection of circulating tumor DNA by fragment size analysis.*

         Sci. Transl. Med **10**, (2018).

40.   Jiang, P. *et al.* Lengthening and shortening of plasma DNA in hepatocellular carcinoma

         patients. *Proc. Natl. Acad. Sci. U. S. A.* **112**, E1317–E1325 (2015).

41.   Underhill, H. R. *et al.* Fragment Length of Circulating Tumor DNA. 1–24 (2016).

         doi:10.1371/journal.pgen.1006162

42.   Marrone, M., Potosky, A. L., Penson, D. & Freedman, A. N. A 22 gene-expression assay,

         decipher® (GenomeDx biosciences) to predict five-year risk of metastatic prostate cancer

in men treated with radical prostatectomy. *PLoS Curr.* **7**, 1–8 (2015).

43.   Cooperberg, M. R., Hilton, J. F. & Carroll, P. R. The CAPRA-S score: a straightforward

tool for improved prediction of outcomes after radical prostatectomy. *Cancer* **117**, 5039–

5046 (2011).

44.   Cooperberg, M. R. *et al.* Combined value of validated clinical and genomic risk

stratification tools for predicting prostate cancer mortality in a high-risk prostatectomy

cohort. *Eur. Urol.* **67**, 326–333 (2015).

45.   Liu, H. *et al.* Prognostic significance of six clinicopathological features for biochemical

recurrence after radical prostatectomy: A systematic review and meta-analysis. *Oncotarget*

9, 32238–32249 (2018).

46.   Wyatt, A. W. *et al.* Concordance of Circulating Tumor DNA and Matched Metastatic

Tissue Biopsy in Prostate Cancer. doi:10.1093/jnci/djx118

47.   Lapin, M. *et al.* Fragment size and level of cell-free DNA provide prognostic information

in patients with advanced pancreatic cancer. *J. Transl. Med.* **16**, 300 (2018).

48.   Meddeb, R., Pisareva, E. & Thierry, A. R. Guidelines for the preanalytical conditions for

analyzing circulating cell-free DNA. *Clin. Chem.* **65**, 623–633 (2019).

49.   Lee, T.-H., Montalvo, L., Chrebtow, V. & Busch, M. P. Quantitation of genomic DNA in

plasma and serum samples: higher concentrations of genomic DNA found in serum than in

plasma. *Transfusion* **41**, 276–282 (2001).

50.   Jung, M., Klotzek, S., Lewandowski, M., Fleischhacker, M. & Jung, K. Changes in

concentration of DNA in serum and plasma during storage of blood samples [5]. *Clin.*

Chem. **49**, 1028–1029 (2003).

51.   Devonshire, A. S. *et al.* Towards standardisation of cell-free DNA measurement in plasma:

Controls for extraction efficiency, fragment size bias and quantification. *Anal. Bioanal. Chem.* **406**, 6499–6512 (2014).

52. Prakash, K., Aggarwal, S., Bhardwaj, S., Ramakrishna, G. & Pandey, C. K. Serial perioperative cell-free DNA levels in donors and recipients undergoing living donor liver transplantation. *Acta Anaesthesiol. Scand.* **61**, 1084–1094 (2017).

53. Sozzi, G. *et al.* Quantification of free circulating DNA as a diagnostic marker in lung cancer. *J. Clin. Oncol.* **21**, 3902–3908 (2003).

54. Boorjian, S. A. *et al.* Long-Term Outcome After Radical Prostatectomy for Patients With Lymph Node Positive Prostate Cancer in the Prostate Specific Antigen Era. *J. Urol.* **178**, 864–871 (2007).

55. Litwin, M. S. & Tan, H. J. The diagnosis and treatment of prostate cancer: A review. *JAMA - J. Am. Med. Assoc.* **317**, 2532–2542 (2017).

56. Caswell, D. R. & Swanton, C. The role of tumour heterogeneity and clonal cooperativity in metastasis, immune evasion and clinical outcome. *BMC Med.* **15**, 133 (2017).

57. Cooper, C. S. *et al.* Analysis of the genetic phylogeny of multifocal prostate cancer identifies multiple independent clonal expansions in neoplastic and morphologically normal prostate tissue. *Nat. Genet.* **47**, 367–372 (2015).

58. Cancer Genome Atlas Research Network, T. *et al.* The Molecular Taxonomy of Primary Prostate Cancer. *Cancer Genome Atlas Res. Netw. Cell* **163**, 1011–1025 (2015).

59. Lalonde, E. *et al.* Tumour genomic and microenvironmental heterogeneity for integrated prediction of 5-year biochemical recurrence of prostate cancer: a retrospective cohort study. (2014). doi:10.1016/S1470-2045(14)71021-6

60. Price, T. J. *et al.* Prognostic impact and the relevance of PTEN copy number alterations in

patients with advanced colorectal cancer (CRC) receiving bevacizumab. *Cancer Med.* **2**, 277–285 (2013).

61. Boutros, P. C. *et al.* Spatial genomic heterogeneity within localized, multifocal prostate cancer. *Nat. Genet.* **47**, 1–14 (2015).

62. Abbosh, C. *et al.* Phylogenetic ctDNA analysis depicts early stage lung cancer evolution. *Nature* **22364**, 1–25 (2017).

63. Phallen, J. *et al.* Direct detection of early-stage cancers using circulating tumor DNA. **2415**, (2017).

64. Hennigan, S. T., Trostel, S. Y., Terrigino, N. T., Voznesensky, O. S. & Schaefer, R. J. Low Abundance of Circulating Tumor DNA in Localized Prostate Cancer. (2019). doi:10.1200/PO.19

65. Kinde, I., Wu, J., Papadopoulos, N., Kinzler, K. W. & Vogelstein, B. Detection and quantification of rare mutations with massively parallel sequencing. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 9530–9535 (2011).

66. Cario, C. L. *et al.* A machine learning approach to optimizing cell-free DNA sequencing panels: with an application to prostate cancer. *BMC Cancer* 1–9 (2020). doi:10.1101/2020.04.30.069658

67. Pérez-Barrios, C. *et al.* Comparison of methods for circulating cell-free DNA isolation using blood from cancer patients: Impact on biomarker testing. *Transl. Lung Cancer Res.* **5**, 665–672 (2016).

68. Benjamin, D. *et al.* Calling Somatic SNVs and Indels with Mutect2. *bioRxiv* 1–8 (2019). doi:10.1101/861054

69. Ramos, A. H. *et al.* Oncotator: Cancer variant annotation tool. *Hum. Mutat.* **36**, E2423–

E2429 (2015).

70. Adalsteinsson, V. A. *et al.* Abstract LB-136: High concordance of whole-exome sequencing of cell-free DNA and matched biopsies enables genomic discovery in metastatic cancer. *Cancer Res.* **76**, (2016).

71. Lang, S. H. *et al.* A systematic review of the prevalence of DNA damage response gene mutations in prostate cancer. *Int. J. Oncol.* **55**, 597–616 (2019).

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

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

74. Ashouri, A., Sayin, V., Van den Eynden, J., Singh, S. & Larsson, E. Pan-cancer transcriptomic analysis associates long non- coding RNAs with key mutational driver events. *Nat. Commun.* **7**, (2016).

75. Lanzós, A., Carlevaro-Fita, J., Mularoni, L. & Johnson, R. Discovery of Cancer Driver Long Noncoding RNAs across 1112 Tumour Genomes: New Candidates and Distinguishing Features. *Nat. Sci. Reports* **7**, (2017).

76. Fredriksson, N. J., Ny, L., Nilsson, J. A. & Larsson, E. Systematic analysis of noncoding somatic mutations and gene expression alterations across 14 tumor types. *Nat. Genet.* **46**, 1258–1263 (2014).

77. He, F. *et al.* Integrative Analysis of Somatic Mutations in Non-coding Regions Altering RNA Secondary Structures in Cancer Genomes. *Sci. Rep.* **9**, 1–12 (2019).

78. Abbosh, C., Swanton, C. & Birkbak, N. J. Clonal haematopoiesis: A source of biological noise in cell-free DNA analyses. *Annals of Oncology* **30**, 358–359 (2019).

79. Liu, J. *et al.* Biological background of the genomic variations of cf-DNA in healthy individuals. *Ann. Oncol.* **30**, 464–470 (2019).

80. Melijah, S. *et al.* The Evolutionary Landscape of Localized Prostate Cancers Drives Clinical Aggression. *Cell* **173**, (2018).

Chapter II

Cell-free DNA as a Biomarker for Prostate Cancer: Elevated Concentration and Decreased Fragment Size

2.1 Abstract

Purpose: Prostate cancer is the most commonly diagnosed neoplasm in American men. Although existing biomarkers may detect localized prostate cancer, additional strategies are necessary for improving detection and identifying aggressive disease that may require further intervention. One promising, minimally invasive biomarker is cell-free DNA (cfDNA), which consist of short DNA fragments released into circulation by dying or lysed cells that may reflect underlying cancer. Here we investigated whether differences in cfDNA concentration and cfDNA fragment size could improve the sensitivity for detecting more advanced and aggressive prostate cancer.

Materials and Methods: This study included 268 individuals: 34 healthy controls, 112 men with localized prostate cancer who underwent radical prostatectomy (RP), and 122 men with metastatic castration-resistant prostate cancer (mCRPC). Plasma cfDNA concentration and fragment size were quantified with the Qubit 3.0 and the 2100 Bioanalyzer. The potential relationship between cfDNA concentration or fragment size and localized or mCRPC prostate cancer was evaluated with descriptive statistics, logistic regression, and area under the curve analysis with cross-validation.

Results: Plasma cfDNA concentrations were elevated in mCRPC patients in comparison to localized disease ($OR_{5 \, ng/mL}$ = 1.34, P = 0.027) or to being a control ($OR_{5 \, ng/mL}$ = 1.69, P = 0.034).

Decreased average fragment size was associated with an increased risk of localized disease compared to controls ($OR_{5bp} = 0.77$, $P = 0.0008$).

Conclusion: This study suggests that cfDNA concentration and average cfDNA fragment size may provide a quick, cost-effective approach to help determine which patients will benefit most from further screening and/or disease monitoring to help improve prostate cancer outcomes.

2.2 Introduction

Prostate cancer accounts for approximately 20% of all new cancer diagnoses in American men. While individuals diagnosed with localized disease have a 98% 5-year survival rate, an estimated 33,330 men will die from aggressive and metastatic disease in 2020[29]. There are a number of existing biomarkers routinely used for prostate cancer diagnosis and monitoring, including prostate-specific antigen (PSA), PHI, 4Kscore, PCA3 expression, parametric MRI, and hypermethylation of GSTP1, APC, and RASSF1[30,31]. These have varying levels of sensitivity and specificity, and additional biomarkers for prostate cancer are necessary to reduce over-diagnosis and over-treatment of this common, but complex disease.

Cell-free DNA (cfDNA) is a promising, minimally invasive biomarker that may originate from cell lysis, apoptosis, necrosis, and active release of DNA fragments into circulation[17–20]. In healthy individuals, cfDNA is predominantly of hematopoietic origin[22]. In cancer patients, cfDNA includes DNA of hematopoietic origin, as well as circulating tumor DNA (ctDNA) derived from tumor cells. Two commonly used methods to profile cfDNA are: 1) quantification of cfDNA based on spectrophotometry, electrophoresis, or quantitative PCR (qPCR); and 2) genomic interrogation of ctDNA fragments with next-generation sequencing, BEAMing (beads, emulsion, amplification, and magnetics), or droplet digital PCR (ddPCR).

While genomic interrogation allows for the detection of cancer-specific fragments, this can have a number of challenges (e.g., sufficient cfDNA, sequencing depth, and mutation panel selection). In contrast, quantification of overall cfDNA concentrations and assessment of cfDNA fragment size may provide a quick, cost-effective method in addition to other biomarkers such as PSA, and

can deliver insight into whether a patient should undergo further biopsy and potentially genomic testing.

Elevated concentrations of cfDNA were initially reported in patients with leukemia and autoimmune disease[32,33]. Subsequent studies have also determined that high concentrations of cfDNA are typically associated with poor survival in several cancers[34,35]. For prostate cancer, increased plasma cfDNA concentrations were found in patients with lymph node and distant metastases[36]. Elevated preoperative serum cfDNA concentrations in men with localized prostate cancer who underwent RP have been associated with PSA recurrence, independent of surgical margin and lymph node status, as well as Gleason score and pathologic stage[25].

In addition to overall cfDNA concentrations, cfDNA fragment size may provide diagnostic and prognostic value. DNA integrity, which measures the ratio of all cfDNA fragments (ALU 247bp) to shorter fragments (ALU 115bp) has distinguished prostate cancer from benign prostatic hyperplasia (BPH)[37]. Pre-treatment cfDNA concentration and fragment size were predictive of advanced pancreatic cancer progression-free survival and overall survival[38]. Furthermore, tumor fragments in cfDNA appeared shorter in size than fragments that originated from non-malignant cells[39–41].

Here, we evaluate whether baseline plasma cfDNA concentrations and cfDNA fragment size can differentiate among: 1) men with prostate cancer and controls; and 2) clinical characteristics or biochemical recurrence among men with localized disease (i.e., PSA at diagnosis, Gleason, organ confinement, extraprostatic extension, seminal vesicle invasion, lymph node invasion, and RNA

gene expression). While cfDNA concentration data was available for mCRPC and localized prostate cancer groups, cfDNA fragment size data was only available for patients with localized disease.

2.3 Materials and Methods

2.3.1 Patient Cohort

From August 2015 to November 2019, biological samples from a total of 268 individuals were included in this study: 34 healthy donors, 112 patients with localized prostate cancer, and 124 mCRPC patients (Table 1). Twenty-eight healthy donor samples were obtained from StemCell (StemCell Technologies, Seattle, WA), and six healthy samples were collected from volunteers at UCSF. All patients with localized disease underwent radical prostatectomy (RP) at UCSF, and 35/112 patients had a Decipher score of RNA gene expression available (GenomeDx Biosciences, Vancouver, British Columbia, Canada)[42]. For the patients with localized disease, blood samples were collected the day of surgery before RP, and for five of these men blood samples were collected two months after surgery. Clinicopathologic variables that play an important role in surgical management after prostatectomy were also collected, including clinical T stage, pathologic Gleason score, preoperative PSA, and risk prediction models including the Cancer of the Prostate Risk Assessment Postsurgical (CAPRA-S) score and the Decipher score[43,44]. Known predictors of biochemical recurrence (BCR), including organ confinement, extraprostatic extension, seminal vesicle invasion, and lymph node involvement were also identified[45]. Biochemical recurrence was defined as two consecutive PSA levels of ≥0.2 ng/mL eight weeks after surgery.

While most of the cohort was collected prospectively, a subset of 110 mCRPC patients were recruited through the Stand Up 2 Cancer/Prostate Cancer Foundation-funded West Coast Prostate Cancer Dream Team Project (IRB 12-10340). Fourteen mCRPC patients were recruited through UCSF[46]. For mCRPC patients, blood samples were collected prior to treatment initiation. Clinicopathologic characteristics were collected for all patients (Table 1). Approval for this study was granted by the local ethics review board (IRB 11-05226 and IRB 12-09659), and written informed consent was obtained from all patients.

2.3.2 cfDNA Extraction from Blood

For healthy controls, whole peripheral blood samples were collected from individuals in PAXgene Blood ccfDNA tubes (Qiagen, Redwood City, CA). Healthy samples collected by StemCell were shipped at room temperature, arriving within 7 days for sample processing. Whole peripheral blood samples were collected immediately before surgery for patients with localized disease or at the time of follow-up and before treatment initiation for mCRPC patients. Plasma was generated from whole blood samples within 2 hrs for blood collected in K3EDTA tubes or within 7 days for blood collected in PAXgene Blood ccfDNA tubes with a two-step centrifugation protocol: first centrifuging the blood at 1,900g for 10 min at 21°C, followed by centrifugation of the supernatant at 16,000g for 10 min to remove leukocytes and cellular debris. DNA was extracted from 7 to 55 mL of plasma using the Qiagen QIAamp Circulating Nucleic Acid Kit (Qiagen, Redwood City, CA), and double eluted with 40 μL of Qiagen Elution Buffer. Extracted DNA was stored at -20°C prior to further analysis.

2.3.3 cfDNA Fragment Size and Concentration

Extracted DNA was quantified with a Qubit 3.0 Fluorometer and a DNA dsDNA HS Assay Kit (Life Technologies, Carlsbad, CA), as well as on the 2100 Bioanalyzer with High Sensitivity DNA Chips (Agilent Technologies, Santa Clara, CA) for assessment of sample purity, concentration, and fragment size distribution according to the manufacturer's instructions. The average fragment size was determined with the Agilent 2100 Bioanalyzer Expert software, and calculated across the first three peaks 75–675bp corresponding to the length of nucleosomal footprints and linkers derived from apoptotic cells (Supplementary Figure 1). The final plasma cfDNA concentrations were calculated by adjusting for the initial plasma and final elution volumes, and quantified with a Qubit 3.0 for a subset of patients (Supplementary Table 1). Assessment of cfDNA fragment size and concentration was performed without prior knowledge of clinical data. Average cfDNA fragment size was not available for mCRPC patients, since samples were not run on the 2100 Bioanalyzer.

2.3.4 Statistical Analysis

Our primary analysis assessed the relationship between cfDNA concentration or average fragment size and prostate cancer, comparing three groups: healthy controls, men with localized disease, and men with mCRPC. Here we used descriptive statistics, logistic regression, and receiver operating characteristic (ROC) curves. Since cfDNA concentration and average fragment size were not normally distributed ($P < 0.001$, Shapiro–Wilk test), we evaluated the difference in descriptive statistics across prostate cancer diagnoses using the Mann-Whitney non-parametric test. We also evaluated differences in cfDNA concentration quantified between 90–150bp, which is known to be enriched for circulating tumor DNA fragments specifically[39]. Then, we further

investigated the potential relationship between cfDNA concentration and prostate cancer diagnoses using logistic regression models (crude, and then adjusting for age at time of blood draw and baseline PSA when available). The ability of cfDNA concentration to discriminate between prostate cancer diagnoses was further assessed based on the area under the curve (AUC) from a Receiver Operating Characeristics (ROC) curve analysis with k-fold cross-validation (a minimum of ten observations per fold) and bootstrap resampling (n = 100). These analyses were also performed with average cfDNA fragment size to distinguish patients with localized disease from controls.

We also undertook secondary analyses investigating the relationship between baseline cfDNA concentration or fragment size and clinical characteristics among patients with localized disease. For continuous characteristics, comparisons were made using cfDNA concentration and Pearson correlation coefficients (i.e., age at diagnosis, PSA at diagnosis, Decipher score, time to salvage therapy, average cfDNA fragment size, and postoperative CAPRA-S score). For categorical clinical features, we assessed the potential relationship between log-transformed cfDNA concentration (for normality) and other clinical features (pathologic Gleason score, organ confinement, extraprostatic extension, seminal vesicle invasion, pathologic lymph node status, biochemical recurrence, and clinical T stage) with Student's t–tests. Mann-Whitney U-tests were used to assess the association between the average cfDNA fragment size and the same clinical features. As with localized disease, we also evaluated the relationship between cfDNA concentration and age at time of blood draw for healthy individuals and mCRPC patients. Finally, we evaluated the association between cfDNA concentration or fragment size and biochemical

recurrence-free survival with Cox proportional hazards models for patients with localized disease. All data analyses were performed using R version 3.6.1.

2.4 Results

2.4.1 cfDNA Concentration and Prostate Cancer

The median cfDNA concentration was 7.9 ng/mL (IQR, 4.0 ng/mL) for controls, 6.7 ng/mL (IQR, 5.8 ng/mL) for patients with localized disease, and 13.8 ng/mL (IQR, 28.1 ng/mL) for patients with mCRPC (Table 1; Figure 1). The average cfDNA levels in mCRPC patients were statistically significantly higher than those observed in controls ($P < 0.0001$) or those with localized prostate cancer ($P < 0.0001$).

These observations were further supported by results from the logistic regression models, including those adjusting for age and PSA levels (Table 2). A 5 ng/mL increase in cfDNA concentration was positively associated with mCRPC in comparison to localized disease (OR_{crude} = 1.47, P = 0.0017; $OR_{adjusted}$ = 1.34, P = 0.027) or to being healthy (OR_{crude} = 1.93, P = 0.0025; $OR_{adjusted}$ = 1.69, P = 0.034). Plasma cfDNA concentration was not associated with having localized disease in comparison to healthy individuals (OR_{crude} = 1.10, P = 0.64; $OR_{adjusted}$ = 1.05, P = 0.72).

In our ROC curve analysis, plasma cfDNA concentration was able to distinguish between mCRPC patients from healthy individuals and those with localized disease (Figure 2), with an estimated AUC of 0.83 (95% CI, 0.72–0.91) and 0.81 (95% CI, 0.74–0.87), respectively.

2.4.2 cfDNA Fragment Size and Prostate Cancer

The median of the average cfDNA fragment size in patients with localized disease was 173bp

(range, 135-280bp), and in controls was 177.5bp (range, 142-265bp) (Table 1). This lower average

cfDNA fragment size in patients with localized disease was statistically significantly different

from that observed in controls (P = 0.0009, Figure 3). Results from the logistic regression analysis

further indicate that average fragment size was inversely associated with localized prostate cancer

(in comparison to healthy individuals): for a 5bp increase in fragment size, the OR_{crude} = 0.86 (P =

0.003) and $OR_{adjusted}$ = 0.77 (P = 0.0008; Table 2). The estimated ROC AUC for distinguishing

between healthy individuals and patients with localized prostate cancer using average cfDNA

fragment size was 0.64 as defined by k-fold cross-validation. There was no difference in cfDNA

concentration quantified across 90–150bp between healthy individuals and patients with localized

disease (Supplementary Figure 4)[39].

2.4.3 cfDNA Concentration / Fragment Size and Clinical Characteristics in Localized Prostate

Cancer

There were no statistically significant differences in cfDNA concentration or fragment size for the

clinical characteristics / outcomes evaluated here (Supplementary Table 2; Supplementary Table

3). Specifically, cfDNA concentration or fragment size did not appear to substantively differ

across: pathologic Gleason score, organ confinement, extraprostatic extension, seminal vesicle

invasion, time to biochemical recurrence, average cfDNA fragment size, clinical T stage,

pathologic lymph node status, age at diagnosis, PSA at diagnosis, Decipher score, time to salvage

therapy, and CAPRA-S score (Supplementary Figure 2; Supplementary Table 2; Supplementary

Table 3). Additionally, no clear correlation was observed between cfDNA concentration and age

at time of blood draw for healthy individuals, patients with localized disease, or patients with mCRPC (Supplementary Figure 3).

2.5 Discussion

This study found that plasma cfDNA concentrations and fragment size may have diagnostic and prognostic value for detecting and profiling prostate cancer. Specifically, plasma cfDNA concentrations may help identify patients with advanced disease, while cfDNA fragment size may be used to distinguish patients with early stage disease from healthy individuals.

In the multivariate model that adjusted for age and PSA, plasma cfDNA concentration remained an independent predictor of mCRPC, indicating that cfDNA concentration may capture different biological processes than PSA and provides additional information (Table 2). Average cfDNA fragment size was predictive of localized disease. Looking at follow-up fragment size measures available for five patients, we found that one patient had a shorter average fragment size two months after surgery, and also exhibited post-treatment elevated PSA (Table 2; Supplementary Figure 6). In combination, these findings suggest that quantification of cfDNA overall may be valuable in identifying prostate cancer patients.

While the biological mechanisms underlying decreased fragment size in cancer patients are not well-understood, differences in nucleosome positioning and DNA methylation may result in varied DNA degradation. Our finding that the overall average fragment size in localized patients was shorter and more fragmented than in healthy individuals is consistent with the findings of studies assessing fragment size in patients with hepatocellular carcinoma and advanced pancreatic

cancer[40,47]. The proportion of cfDNA fragments shorter than 150bp is also increased for multiple cancer types when compared to healthy fragment sizes with shallow genome-wide sequencing[39]. Localized prostate cancers are characterized by initial accumulation of clonal point mutations and deletions, with subsequent branching copy number gains where amplified regions are relatively enriched for tumor DNA, possibly modifying intracellular DNA degradation processes and mechanisms of DNA release and contributing to the size differences that were observed[16].

The relatively short follow-up time for a protracted disease like prostate cancer is a limitation in this study. Of the 112 patients who underwent surgery, 24 patients experienced biochemical recurrence with a median follow-up time of three years (range, 9–1704 days), and it was not feasible to identify patients with localized disease who may have progressed to metastatic disease. This study did not include patients with metastatic disease who were not resistant to hormone therapy (i.e. castration sensitive), limiting the generalizability of these results to the full spectrum of this disease.

We processed samples in a manner that maximized the quality and quantity of extracted cfDNA[27,48–51]. An initial low-speed centrifugation step followed by high-speed centrifugation was used to reduce the amount of cellular debris and genomic DNA in the sample. Importantly, there was no significant difference in cfDNA concentration for samples collected in K3EDTA and PAXgene Blood ccfDNA tubes in the localized cohort (Supplementary Figure 5). However, the slightly increased cfDNA concentrations observed in controls may be due to cell lysis during transit, since whole blood was collected from individuals at a donor center in Kent, Washington and shipped overnight to San Francisco, California, whereas the patient samples were collected

and processed onsite. Additionally, the lower overall cfDNA concentrations found in the localized cohort may be due to the subduing effect of anesthetic agents on cell death, which were administered prior to blood sample collection before surgery[52]. While data comparison across studies is difficult due to differences in sample collection and processing, most studies demonstrate the diagnostic role of cfDNA[35]. To quantify cfDNA, the Qubit 3.0 and the Agilent 2100 Bioanalyzer were used as a straightforward approach, albeit potentially less accurate than sequencing. In a clinical setting, an affordable, rapid, and straightforward test is critical to minimizing disruption to standard workflows while providing additional information. However, cfDNA quantification is a complementary approach that could help identify patients who may benefit from cfDNA sequencing.

Bastian et al. observed significant associations between cfDNA concentration and clinical characteristics in a cohort of patients with localized disease that experienced biochemical recurrence, supporting the hypothesis that cfDNA quantification may have more utility in the management of more advanced disease[25]. The lack of associations observed between cfDNA concentration and clinical characteristics in our study may be due to differences in the study cohorts. While all patients in the Bastian et al. study experienced BCR, only 24 of 112 patients experienced BCR in our study[25].

A previous study demonstrated that in pre-treatment specimens, shorter cfDNA fragment size and elevated cfDNA concentrations were associated with shorter progression-free survival and overall survival in patients with advanced pancreatic cancer[47]. Due to the relatively short follow-up time in this study, future longitudinal studies evaluating disease progression from localized to metastatic

disease are necessary to elucidate the value of analyzing cfDNA concentration and fragment size in the context of prostate cancer.

While the exact mechanism of cfDNA release into circulation remains unknown, apoptosis, lysis, necrosis, and active secretion have been identified as potential routes[20,53]. The cfDNA found in healthy individuals originates from hematopoietic cells, and likely reflects the processes of regulated cell turnover in these cells[22]. In patients with cancer, cfDNA includes both DNA fragments from hematopoietic cells, as well as fragments from tumor cells. Future studies evaluating the mechanisms of release will help elucidate the underlying biology of this biomarker, especially in combination with diagnostic and prognostic information over longer periods of time.

2.6 Conclusion

Collectively, our data demonstrate the potential applications of plasma cfDNA concentration and cfDNA fragment size in both advanced and localized prostate cancer. Patients with advanced mCRPC had higher cfDNA concentration than men with localized disease or healthy controls, and those with localized disease had shorter average fragment sizes than controls. Importantly, cfDNA concentration and fragment size remained independent predictors after adjusting for age and PSA. Future studies assessing both cfDNA concentration and fragment size will be necessary to clarify the utility of plasma cfDNA in the context of diagnosis, prognosis, and disease monitoring.

2.7 Acknowledgments

The authors would like to thank all participating patients and their families.

2.8 Ethics Statement

This study was carried out in accordance to the local ethics review board (IRB 11-05226 and IRB 12-09659). All subjects gave written informed consent in accordance with the Declaration of Helsinki.

2.9 Tables

Table 2.1 Clinical characteristics of individuals included in the study at baseline.

	Healthy N = 34	Localized N = 112	mCRPC N = 122
Age (years)			
Median $\pm$ IQR	60 $\pm$ 17	65 $\pm$ 10	71 $\pm$ 9
Range	41 – 74	43 – 78	47 – 91
Pathologic Gleason*			
6	-	6	-
7	-	75	-
8-10	-	30	-
Pathologic Stage			
Organ confined (pT2)	-	46	-
Not organ confined (pT3, pT4)	-	66	-
Extraprostatic extension (pT3a)	-	43	-
Seminal vesicle invasion (pT3b)	-	15	-
Lymph node involvement (N1)	-	17	-
PSA (ng/mL)			
Median $\pm$ IQR	-	6.9 $\pm$ 7.4	50.8 $\pm$ 128
Range	-	1.21 – 70.0	0 – 3007
cfDNA Concentration (ng/mL)[†]			
Median $\pm$ IQR	7.9 $\pm$ 4.0	6.7 $\pm$ 5.8	13.8 $\pm$ 28.1
Range	0.29 – 16.9	1.22 – 53.9	1 – 1380
cfDNA Fragment Size (bp)			
Median $\pm$ IQR	177.5 $\pm$ 29.5	173 $\pm$ 6	-
Range	142 – 265	135 – 280	-

* One man with unknown data in the cohort.
[†] Concentration data collected on a subset of individuals (31 healthy, 45 localized, and 122 mCRPC).

Table 2.2 Association between increase in cfDNA concentration (5 ng/mL) or in cfDNA fragment size (5bp) and prostate cancer status. Results from crude univariate and adjusted multivariate logistic regression analyses (adjusted for age, and PSA when available).

cfDNA Measure	Prostate Cancer Status	Crude Odds Ratio (95% CI)	P-value	Adjusted Odds Ratio (95% CI)	P-value
cfDNA Concentration (5 ng/mL)					
	Localized vs Healthy	1.10 (0.82 – 1.61)	0.64	1.05 (0.77 – 1.69)*	0.72
	mCRPC vs Healthy	1.93 (1.34 – 3.18)	0.0025	1.69 (1.16 – 2.93)*	0.034
	mCRPC vs Localized	1.47 (1.22 – 2.01)	0.0017	1.34 (1.05 – 1.76)†	0.027
cfDNA Fragment Size (5bp)					
	Localized vs Healthy	0.86 (0.73 – 0.9)	0.003	0.77 (0.66 – 0.90)*	0.0008

* Adjusted for age
† Adjusted for age and PSA

33

2.10 Figures
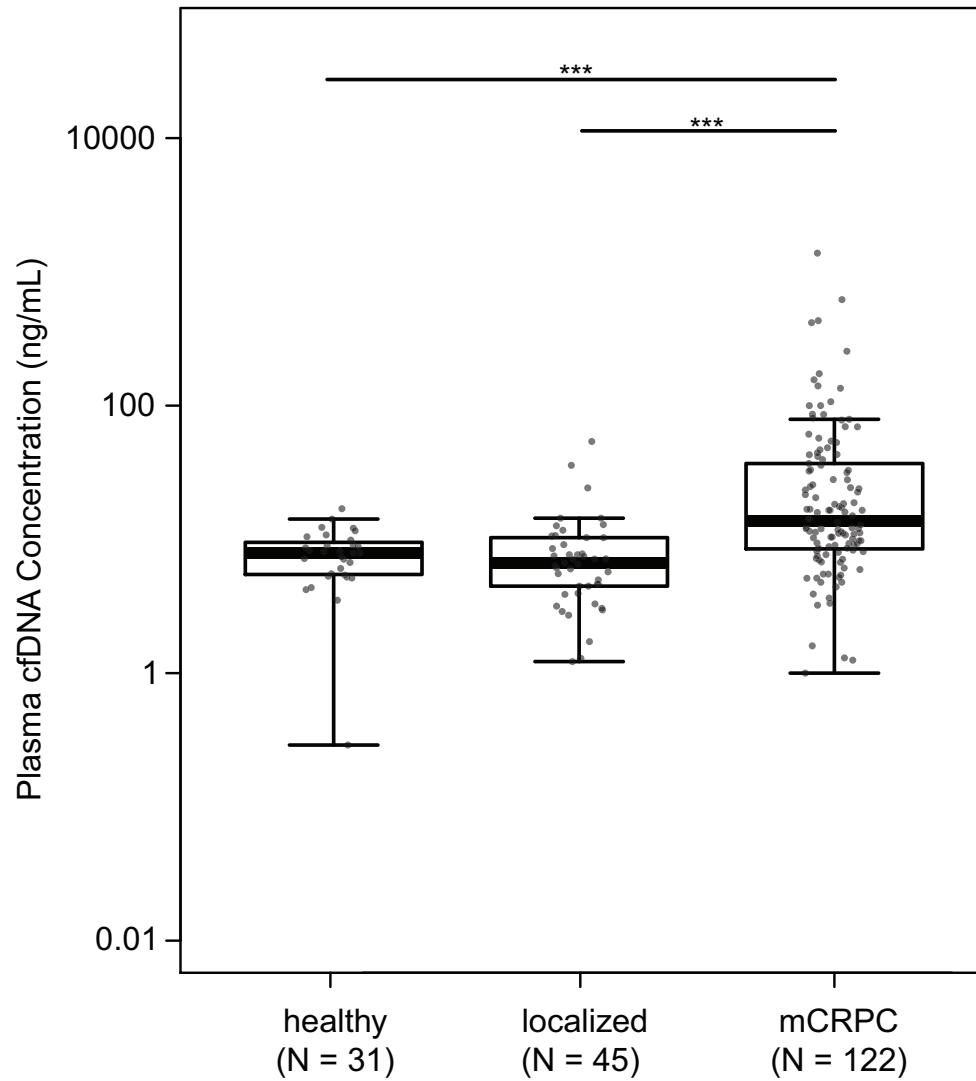
Figure 2.1 Distribution of plasma cfDNA concentration in healthy individuals, patients with localized disease, and patients with mCRPC. Boxplots and points identify the minimum, interquartile range, median, and maximum values for each group. The Mann-Whitney test was applied to test differences in cfDNA concentration between groups.

*** P < 0.0001
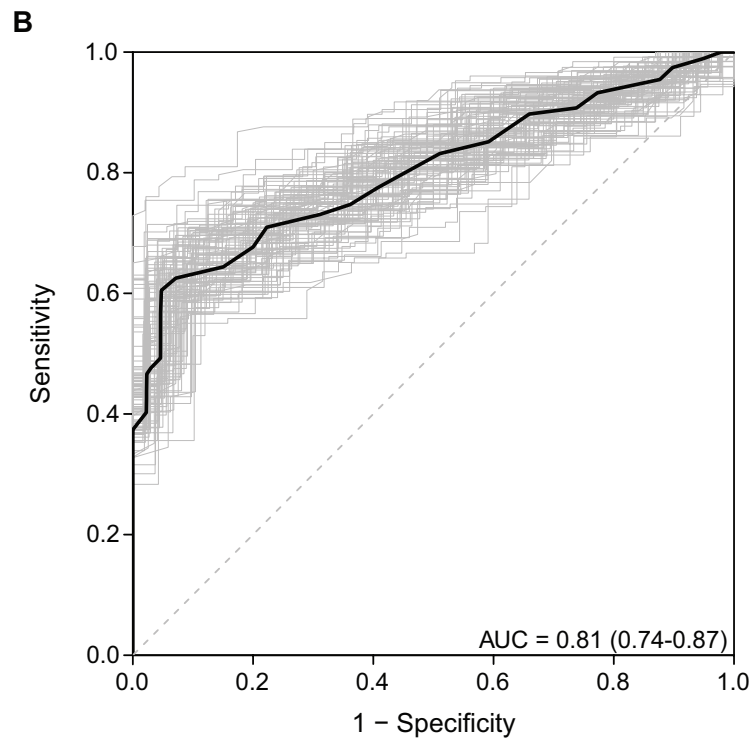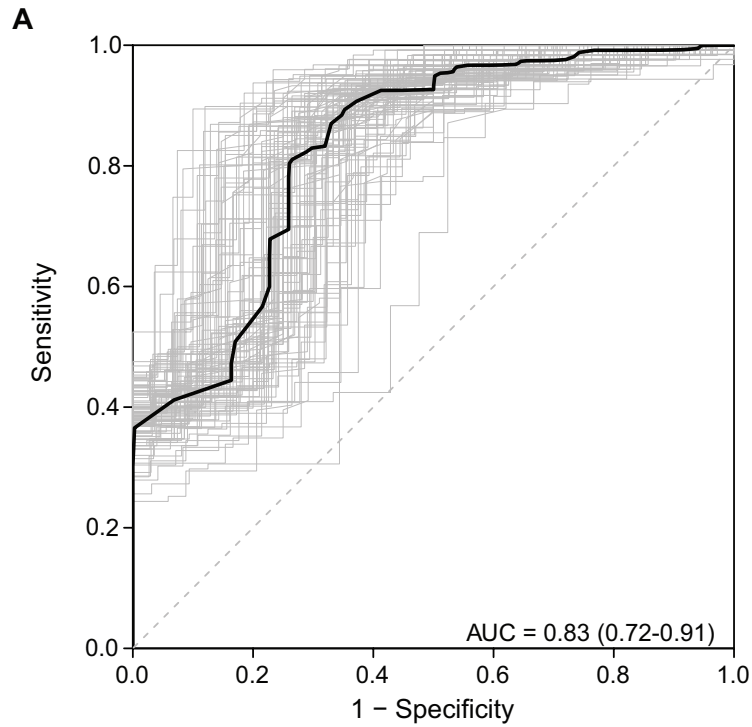
Figure 2.2 Receiver operating characteristic (ROC) curves for cfDNA concentration comparison between A) healthy individuals and mCRPC, and B) patients with localized disease and mCRPC. Area under the curve (AUC) and 95% CI were estimated with k-fold cross-validation and bootstrap resampling.
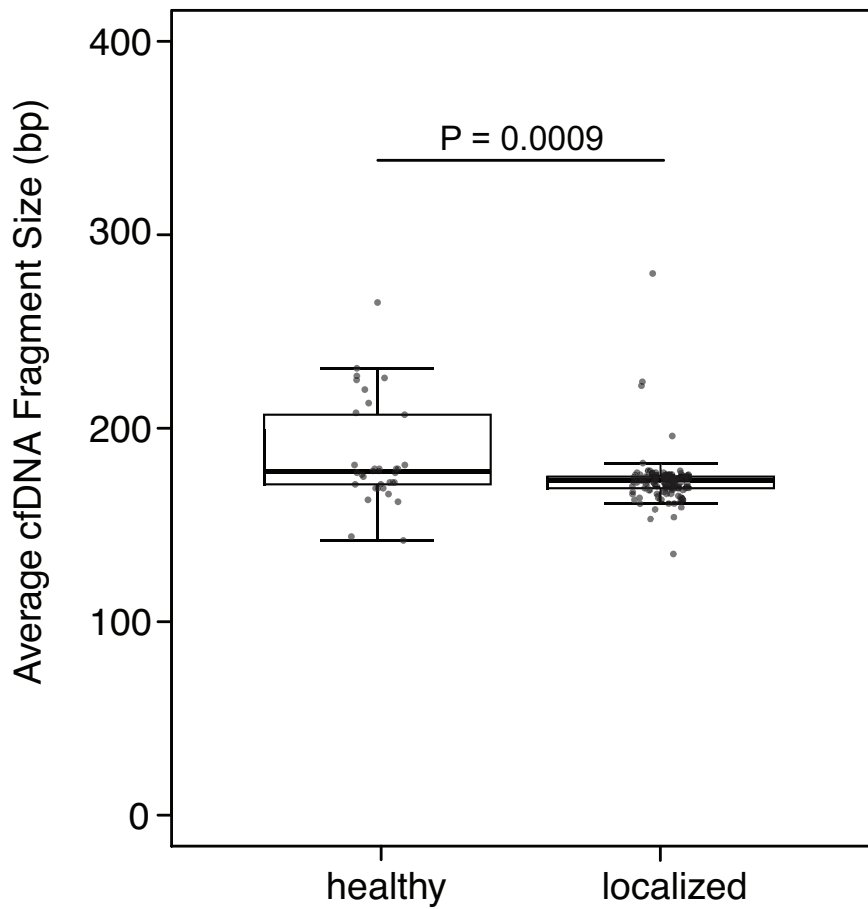
Figure 2.3 Distribution of average fragment size in healthy individuals and patients with localized disease (P = 0.0009). Boxplots and points identify the minimum, interquartile range, median, and maximum values for each group. A Mann-Whitney U-test was performed to test the difference in cfDNA fragment size.

2.11 Supplementary Materials
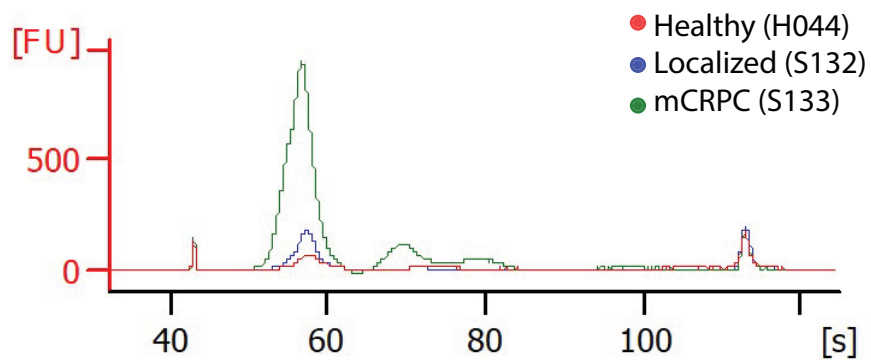
Figure S2.1 Plasma cfDNA concentration and distribution of cfDNA fragment size with representative traces for a healthy individual, patient with localized prostate cancer, and mCRPC patient.
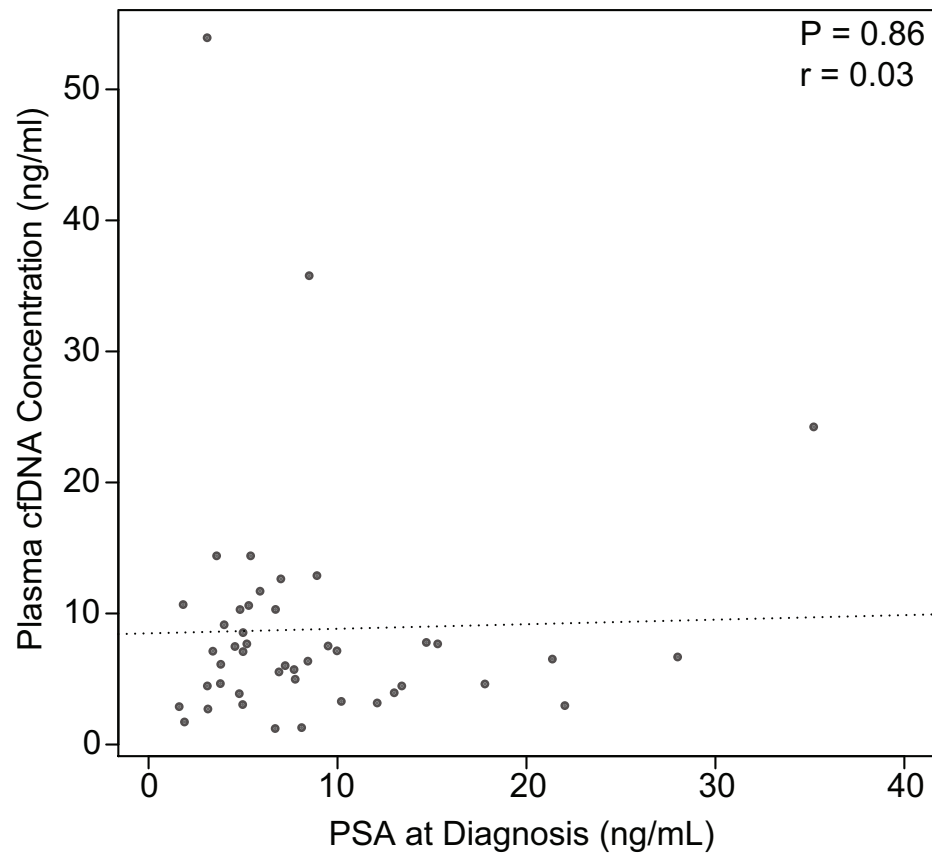
Figure S2.2 Assessment of the relationship between cfDNA concentration and PSA for patients with localized disease (P = 0.86).
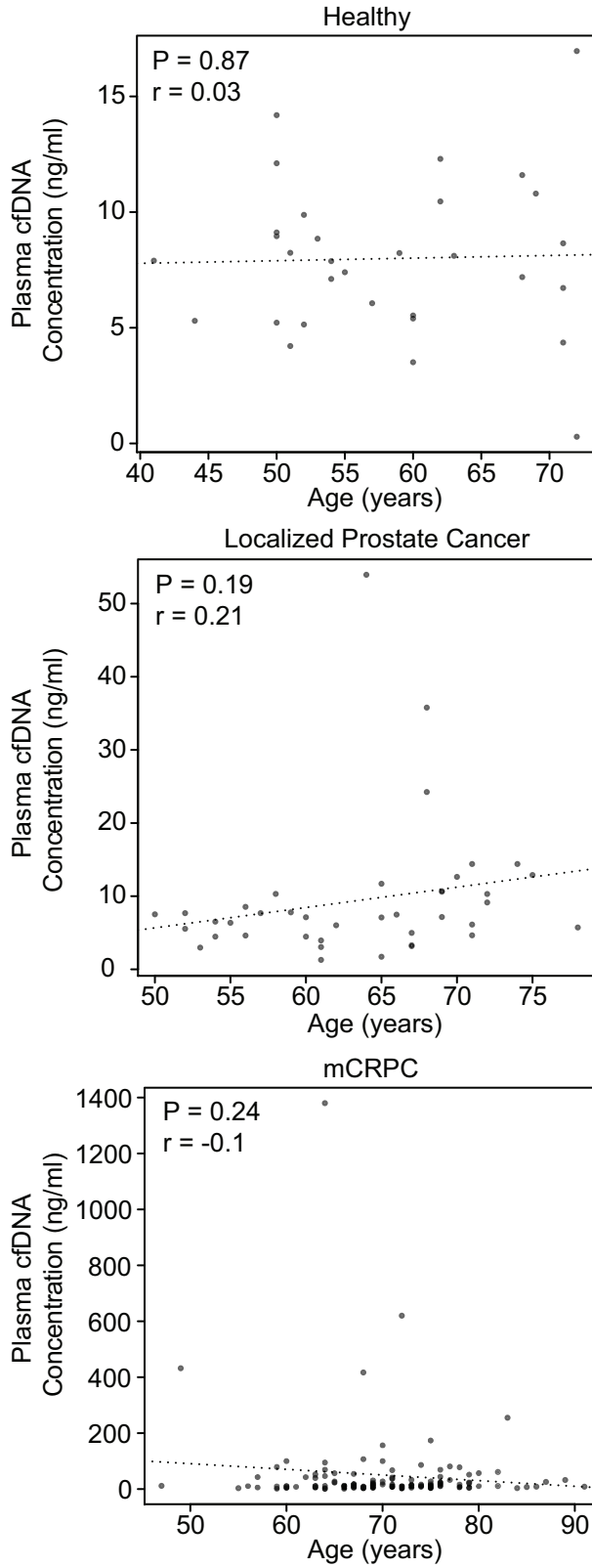
Figure S2.3 Assessment of the relationship between cfDNA concentration and age at time of blood draw.
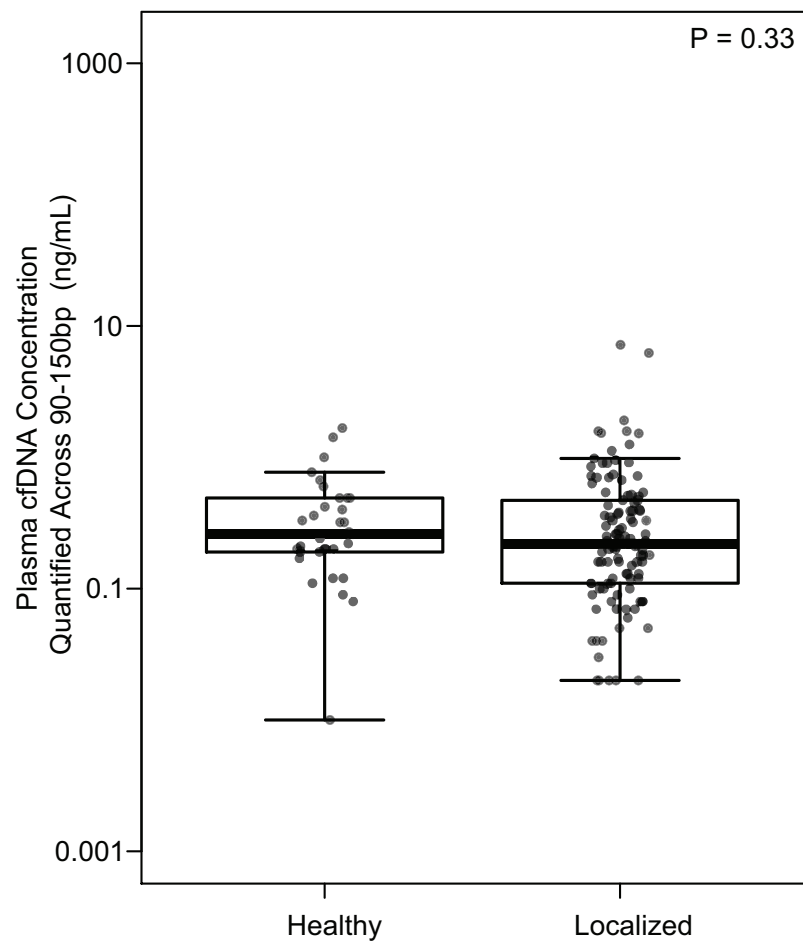
Figure S2.4 Distribution of plasma cfDNA concentration quantified across 90–150bp in healthy individuals and patients with localized disease with the 2100 Bioanalyzer. Boxplots and points identify the minimum, interquartile range, median, and maximum values for each group. A Mann-Whitney U-test was performed to test the difference in cfDNA yields.
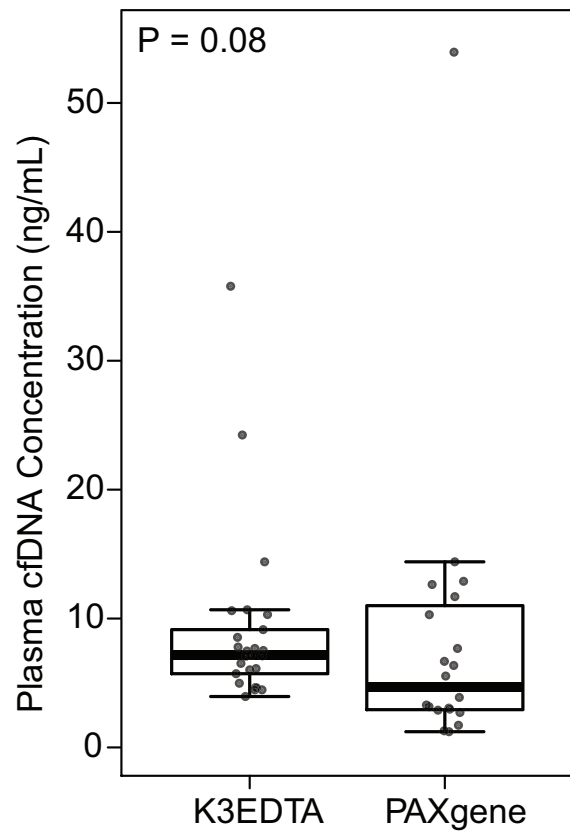
Figure S2.5 Distribution of cfDNA concentration for K3EDTA and Qiagen PAXgene tube types for patients with localized disease.
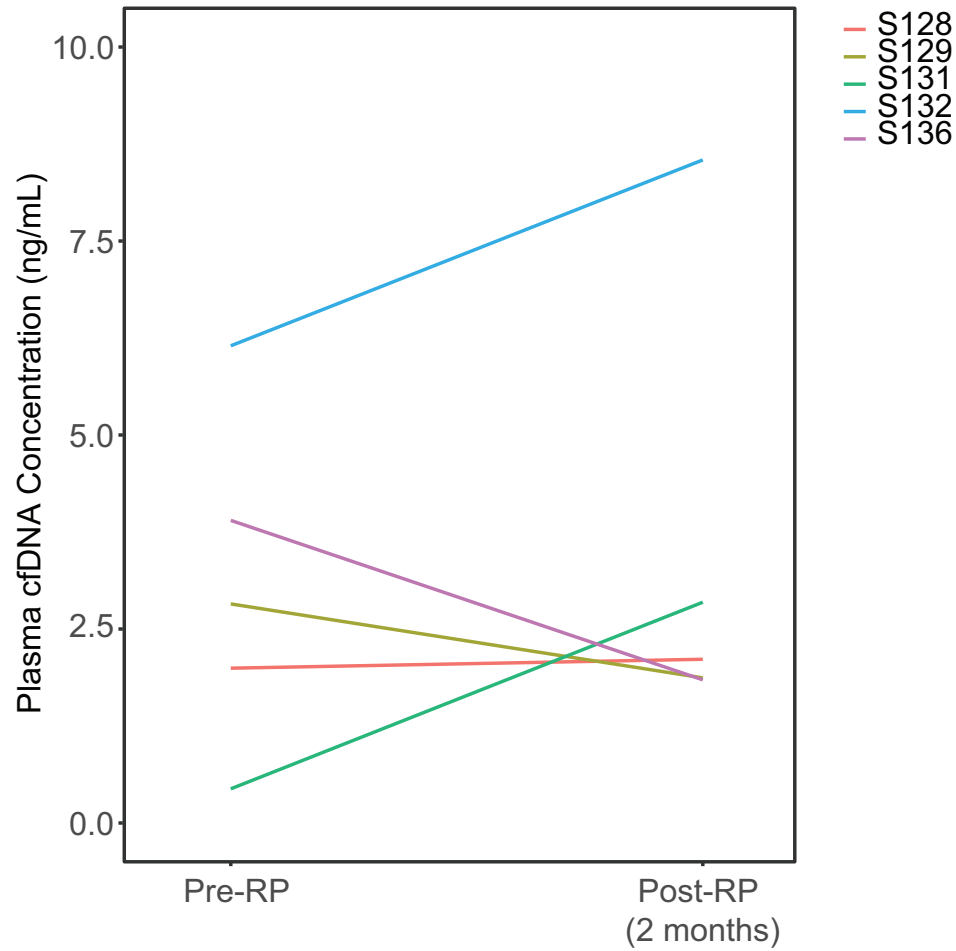
Figure S2.6 Blood samples were collected two months after RP for five patients. Plasma cfDNA concentration increased for two patients (S131 and S132). Patient S131 exhibited a decrease in average cfDNA fragment size from 206bp to 167bp, as well as an elevated PSA of 0.36 ng/mL 64 days after surgery, while the average cfDNA fragment size for patient S132 remained the same at 181bp and PSA levels remained low at 0.03 ng/mL.

Table S2.1 Clinical characteristics of individuals included in the cfDNA concentration analysis.

	Healthy N = 31	Localized N = 45	mCRPC N = 122
Age (years)			
Median $\pm$ IQR	57 $\pm$ 15	65 $\pm$ 10	71 $\pm$ 9
Range	41 – 72	43 – 78	47 – 91
Pathologic Gleason*			
6	-	4	-
7	-	28	-
8-10	-	12	-
Pathologic Stage			
Organ confined (pT2)	-	24	-
Not organ confined (pT3, pT4)	-	21	-
Extraprostatic extension (pT3a)	-	14	-
Seminal vesicle invasion (pT3b)	-	3	-
Lymph node involvement (N1)	-	5	-
PSA (ng/mL)			
Median $\pm$ IQR	-	6.7 $\pm$ 5.4	50.8 $\pm$ 128
Range	-	1.62 – 35.2	0 – 3007
cfDNA Concentration (ng/mL)			
Median $\pm$ IQR	7.9 $\pm$ 4.0	6.7 $\pm$ 5.8	13.8 $\pm$ 28.1
Range	0.29 – 16.9	1.22 – 53.9	1 – 1380

* One man with unknown data in the cohort.

Table S2.2 Association between log transformed cfDNA concentration or average cfDNA fragment size and categorical clinical features for patients with localized disease.

Clinical Feature	Number of Patients	Average cfDNA Concentration (ng/mL)	Average cfDNA Fragment Size (bp)
Pathologic Gleason score			
≤ 7 (3+4)	25	8.90	169.8
≥ 7 (4+3)	20	8.65	174.7
P		0.19	0.09
Organ confinement			
Confined (pT2)	24	8.25	172.7
Not confined (pT3, pT4)	21	9.40	172.8
P		0.80	0.69
Extraprostatic extension			
Negative (pT2)	24	8.25	172.7
Positive(pT3a)	14	12.33	170.7
P		0.29	0.67
Seminal vesicle invasion			
Negative (pT2, pT3a)	38	9.75	171.7
Positive (pT3b)	3	3.74	171.9
P		0.15	0.45
Pathologic lymph node			
Negative (pN0)	23	8.45	172.2
Positive (pN1)	5	8.49	173.8
P		0.90	0.69
Biochemical Recurrence			
No	37	7.62	172.5
Yes	5	20.3	171.5
P		0.24	0.41
Clinical T stage			
T1c*	20	9.47	169.5
T2a-c†	21	9.19	174.1
P		0.24	0.29

* Tumor identified by needle biopsy, but clinically inapparent and not palpable
† Tumor is palpable

Table S2.3 Association between cfDNA and continuous clinical features for patients with localized disease.

Clinical Feature		cfDNA Concentration	Average Fragment Size
Age at Diagnosis (years)			
Number of Patients	45		
Average (SD)	62.7 (7)		
P		0.35	0.60
PSA at Diagnosis (ng/mL)			
Number of Patients	45		
Average (SD)	8.7 (7)		
P		0.86	0.85
Decipher Score			
Number of Patients	12		
Average (SD)	0.66 (0.2)		
P		0.81	0.65
Time to Salvage Therapy (days)			
Number of Patients	39		
Average (SD)	1044 (926)		
P		0.63	0.25
Average Fragment Size (bp)			
Number of Patients	45		
Average (SD)	176 (21)		
P		0.67	-
CAPRA-S Score			
Number of Patients	45		
Average (SD)	3.7 (3)		
P		0.50	0.12

Chapter III

Cell-free DNA detection of tumor mutations in heterogeneous, localized prostate cancer via targeted, multi-region sequencing

3.1 Abstract

Purpose: Cell-free DNA (cfDNA) may allow for minimally invasive identification of biologically relevant genomic alterations and genetically distinct tumor subclones. Although existing biomarkers may detect localized prostate cancer, additional strategies interrogating genomic heterogeneity are necessary for identifying and monitoring aggressive disease that may require further intervention. In this study we aimed to evaluate whether circulating tumor DNA (ctDNA) can detect genomic alterations present in multiple regions of localized prostate tumor tissue.

Methods: Low-pass whole-genome and targeted sequencing with a machine-learning guided 2.5Mb targeted panel were used to identify single nucleotide variants (SNVs), small insertions and deletions (indels), and copy number alterations (CNAs) in cfDNA from 45 individuals: 15 healthy controls, 21 men with localized prostate cancer who underwent radical prostatectomy (RP), and 9 men with metastatic castration-resistant prostate cancer (mCRPC). Plasma cfDNA was barcoded with duplex Unique Molecular Identifiers (UMIs) to construct consensus reads and improve variant detection by leveraging duplicate reads and sequence complementarity of the two DNA strands. For localized cases, matched tumor tissue was collected from multiple regions (1 to 9 samples per patient) for comparison.

Results: Plasma cfDNA mutational burden was found to track with disease severity, and somatic variants present in heterogeneous tumor foci from patients with localized disease were detected in cfDNA. Somatic tissue alterations were identified in cfDNA, including nonsynonymous variants in *FOXA1, PTEN, MED12, and ATM,* and copy number loss in *CHD1*. Detection of these overlapping variants was associated with seminal vesicle invasion (P = 0.019) and with the number of variants initially found in the matched tumor tissue samples (P = 0.0005).

Conclusion: Our findings demonstrate the potential of targeted cfDNA sequencing to detect somatic tissue alterations in heterogeneous, localized prostate cancer, especially in a setting where matched tumor tissue may be unavailable (i.e. active surveillance or treatment monitoring).

3.2 Introduction

Prostate cancer accounts for approximately 20% of all new cancer diagnoses in American men, and an estimated 33,330 men will die from this disease in 2020[11]. While the majority of prostate cancer is diagnosed when the disease is still localized and are successfully treated, an estimated 2.9% of patients develop bone metastases and 2.4% die of prostate cancer within six years[12,54,55]. This is believed in part to be due to pathologically heterogeneous and genetically multiclonal disease, which likely determines available tumor escape mechanisms that allow for tumor survival and proliferation, subsequently driving disease progression and treatment outcome[9,16,56,57]. A variety of existing biomarkers in addition to prostate-specific antigen (PSA) have been critical in predicting treatment outcomes, including nomograms that incorporate clinical and pathologic factors, parametric MRI, and molecular testing[30,31,42]. However, tissue biopsies may miss some of the tumor and can lead to underestimation of disease grade and stage, motivating the need for additional modalities to comprehensively assess disease heterogeneity. This may be particularly important if tumor tissue is unavailable during active surveillance, or during disease monitoring after surgery—or other treatments—for detection of minimal residual disease or progression.

Multiple studies have tracked the evolutionary trajectory of localized prostate cancers, and have found a number of genomic factors to be predictive of poor outcomes. Specifically, genomic instability resulting from recurrent copy number alterations (CNAs) in genes such as *MYC, NKX3-1,* and *PTEN* is prognostic for biochemical recurrence following surgery or radiotherapy[13,58–60]. Prostate tumors also harbor a large proportion of somatic single nucleotide variants (SNVs) that are not protein-altering, and exhibit extensive intrafocal heterogeneity[61]. While nonsynonymous mutations have been found in *SPOP, FOXA1*, *MED12* and *ATM*, these recurrently mutated genes

are only found in less than 10% of patients[15,58]. Overall, polyclonal tumors with multiple tumor populations originating from a single clone are more likely to result in adverse outcomes[16].

Plasma cell-free DNA (cfDNA) remains a promising tool for directly detecting tumor DNA that is shed into the bloodstream. Both droplet digital PCR (ddPCR) and next-generation sequencing (NGS) have been successfully used to depict clonal evolution and identify genomic alterations in the context of early detection and disease monitoring. For example, personalized multiplex-PCR NGS of cfDNA has been used to derive tumor phylogenetic trees and characterize post-operative non-small cell lung cancer (NSCLC) relapse[62]. Targeted error correction sequencing (TEC-Seq) with dual-index barcodes to leverage duplicate fragments has been used to discover somatic alterations in early-stage cancers, including breast cancer, colorectal cancer, lung cancer, and ovarian cancer[63]. Detection of *BRCA2* reversion mutations in cfDNA have been associated with resistance to PARP inhibitors, allowing for monitoring of treatment resistance in patients with mCRPC[26]. Recently, ultra-low-pass whole-genome sequencing (ULP-WGS) and targeted resequencing were unable to detect tumor fragments in cfDNA in patients with localized prostate cancer[64]. While detection of somatic alterations in cfDNA in the localized disease setting may be more challenging due to low disease burden, this study utilizes advances in sample processing, library preparation, targeted panel design, and bioinformatic tools for broad yet sensitive variant detection in the cfDNA of a heterogenous disease such as localized prostate cancer[65].

In this study, extensive tissue sampling was used to capture tumor heterogeneity and provide a patient-specific gold standard for comparison of matched cfDNA. We performed both targeted and low-pass whole-genome sequencing of cfDNA and multi-region tumor tissue from surgically

resected prostate tissue from 21 men with localized prostate cancer, who underwent radical prostatectomy (RP) as primary treatment. This allowed for assessing the genomic heterogeneity of localized prostate cancer that is a result of clonal evolution. Next, we used a 2.5Mb targeted panel to also assess the mutational burden found in 15 healthy controls and 9 men with metastatic castration-resistant prostate cancer (mCRPC). Importantly, the targeted panel, as previously described, was generated without *a priori* patient-specific tumor mutational information in an attempt to capture a wide range of potentially important mutations, as well as reflect the scenario of repeated cfDNA blood collection in a clinical setting when tumor tissue biopsy is not possible[66]. Additionally, plasma cfDNA was barcoded with duplex Unique Molecular Identifiers (UMIs) to improve variant detection in a setting where the fraction of ctDNA can be low.

3.3 Materials and Methods

3.3.1 Patient Cohort

A total of 45 individuals recruited between August 2015 to November 2019 were included in this study: 15 healthy donors, 21 patients with localized prostate cancer, and 9 mCRPC patients (Table 1). Healthy control samples were collected from volunteers at UCSF, and all patients with localized disease underwent radical prostatectomy (RP) at UCSF. For the patients with localized disease, blood samples were collected the day of surgery before RP. Patients with mCRPC were included from the UCSF PROMOTE study investigating predictive markers of response. Blood samples, matched tissue from adjacent normal seminal vesicles, and multiple tumor regions (1 to 9 samples per patient) were collected from patients undergoing radical prostatectomy.

Clinicopathologic variables that play an important role in surgical management after prostatectomy were also collected for patients with localized disease, including pathologic T stage, tumor volume, and the Cancer of the Prostate Risk Assessment Postsurgical (CAPRA-S) score, which is a prediction model used to assess risk of recurrence after surgery and encompasses pre-surgical PSA level, pathologic Gleason score, presence of positive surgical margins, extracapsular extension, seminal vesicle invasion, and lymph node involvement[43]. Biochemical recurrence was defined as two consecutive PSA levels of >0.2 ng/mL at least eight weeks after surgery. Approval for this study was granted by the local ethics review board (IRB 11-05226, IRB 12-09659, and IRB 12-10340), and written informed consent was obtained from all patients.

3.3.2 Tissue Sample Collection

For the men with localized prostate cancer undergoing RP, multiple 3-mm punches were collected from the index lesion, regions with varying histology or Gleason grade, as well as other spatially distinct tumors from surgically resected prostates. Prostate tissue was cut into quadrants and snap-frozen in Optimal Cutting Temperature (OCT) compound, using isopentane chilled by dry ice. A cryostat was used to create 5μm sections. Tumor locations, Gleason grade, and tumor content were verified by a pathologist's examination of H&E stained sections of prostate tissue. Tissue punches were stored at -80°C. Matched normal tissue samples were also collected from nearby seminal vesicles or peripheral whole blood when the prior was unavailable. For the mCRPC patients, tissue samples were obtained using image-guided core needle biopsy of the metastatic lesion in the bone or soft tissue, and formalin-fixed, paraffin-embedded (FFPE) for histopathological review.

3.3.3 Tissue Processing and Sequencing

DNA was extracted from normal and tumor tissue using the Qiagen DNeasy Blood and Tissue Kit (Qiagen, Redwood City, CA). For each sample, genomic DNA was sheared with a Covaris M220 ultrasonicator to a target size of ~300bp, and assembled into a library with Illumina TruSeq adapters. For each sample, 10-100 ng of DNA was used for targeted and whole-genome sequencing (see Data Supplement). A set of custom MyBaits (Arbor Biosciences, Ann Arbor, MI) hybrid capture probes +175bp and tiled 3X were designed to target mutations in a custom panel (described below) and applied by Arbor Biosciences prior to sequencing to a target depth of 200X on an Illumina HiSeq 2500 at the CLIA-certified laboratory of the UCSF Institute for Human Genetics (IHG) Genomics Core Facility (CLIA #05D2080584). For mCRPC patients, four of the nine men had tumor tissue samples available for sequencing as a part of the UCSF PROMOTE study (NCT02735252).

3.3.4 Targeted Panel Design

We used a custom designed 2.5Mb targeted panel that included 7,034 mutations identified using a Support Vector Machine classification and ranking model, as previously described[66]. Briefly, this model was trained on whole-genome sequence data from 550 primary prostate tumors from the International Cancer Genome Consortium (ICGC; release 23), along with biological feature annotations (i.e. CADD and PhyloP deleterious measures, annotation in KEGG, amino acid identity, evolutionary conservation). The resulting panel included single point mutations as well as small <200 bp indels in both coding (83%) and noncoding regions (17%).

3.3.5 cfDNA Extraction and Quantification

For healthy controls, 20 mL of whole peripheral blood was collected in PAXgene Blood ccfDNA tubes (Qiagen, Redwood City, CA). From 13 to 25 mL of whole peripheral blood was collected immediately before surgery for patients with localized prostate cancer and before treatment initiation for mCRPC patients. Plasma was generated from whole blood samples within 2 hours for blood collected in K3EDTA tubes or within 7 days for blood collected in PAXgene Blood ccfDNA tubes (Qiagen, Redwood City, CA). We used a two-step centrifugation protocol: first centrifuging the blood at 1,900g for 10 min at 21°C, followed by centrifugation of the supernatant at 16,000g for 10 min to remove leukocytes and cellular debris. DNA was extracted from 7 to 29 mL of plasma using the Qiagen QIAamp Circulating Nucleic Acid Kit (Qiagen, Redwood City, CA), double eluted with 40 μL of Qiagen Elution Buffer, and analyzed on the Agilent 2100 Bioanalyzer with High Sensitivity DNA Chips (Agilent Technologies, Santa Clara, CA) for assessment of sample purity, concentration, and fragment size distribution according to the manufacturer's instructions. Plasma cfDNA concentration was determined with the Agilent 2100 Bioanalyzer Expert software, and calculated across the first three peaks (between 75–675bp) corresponding to the length of nucleosomal footprints and linkers derived from apoptotic cells[67].

3.3.6 cfDNA Library Preparation and Sequencing

A minimum of 10 ng of cfDNA from each sample was used to prepare sequencing libraries by concentrating the cfDNA with a Zymo Clean and Concentrator Kit (Zymo Research, Irvine, CA) and tagging molecules with unique molecular identifiers (UMI) with the ThruPLEX Tag-Seq 48S kit (Takara Bio, Mountain View, CA) prior to PCR amplification (7-11 cycles). The UMIs included two 6 nucleotide barcodes and two 8-11 nucleotide stems on each end of the insert, with

an 8 nucleotide Sanger index on the 3' end. Finally, samples were again analyzed on the Agilent 2100 Bioanalyzer after AMPure XP bead cleanup for quality control (Beckman Coulter, San Jose, CA). Hybrid capture with custom MyBaits (Arbor Biosciences, Ann, Arbor, MI) were applied to the libraries prior to sequencing to a target depth of 2,500X on the Illumina HiSeq 2500. Samples were also whole-genome sequenced to a target depth of 4X (see Data Supplement).

Due to the low tumor fractions typically found in localized prostate cancer, special consideration was given to UMI-tagged cfDNA sequencing depth calculations. The average sequencing depth can be defined theoretically as LN/G, where L is the read length, N is the number of reads, and G is the haploid genome length. For sequencing with the targeted panel, the on-target hybrid capture efficiency was estimated to be 40% with 10% duplicates. The number of total reads was found by identifying the minimum number of raw reads per UMI family necessary to generate consensus reads for variant calling, where a UMI family is a set of reads constructed from both strands of the original dsDNA molecule.

3.3.7 Tissue Sequencing Data Analysis

Quality assessment of sequence reads was first evaluated using FastQC, which includes metrics on per base quality, GC content, sequence length distribution, and overrepresented sequences. Whole-genome and targeted sequencing data were then analyzed using the pipelines developed by the Broad Institute of MIT and Harvard on the Terra platform with the GATK4.1 tools release.

All tissue sequencing data was preprocessed to produce analysis-ready BAM files prior to somatic variant calling and copy number analysis. Raw paired-end reads (150bp) in FASTQ format were

merged and aligned to the Genome Reference Consortium Human Build 37 (GRCh37) with BWA. Bases with a Phred quality score <20 were filtered out to remove poor quality reads, likely due to sequencing errors. Picard tools were used to sort, index, and merge files, as well as mark and remove duplicate reads that originated from the same DNA fragment, which are non-independent observations. Base quality scores were also recalibrated to correct for systematic errors to produce a final BAM file for further analysis.

For the samples sequenced with the targeted panel, MuTect2 was used to perform somatic variant calling on matched tumor-normal BAMs to detect SNVs and small INDELS, which utilized annotation files contained in the GATK bundle[68]. Variant filtering was performed to remove potential technical or germline artifacts, including cross-sample contamination. Variants were retained if the filter parameter was designated as "PASS", and subsequently annotated with Oncotator[69]. Manual review of the variants was performed with Integrative Genomics Viewer (IGV) to inspect variants for sequencing evidence.

Somatic copy number alterations (CNAs) in tumor tissue were identified in whole-genome sequence data using GATK ModelSegments, utilizing a panel of normals (PoN) generated from whole-genome normal samples sequenced at the Broad Institute Genomics Platform. When creating a genomic intervals list to define the resolution of the analysis, bin lengths were set to 1,000bp. Read coverage data is denoised against the PoN using principal component analysis, and both kernel segmentation and Markov-chain Monte Carlo are used with copy ratio and allelic counts data to group contiguous segments and make calls. Genomic instability was assessed with

the percentage of genome altered (PGA) metric, which was calculated by dividing by the number of base pairs affected by copy number changes by the total length of the genome.

3.3.8 cfDNA Sequencing Data Analysis

Plasma cfDNA barcoded with UMIs and sequenced with the targeted panel underwent variant calling using the Curio Genomic platform, which was specifically designed for processing UMI-tagged sequences prepared with the Takara ThruPlex Tag-seq kit. Raw paired-end reads were merged and aligned to GRCh37 with BWA, and the 6nt UMIs were extracted for downstream analysis. Duplex UMI processing was enabled to group reads from both strands of the original dsDNA molecule into UMI families. Consensus reads were generated from UMI families prior to variant calling with Curio version 1.4.1 with the following parameter thresholds: 1) a minimum base quality Phred score of 30, corresponding to 99.9% base call accuracy; 2) a minimum of four reads in every UMI family to filter out smaller families with few reads; 3) a minimum of 75% of the reads with the same base call in a UMI family at a given position; 4) an allowable UMI hamming distance of four bases to differ across both the read and its paired-end mate; 5) a minimum non-reference coverage or number of unique UMI families supporting the variant was set to three; and 6) an allele frequency less than 20% to exclude potentially homozygous and heterozygous germline variants.

Low-pass whole-genome sequencing data was used to identify large-scale CNAs and estimate the fraction of tumor in cfDNA using HMMcopy and ichorCNA[70]. Briefly, whole genomes were binned into 1Mb windows, and a Hidden Markov Model was used to segment the copy number

profile into regions predicted to be generated by the same copy number variant event, as well as identify copy number alterations for each segment.

3.3.9 Statistical Analysis

Since cfDNA variant counts were not normally distributed (P < 0.001, Shapiro–Wilk test), we evaluated the difference across healthy and prostate cancer groups using the Mann-Whitney non-parametric test. Correlations between clinical categorical variables (i.e. biochemical recurrence, seminal vesicle invasion, Gleason score ≤ 3+4 vs. ≥ 4+3) and somatic tumor tissue variant detection in cfDNA was assessed with Fisher's exact test. The non-parametric Mann-Whitney U-test was used to evaluate differences in cfDNA variant detection for continuous variables (i.e. starting amount of extracted cfDNA, number of tumor tissue or cfDNA variants, CAPRA-S score, pathological tumor volume, and sequencing depth of coverage). All statistical analyses were performed using R version 3.6.1.

3.4 Results

3.4.1 Plasma cfDNA mutational burden and prostate cancer

The median cfDNA variant count was 1,089 (IQR = 761) for non-diseased controls, 1,843 (IQR = 605) for men with localized prostate cancer, and 5,081 (IQR = 716) for men with mCRPC. The average cfDNA variant count for men with localized prostate cancer was statistically significantly higher than those observed in non-diseased controls (Table 1; Figure 2; P < 0.01). Men with mCRPC had a statistically significantly higher cfDNA variant count than men with localized disease (P < 0.0001) or controls (P < 0.0001).

3.4.2 Genomic heterogeneity in localized prostate cancer

A total of 21 men had between 1 and 9 samples collected from distinct tumor foci (71 specimens total), which were then sequenced with our targeted panel in an effort to capture regions with varying histology and identify potential clonal and subclonal mutations. Targeted sequencing identified somatic variants in all 71 prostate cancer specimens. A median of 8 variants with a range from 1 to 1,091 variants (IQR = 8) were identified across all foci (Figure 3a). While 17% of the target panel was comprised of noncoding variants, 88% of the tissue variants identified were in noncoding regions.

Nonsynonymous variants were identified in 22 of the 71 tissue specimens (Figure 3b). Alterations were discovered among commonly mutated genes in localized prostate cancer, including *FOXA1, PTEN, MED12,* and *ATM*. The majority of somatic mutations observed in patients with multiple regions sequenced were private to each tumor focus, with a subset of mutations present in all regions. In one patient, the potentially clonal mutation identified in all six tissue regions was found in *ATM* (Figure 4).

Tissue samples collected from five patients underwent WGS at 4X coverage, and were found to harbor a median of 18 copy number alterations with a range from 2 to 626 copy number alterations (IQR = 153). Across all patients, 2% to 25% of the copy number changes were likely clonal and found in all foci for a given patient. For each patient, copy number alterations are shown in Supplementary Figure 1, with multiple tracks overlaid for different samples. Samples had a median PGA of 9% with a range from 0.002% to 17.8% (IQR = 4.8%). Loss of *CHD1, NKX3-1, CDKN1B, PTEN,* and *TP53* were found in all sequenced tissue regions for two of the five patients. Notably,

MYC amplification was found in a subset of the regions for three patients, and found to co-occur with PTEN loss in one patient. However, somatic copy number alterations were not detected in the cfDNA of patients with localized prostate cancer.

3.4.3 Somatic tumor tissue variants identified in cfDNA with 2.5Mb targeted panel sequencing in localized prostate cancer

Somatic tissue variants were identified in cfDNA sequenced without prior knowledge of the variants present in tumor tissue. A matched source of normal tissue was used to exclude germline variants from the analysis. Overlapping variants were found in 12 of the 21 patients, with a range of 1 to 62 variants (Figure 5). For the majority of the patients, overlapping variants were subclonal and found in a subset of the tumor tissue regions sequenced. While the targeted panel was composed of 17% noncoding variants, 85% of the overlapping variants are found in intronic and intergenic regions, which is comparable to the 88% noncoding variants found in tumor tissue. The cfDNA variant allele frequency (VAF) for the overlapping variants ranged from 0.9% and 19% (Figure 6).

3.4.4 Determinants of somatic tissue variant detection in cfDNA for localized prostate cancer

Our ability to detect some of the observed somatic variants in cfDNA was positively associated with the number of variants co-identified in tumor tissue ($P = 0.005$) and with seminal vesicle invasion ($P = 0.019$). There was no clear pattern of association observed with the other clinical factors. However, the small number of patients with overlapping variants may not be sufficiently powered to detect a difference between groups (Table 2; Table 3).

3.5 Discussion

We found that targeted sequencing of cfDNA—without *a priori* patient-specific tumor mutation information—identified somatic alterations found in matched tumor tissue from multiple regions, potentially allowing for dynamic monitoring of emerging resistant subclones throughout the course of disease. Detection of these concordant variants was associated with seminal vesicle invasion and the number of somatic variants initially found in the tumor tissue samples, predicating its use for patients with poor prognostic factors in a localized setting. Our study demonstrates the ability of cfDNA to detect a portion of the genomic heterogeneity present in localized prostate cancer, with implications in clonal architecture reconstruction for identification of patients with increased risk of relapse.

Interestingly, clinicopathologic factors may be important in the detection of somatic tissue variants in plasma cfDNA. Specifically, tissue variants were found in the cfDNA of three men with the highest diagnostic PSA levels (ranging from 39 ng/mL to 70 ng/mL) who also had high Gleason scores of nine. Two of these patients had metastasis to nearby lymph nodes and pathologic T4 staging, with the tumor invading nearby structures beyond adjacent seminal vesicles. Notably, two of these patients had the two highest tumor volumes measured, with the third patient's measurement unavailable for comparison.

Technical considerations related to library preparation and sequencing are also a factor in cfDNA detection of tissue variants. A majority of the tumor tissue variants that were not detected in cfDNA had zero read coverage, suggesting that cfDNA fragments with the variants may not have been present initially during cfDNA extraction or were present, but either not captured during the

hybridization step or failed to bind to the flow cell prior to sequencing. Other tissue variants that were not detected in cfDNA had sufficient total coverage, but not enough UMI family coverage supporting the alternate allele to be called as a variant (Figure 7).

Importantly, many of the somatic tumor tissue alterations identified in this study's cohort of localized cases were also previously identified in large cohorts of patients with localized prostate cancer, providing a patient-matched gold standard for comparison with cfDNA variants detected in this study. A fraction of these alterations were found to be common across all tissue samples, with others identified in only a subset of the samples, confirming the importance of comprehensive sampling for accurate grading and staging.

Nonsynonymous tumor tissue variants were identified in several well-characterized genes, including in *ATM, FOXA1, PTEN,* and *MED12*. In one patient, a likely clonal mutation in *ATM* was detected in all six tissue regions sampled; this gene is involved in cell cycle regulation and maintenance of genomic integrity as a part of the DNA repair pathway[71]. Mutations in *FOXA1* have been associated with increased AR-driven transcription, while *PTEN* is known to play a role in cell migration and DNA repair as an effector of the PI3K signaling pathway. Mutations in *MED12* have also been found to perturb CDK8-dependent modulation of transcription related to p53 and androgen signaling[72]. For patients with mCRPC, nonsynonymous variants were found in *TP53, CDK12, PTEN,* and *AR*, which have also been previously reported[73].

Copy number loss at a locus on chr5 q15–q21.1 containing *CHD1* was also identified in both available tissue samples for two of the patients with localized disease. Loss in *CHD1* has been

shown to sensitize cells to DNA double-strand breaks, leading to increased sensitivity to DNA damaging therapies such as PARP inhibitors, resulting in a synthetic lethal response. In another patient, one of the three tissue regions sampled harbored both *MYC* amplification and *PTEN* loss, which is associated with increased genomic instability and is prognostic for biochemical recurrence[14]. This patient's tumor had the highest PGA at 17.8%.

While 17% of our 2.5Mb targeted panel was comprised of noncoding variants, 88% of the variants found in the tumors of men with localized disease were in noncoding regions. While the impact of these specific alterations remains largely unknown, prior studies have discovered cancer driver noncoding elements in regulatory regions (i.e. promoters and enhancers) and noncoding SNVs that alter RNA secondary structures[74–77]. In a recent study, noncoding mutations were found to target cis-regulatory elements of *FOXA1* and modulate the binding of transcription factors, exposing a potential therapeutic target and highlighting the importance of mutations in untranslated regions[77]. Similarly to the percentage of noncoding variants detected in the tumors of men with localized prostate cancer, 85% of the overlapping variants identified in both cfDNA and matched tumor tissue were found in noncoding regions. Both likely clonal and subclonal mutations were identified, supporting the ability of cfDNA to capture somatic alterations from multiple tumor cell populations and detecting intrapatient heterogeneity.

To ensure that the tissue variants identified in cfDNA were not clonal hematopoietic mutations of indeterminate potential (CHIP), which typically accumulate during the aging process, we looked for the presence of alterations in genes commonly associated with CHIP[78]. Variants were found in *DNMT3A, ASXL1, TET2, and NOTCH2* in the cfDNA of healthy patients, in white blood cells

from patients with localized disease, and in cfDNA from patients with localized disease, which is expected since the majority of cfDNA is derived from hematopoietic cells; however, variants overlapping between cfDNA and localized tumor tissue were not found in these genes, suggesting that the CHIP effect does not explain our findings[79]. Interestingly, 5% of the cfDNA variants found in coding regions that were not present in tumor tissue were identified in genes previously found to be mutated in prostate cancer, including *AR, ATM, BRCA2, BRAF, CDK12, CHEK2, IDH1, PIK3CA, MYC,* and *FOXA1*[58].

In undertaking this study we leveraged a number of methods to ensure broad and sensitive detection of cfDNA variants for patients with localized prostate cancer. Relatively large blood volumes, between 13 and 25 mL, were collected and centrifuged with an initial low spin at 1,900g followed by a high spin at 16,000g to pellet and remove leukocytes and cellular debris. During library preparation, UMIs were used to barcode individual cfDNA fragments and take advantage of sequence complementarity of the double-stranded DNA and duplicates that arise during amplification. Importantly, the 2.5Mb targeted panel used in this study was generated by using a classification and ranking model trained on WGS data from 550 prostate tumors in ICGC, and included both coding and noncoding regions. This optimized the composition of the panel to capture the heterogeneity previously seen in localized prostate cancer, while limiting the panel in size to allow for higher coverage at a lower sequencing cost. Additionally, analysis of cfDNA variants was performed with matched normal samples to filter out germline and clonal hematopoiesis variants, and compared to matched tumor tissue from multiple regions to confirm variant calls.

There are a number of limitations to this study that merit consideration. First, the relatively short follow-up time for a protracted disease like localized prostate cancer means detection of relapse may be difficult. After surgery, 4 of the 21 patients experienced biochemical recurrence with a median follow-up time of 2.34 years and a range of 66 to 1502 days. Second, somatic tissue variants were detected in cfDNA for only a subset of our study subjects. While factors including seminal vesicle invasion and tumor mutational burden were predictors of detection, with improved cfDNA detection likely due to the increased number of variants present in the tissue, patients with earlier stage disease may have fewer tissue variants detected in their cfDNA. Another factor is the selection of a matched source of normal tissue for the exclusion of germline and CHIP variants during analysis. We used normal tissue from nearby seminal vesicles for 17 of 21 patients, which may have a genomic profile that is more similar to the prostate tissue than to patient-matched whole blood. Consequently, CHIP variants may remain after cfDNA variant calling, and somatic alterations arising from mosaicism, a process where mutations occur during development and propagate to a subset of tissues, may be removed.

In previous studies, features identified in the subclonal architecture of localized prostate cancer have been found to identify patients at higher risk of relapse. In one study, almost two-thirds of men with localized prostate cancer had tumors that harbored multiple subclones, and these men relapsed following treatment at a much higher rate than men with monoclonal tumors[80]. In this study, among men with localized disease, all patients with more than one sequenced tissue region had tumors that harbored subclonal mutations, suggesting that these men have an increased risk of subsequent relapse after surgery. To this end, they may benefit from the use of cfDNA in determining the phylogeny of their tumor mutations to assess evolutionary features associated with

aggressive disease, which may help distinguish between patients who should be treated immediately from those who should stay on active surveillance or continued monitoring.

3.6. Conclusion

In summary, we show that targeted sequencing of cfDNA without prior patient-specific tumor mutation information can be used to identify somatic alterations, with implications for disease monitoring and detection of emerging subclones through repeat sampling. Targeted sequencing of cfDNA molecules can detect both potentially clonal and subclonal somatic tissue variants, with clinicopathologic and technical factors influencing detection. Future studies investigating the regulatory role of noncoding somatic mutations in localized prostate cancer will help elucidate the functional impact of cfDNA detection of these types of alterations. Combined with previous studies, the importance of detecting somatic alterations using cfDNA in localized disease is emerging. Our study supports the use of cfDNA in the assessment of heterogeneous, localized prostate cancer, which will be further strengthened by ongoing technological advances to enrich for tumor fragments found in cfDNA .

3.7 Acknowledgments

3.8 Tables

Table 3.1 Clinical characteristics of individuals included in the study at baseline.

	Healthy N = 15	Localized N = 21	mCRPC N = 9
Age (years)			
Median $\pm$ IQR	33 ± 19	66 ± 10	63 ± 3
Range	$22 - 63$	$50 - 74$	$59 - 75$
Pathologic Gleason			
6	-	1	-
7	-	11	-
8-10	-	9	-
Pathologic Stage			
Organ confined (pT2)	-	6	-
Not organ confined (pT3, pT4)	-	15	-
Extraprostatic extension (pT3a)	-	9	-
Seminal vesicle invasion (pT3b)	-	4	-
Lymph node involvement (N1)	-	6	-
PSA (ng/mL)			
Median $\pm$ IQR	-	9.1 ± 12	$34.5 \pm 67^{\dagger}$
Range	-	$2 - 69.9$	$0 - 263$
Number of Tissue Samples			
Median	-	3	-
Range	-	$1 - 9$	-
Tumor Tissue Variants*			
Median $\pm$ IQR	-	8 ± 8	-
Range	-	$1 - 1091$	-
cfDNA Variants*			
Median $\pm$ IQR	1089 ± 761	1843 ± 605	5081 ± 716
Range	$598 - 2423$	$1172 - 2595$	$4285 - 5938$

† One man with unknown data in the cohort
* Sequenced with 2.5Mb targeted panel

Table 3.2 Fisher's exact test results from investigating the correlation between categorical features and detection of tumor tissue variants from targeted cfDNA sequencing in patients with localized prostate cancer.

Clinical Feature	No Detection in cfDNA	Detection in cfDNA	P-value
Pathologic T Stage			
$\leq$ T3a (No SVI)	9	6	0.019
$\geq$ T3b	0	6	
Gleason			
$\leq$ 7 (3 + 4)	4	5	1.0
$\geq$ 7 (4 + 3)	5	7	
Biochemical Recurrence			
No	8	6	0.087
Yes	1	6	
Number of Tissue Samples			
< 3	4	4	0.67
$\geq$ 3	5	8	

Table 3.3 Mann-Whitney U-test results assessing the association between continuous clinical features and detection of tumor tissue variants from targeted cfDNA sequencing in patients with localized prostate cancer. Tumor tissue variant count was significantly associated with detection in cfDNA.

Clinical Feature	No Detection in cfDNA	Detection in cfDNA	P-value
Tumor Tissue Variant Count			
Average	14	597	0.005
Starting Amount of cfDNA (ng)			
Average	18.5	26.1	0.25
CAPRA-S Score			
Average	4	6.6	0.08
Tumor Volume (cc)			
Average	4	4	0.61

3.9 Figures
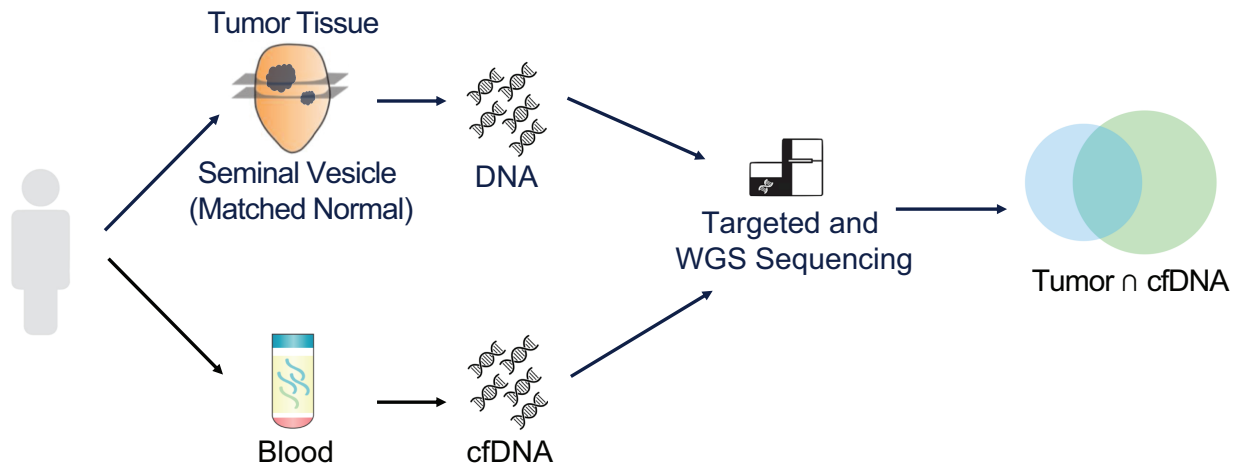
Figure 3.1 A total of 71 tissue specimens from surgically resected prostates were collected from 21 men. For each patient, multiple tumor tissue samples were obtained from regions with varying histology when possible, and matched normal tissue was collected from nearby seminal vesicles or peripheral whole blood when the prior was unavailable. Venous whole blood was drawn in K3EDTA or PAXgene ccfDNA tubes for all patients. DNA extraction and library preparation were then performed prior to targeted sequencing with a 2.5Mb panel or whole-genome sequencing (WGS) to assess genomic heterogeneity among localized prostate tumors. Additionally, cfDNA molecules were barcoded with unique molecular identifiers (UMIs) to group reads from both strands of the original dsDNA molecule into UMI families during variant calling.
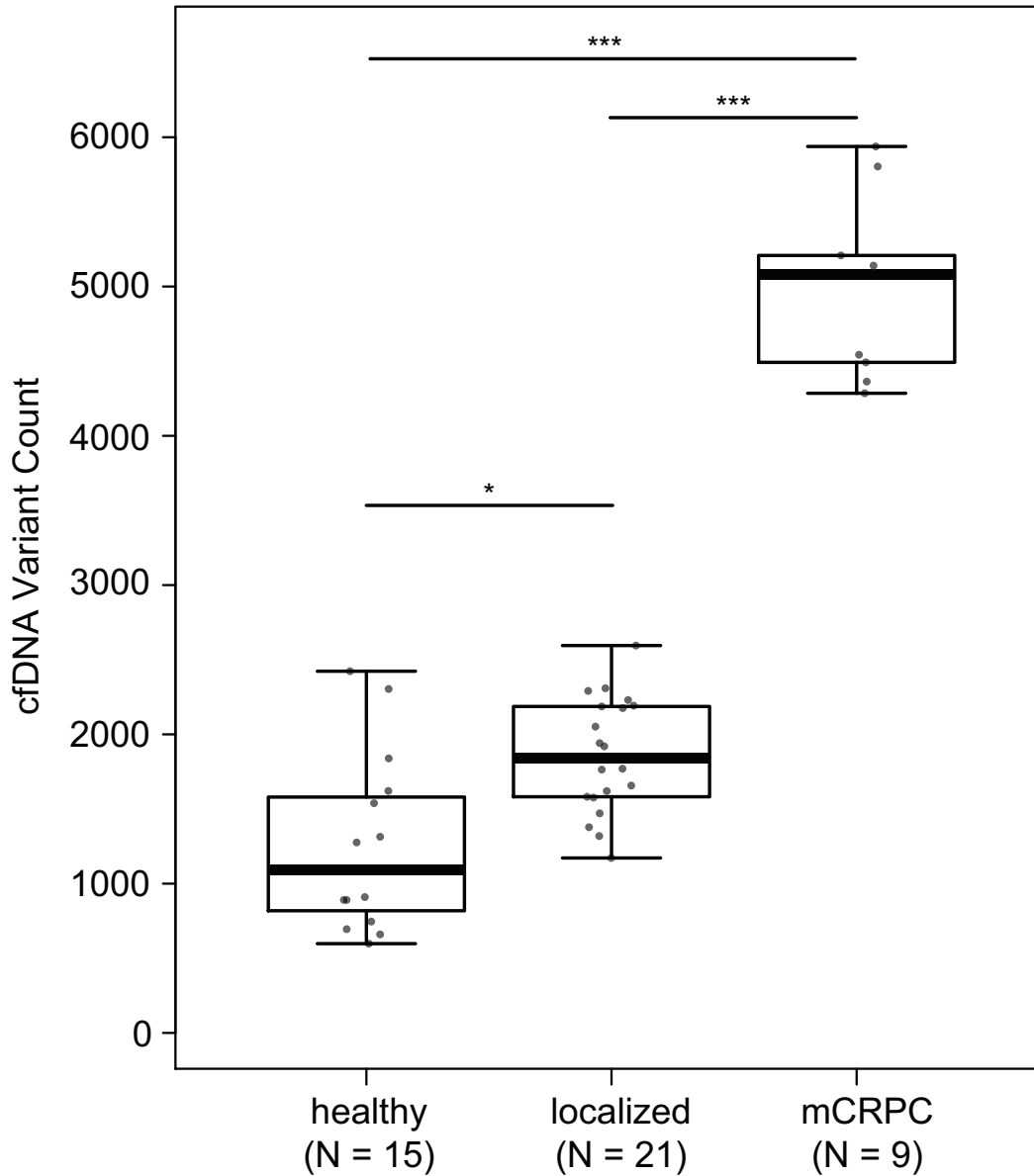
Figure 3.2 Plasma cfDNA mutational burden assessed by targeted sequencing increases with disease severity. Boxplots show the distribution in cfDNA variant count across healthy controls (N = 15), patients with localized disease (N = 21), and patients with mCRPC (N = 9) from blood samples collected at baseline prior to surgery or treatment initiation. Men with localized disease had significantly higher counts than those observed in controls (P < 0.01), and men with mCRPC had significantly higher counts compared to those found in men with localized disease (P < 0.0001) or controls (P < 0.0001).

* P < 0.01 *** P < 0.0001

A
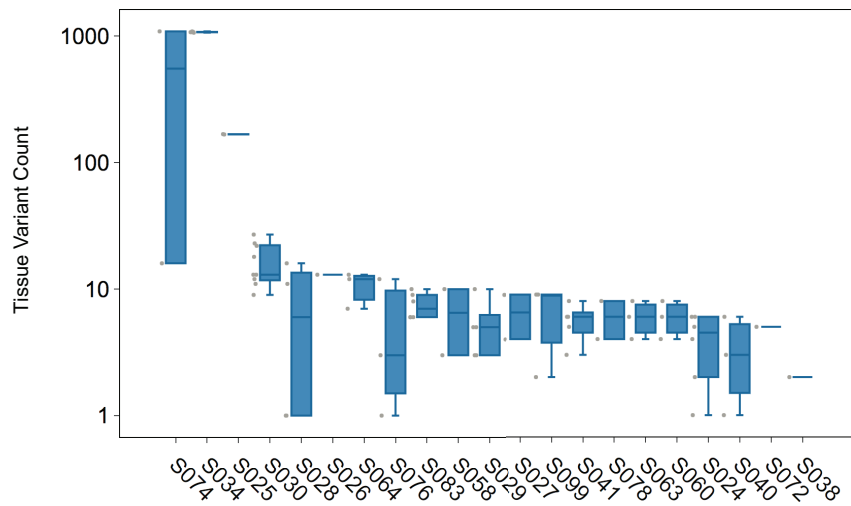
B

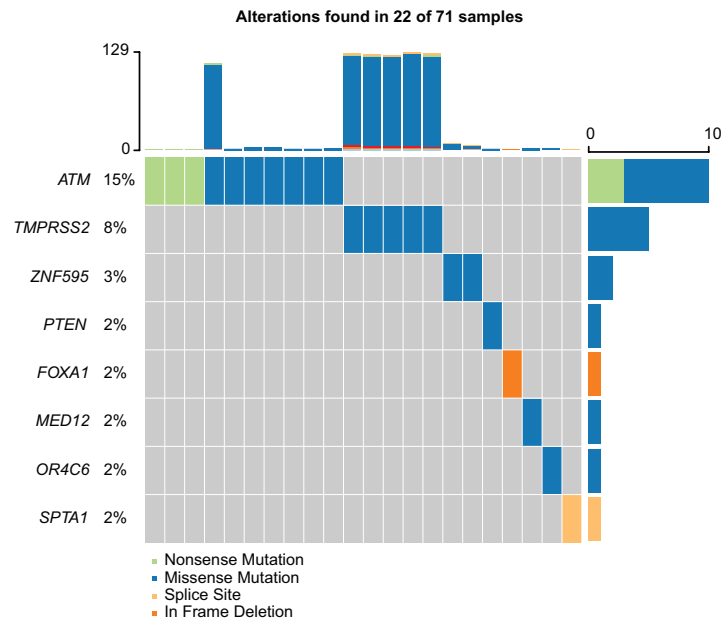**Figure 3.3** Localized prostate cancer tumors harbor a wide range of somatic variants both across patients and within a patient's tumor foci. (a) Distribution of variant counts for a cohort of 21 men with one to nine tumor tissue samples collected from regions of varying histology that were sequenced with a targeted panel. Patients are on the x-axis and each dot represents the variant count for a single tissue sample and tissue variant counts are on the log-scaled y-axis. A median of 8 variants with a range from 1 to 1,091 variants identified across foci. Matched whole blood was used as a source of normal for patients S025, S034, S041, and S060. (b) Nonsynonymous variants in the listed genes were detected in 22 of the 71 tissue samples. Top, the number of variants in each sample. Left, gene names and percentage of samples with mutations in a given gene. Center, mutations colored by coding consequence. Right, number of mutations in given gene.
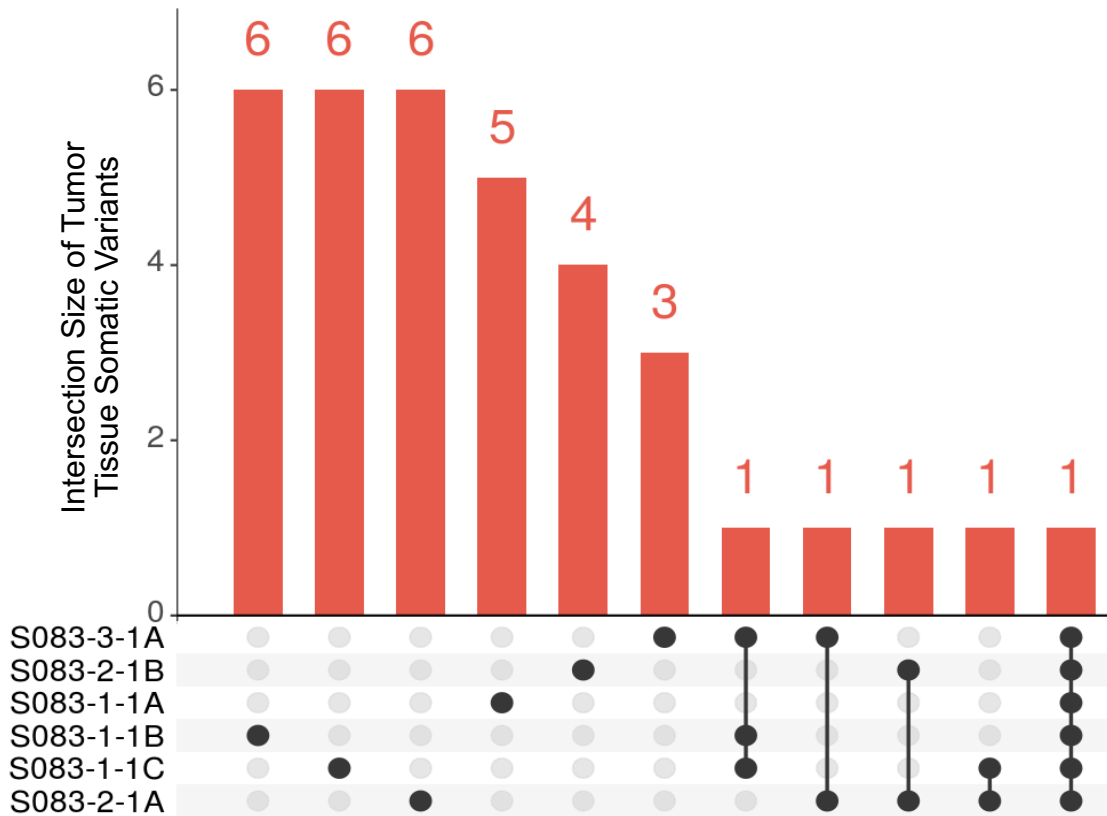
Figure 3.4 Within a single patient, there appears to be likely clonal and subclonal mutations. Rows are tumor tissue samples and bars show the number of variants that are either private to each sample or are shared among multiple tissue samples. For this patient, the mutation shared among all six samples is found in *ATM*, which is commonly found in localized prostate cancer and known to play an important role in cell cycle regulation and maintenance of genomic integrity.
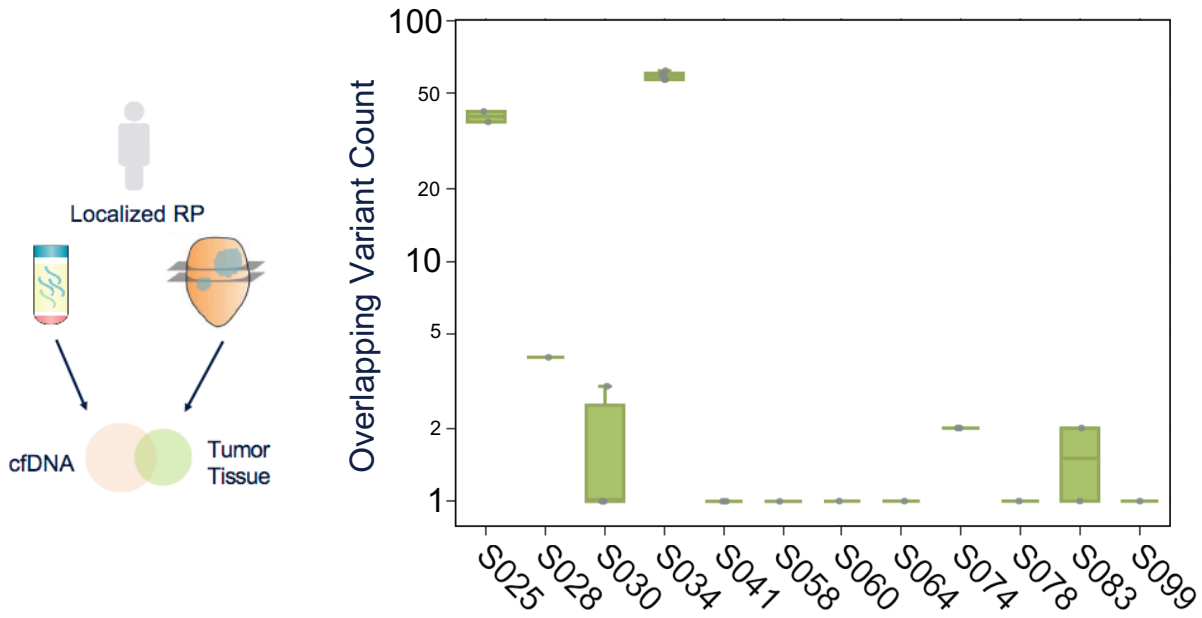
Figure 3.5 Somatic tissue variants were detected through targeted cfDNA sequencing without prior identification of variants present in tumor tissue. Boxplots show the distribution of variants overlapping between cfDNA and tumor tissue. Patients are on the x-axis and each dot represents the overlapping variant count for a single tissue sample, with counts shown on the log-scaled y-axis. Overlapping variants were detected in 12 of the 21 patients sequenced with the targeted panel, and ranged from 1 to 62 variants.
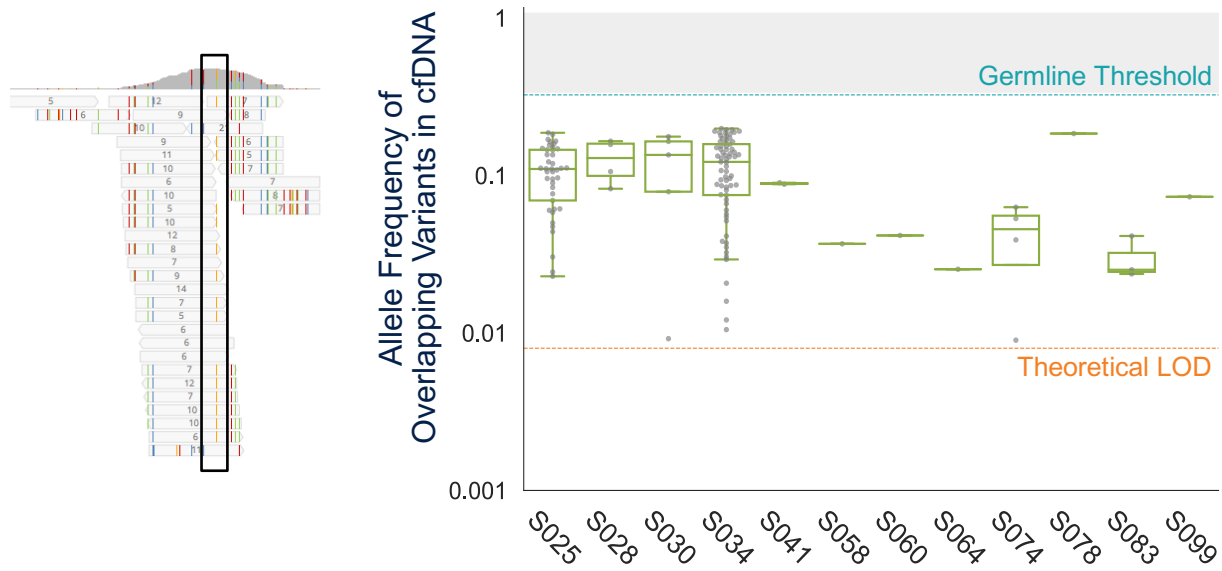
Figure 3.6 Variant allele frequency (VAF) of overlapping variants in cfDNA ranged from 0.9% to 19%. Left, schematic of an example cfDNA variant highlighted by a vertical orange bar and the number of reads supporting the call. Numbers reflect the number of reads in the same UMI family used for consensus calling. Right, boxplots show the allele frequency distribution for overlapping variants on a log-scaled y-axis. Each dot represents a variant identified in a given patient.

Detecton in cfDNA	Chromosome	Position	Total Coverage	Variant Coverage
Yes	chr2	91775688	434	4
Yes	chrY	10037785	74	10
Yes	chrY	10037787	76	6
Yes	chrY	10037790	79	13
Yes	chr19	53357000	17	3
No	chr1	152127865	412	1
No	chr2	152127871	0	0
No	chr3	158450451	0	0
No	chr4	158812248	0	0
No	chr5	97552665	0	0
No	chr5	121739540	0	0
No	chr6	75814974	0	0
No	chr7	100678421	0	0
No	chr8	100678426	0	0

Figure 3.7 Total coverage and variant allele coverage in cfDNA were potential determinants of somatic tissue variant detection in cfDNA. Shown here is a subset of somatic tissue variants for a single patient and whether or not the variants were detected in cfDNA. A majority of variants that were not detected in cfDNA had zero coverage or sufficient total coverage but not enough UMI family coverage supporting the alternate allele to be called as a variant.

3.10 Supplementary Materials
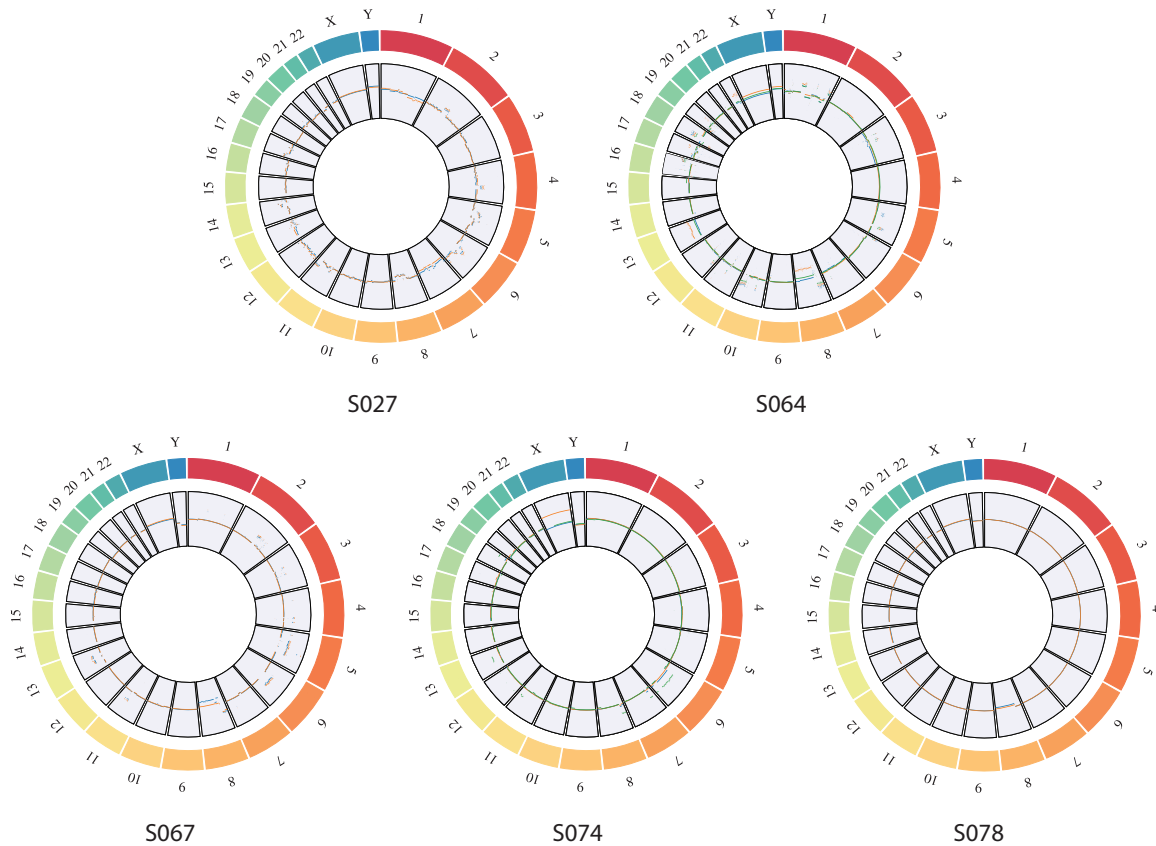
Figure S3.1 Whole-genome sequencing was performed for twelve tissue specimens from five patients with localized prostate cancer. Each Circos plot depicts the genomic location in the outer ring and chromosomal log2 copy number in the inner ring, with multiple samples overlaid for the same patient. Likely clonal and subclonal copy number alterations were identified *CHD1, NKX3-1, CDKN1B, MYC, PTEN,* and *TP53*. In one patients, one of the three tissue samples harbored both *MYC* amplification and *PTEN* loss, which is prognostic for biochemical recurrence and 17.8% of the genome was affected by copy number changes.

References

1. Lewis, R. *Human Genetics: Concepts and Applictaions*. (2010).

2. Crick, F. & Watson, J. Molecular structure of nucleic acids. (1953).

3. Wetterstrand KA. DNA Sequencing Costs: Data from the NHGRI Genome Sequencing Program (GSP) available at www.genome.gov/sequencingcostsdata. *Natl. Hum. Genome Res. Inst.*

4. Metzker, M. L. Sequencing technologies the next generation. *Nat. Rev. Genet.* **11**, 31–46 (2010).

5. Mardis, E. R. A decade's perspective on DNA sequencing technology. *Nature* **470**, 198–203 (2011).

6. Howlader, N. & Cronin, K. SEER Cancer Statistics Review, 1975-2017. *Natl. Cancer Inst.* (2020).

7. Balmain, A. Cancer as a complex genetic trait: tumor susceptibility in humans and mouse models. *Cell* **108**, 145–152 (2002).

8. Amirouchene-Angelozzi, N., Swanton, C. & Bardelli, A. Tumor evolution as a therapeutic target. *Cancer Discov.* **7**, 805–817 (2017).

9. Boutros, P. C. *et al.* Spatial genomic heterogeneity within localized, multifocal prostate cancer. *Nat. Genet.* **47**, 736–745 (2015).

10. Cooper, C. S. *et al.* Analysis of the genetic phylogeny of multifocal prostate cancer identifies multiple independent clonal expansions in neoplastic and morphologically normal prostate tissue. *Nat. Genet.* **47**, 367–72 (2015).

11. Siegel, R. L., Miller, K. D. & Jemal, A. Cancer statistics, 2020. *Cancer J Clin* **70**, 7–30 (2020).

12. Cooperberg, M. R., Broering, J. M. & Carroll, P. R. Risk assessment for prostate cancer metastasis and mortality at the time of diagnosis. *J. Natl. Cancer Inst.* **101**, 878–887 (2009).

13. Locke, J. A. *et al.* NKX3.1 haploinsufficiency is prognostic for prostate cancer relapse following surgery or image-guided radiotherapy. *Clin. Cancer Res.* **18**, 308–316 (2012).

14. Zafarana, G. *et al.* Copy number alterations of c-MYC and PTEN are prognostic factors for relapse after prostate cancer radiotherapy. *Cancer* **118**, 4053–4062 (2012).

15. Fraser, M. *et al.* Genomic hallmarks of localized, non-indolent prostate cancer. (2017). doi:10.1038/nature20788

16. Melijah, S. *et al.* The Evolutionary Landscape of Localized Prostate Cancers Drives Clinical Aggression. *Cell* **173**, (2018).

17. Mandel, P. & Metais, P. Les acides nucleiques du plasma sanguin ches l'homme. *C R Seances Soc Biol Fil.* **142**, 241–243 (1948).

18. Stroun, M. *et al.* Neoplastic characteristics of the DNA found in the plasma of cancer patients. *Oncology* **46**, 318–22 (1989).

19. Jahr, S. *et al.* DNA fragments in the blood plasma of cancer patients: Quantitations and evidence for their origin from apoptotic and necrotic cells. *Cancer Res.* **61**, 1659–1665 (2001).

20. Stroun, M., Lyautey, J., Lederrey, C., Olson-Sand, A. & Anker, P. About the possible origin and mechanism of circulating DNA: Apoptosis and active DNA release. *Clin. Chim. Acta* **313**, 139–142 (2001).

21. Sorenson, G. D. *et al.* Soluble Normal and Mutated DNA Sequences from Single-Copy Genes in Human Blood. *Cancer Epidemiol. Biomarkers Prev.* **3**, 67–71 (1994).

22. Lui, Y. Y. N. *et al.* Predominant hematopoietic origin of cell-free dna in plasma and serum after sex-mismatched bone marrow transplantation. *Clin. Chem.* **48**, 421–427 (2002).

23. Bettegowda, C. *et al.* Detection of Circulating Tumor DNA in Early- and Late-Stage Human Malignancies. *Sci. Transl. Med.* **6**, 224ra24-224ra24 (2014).

24. Bronkhorst, A. J., Ungerer, V. & Holdenrieder, S. The emerging role of cell-free DNA as a molecular marker for cancer management. *Biomol. Detect. Quantif.* **17**, (2019).

25. Bastian, P. J. *et al.* Prognostic value of preoperative serum cell-free circulating DNA in men with prostate cancer undergoing radical prostatectomy. *Clin. Cancer Res.* **13**, 5361–5367 (2007).

26. Quigley, D. *et al.* Analysis of Circulating Cell-free DNA Identifies Multi-clonal Heterogeneity of BRCA2 Reversion Mutations Associated with Resistance to PARP Inhibitors. *Cancer Discov.* CD-17-0146 (2017). doi:10.1158/2159-8290.CD-17-0146

27. Greytak, S. R. *et al.* Harmonizing Cell-Free DNA Collection and Processing Practices through Evidence-Based Guidance. *Clin. Cancer Res.* **26**, 3104–3109 (2020).

28. Kwapisz, D. The first liquid biopsy test approved. Is it a new era of mutation testing for non-small cell lung cancer? *Ann. Transl. Med.* **5**, 1–7 (2017).

29. Siegel, R. L., Miller, K. D. & Jemal, A. Cancer statistics, 2019. *CA. Cancer J. Clin.* **69**, 7–34 (2019).

30. Lilja, H., Ulmert, D. & Vickers, A. J. Prostate-specific antigen and prostate cancer: Prediction, detection and monitoring. *Nat. Rev. Cancer* **8**, 268–278 (2008).

31. Carroll, P. H. & Mohler, J. L. NCCN guidelines updates: Prostate cancer and prostate cancer early detection. *JNCCN J. Natl. Compr. Cancer Netw.* **16**, 620–623 (2018).

32. Koffler, D., Agnello, V., Winchester, R. & Kunkel, H. G. The occurrence of single-

stranded DNA in the serum of patients with systemic lupus erythematosus and other diseases. *J. Clin. Invest.* **52**, 198–204 (1973).

33.   Tan, E. M., Schur, P. H., Carr, R. I. & Kunkel, H. G. Deoxybonucleic acid (DNA) and antibodies to DNA in the serum of patients with systemic lupus erythematosus. *J. Clin. Invest.* **45**, 1732–1740 (1966).

34.   Tissot, C. *et al.* Circulating free DNA concentration is an independent prognostic biomarker in lung cancer. *Eur Respir J* **46**, 1773–1780 (2015).

35.   Ellinger, J. *et al.* The role of cell-free circulating DNA in the diagnosis and prognosis of prostate cancer. *Urol. Oncol. Semin. Orig. Investig.* **29**, 124–129 (2011).

36.   Jung, K. *et al.* Increased cell-free DNA in plasma of patients with metastatic spread in prostate cancer. *Cancer Lett.* **205**, 173–80 (2004).

37.   Feng, J. *et al.* Plasma cell-free DNA and its DNA integrity as biomarker to distinguish prostate cancer from benign prostatic hyperplasia in patients with increased serum prostate-specific antigen. *Int Urol Nephrol* **45**, 1023–1028 (2013).

38.   Lapin, M. *et al.* Fragment size and level of cell-free DNA provide prognostic information in patients with advanced pancreatic cancer. *J. Transl. Med.* **16**, 1–10 (2018).

39.   Mouliere, F. *et al. Enhanced detection of circulating tumor DNA by fragment size analysis. Sci. Transl. Med* **10**, (2018).

40.   Jiang, P. *et al.* Lengthening and shortening of plasma DNA in hepatocellular carcinoma patients. *Proc. Natl. Acad. Sci. U. S. A.* **112**, E1317–E1325 (2015).

41.   Underhill, H. R. *et al.* Fragment Length of Circulating Tumor DNA. 1–24 (2016). doi:10.1371/journal.pgen.1006162

42.   Marrone, M., Potosky, A. L., Penson, D. & Freedman, A. N. A 22 gene-expression assay,

decipher® (GenomeDx biosciences) to predict five-year risk of metastatic prostate cancer in men treated with radical prostatectomy. *PLoS Curr.* **7**, 1–8 (2015).

43.   Cooperberg, M. R., Hilton, J. F. & Carroll, P. R. The CAPRA-S score: a straightforward tool for improved prediction of outcomes after radical prostatectomy. *Cancer* **117**, 5039–5046 (2011).

44.   Cooperberg, M. R. *et al.* Combined value of validated clinical and genomic risk stratification tools for predicting prostate cancer mortality in a high-risk prostatectomy cohort. *Eur. Urol.* **67**, 326–333 (2015).

45.   Liu, H. *et al.* Prognostic significance of six clinicopathological features for biochemical recurrence after radical prostatectomy: A systematic review and meta-analysis. *Oncotarget* **9**, 32238–32249 (2018).

46.   Wyatt, A. W. *et al.* Concordance of Circulating Tumor DNA and Matched Metastatic Tissue Biopsy in Prostate Cancer. doi:10.1093/jnci/djx118

47.   Lapin, M. *et al.* Fragment size and level of cell-free DNA provide prognostic information in patients with advanced pancreatic cancer. *J. Transl. Med.* **16**, 300 (2018).

48.   Meddeb, R., Pisareva, E. & Thierry, A. R. Guidelines for the preanalytical conditions for analyzing circulating cell-free DNA. *Clin. Chem.* **65**, 623–633 (2019).

49.   Lee, T.-H., Montalvo, L., Chrebtow, V. & Busch, M. P. Quantitation of genomic DNA in plasma and serum samples: higher concentrations of genomic DNA found in serum than in plasma. *Transfusion* **41**, 276–282 (2001).

50.   Jung, M., Klotzek, S., Lewandowski, M., Fleischhacker, M. & Jung, K. Changes in concentration of DNA in serum and plasma during storage of blood samples [5]. *Clin. Chem.* **49**, 1028–1029 (2003).

51. Devonshire, A. S. *et al.* Towards standardisation of cell-free DNA measurement in plasma: Controls for extraction efficiency, fragment size bias and quantification. *Anal. Bioanal. Chem.* **406**, 6499–6512 (2014).

52. Prakash, K., Aggarwal, S., Bhardwaj, S., Ramakrishna, G. & Pandey, C. K. Serial perioperative cell-free DNA levels in donors and recipients undergoing living donor liver transplantation. *Acta Anaesthesiol. Scand.* **61**, 1084–1094 (2017).

53. Sozzi, G. *et al.* Quantification of free circulating DNA as a diagnostic marker in lung cancer. *J. Clin. Oncol.* **21**, 3902–3908 (2003).

54. Boorjian, S. A. *et al.* Long-Term Outcome After Radical Prostatectomy for Patients With Lymph Node Positive Prostate Cancer in the Prostate Specific Antigen Era. *J. Urol.* **178**, 864–871 (2007).

55. Litwin, M. S. & Tan, H. J. The diagnosis and treatment of prostate cancer: A review. *JAMA - J. Am. Med. Assoc.* **317**, 2532–2542 (2017).

56. Caswell, D. R. & Swanton, C. The role of tumour heterogeneity and clonal cooperativity in metastasis, immune evasion and clinical outcome. *BMC Med.* **15**, 133 (2017).

57. Cooper, C. S. *et al.* Analysis of the genetic phylogeny of multifocal prostate cancer identifies multiple independent clonal expansions in neoplastic and morphologically normal prostate tissue. *Nat. Genet.* **47**, 367–372 (2015).

58. Cancer Genome Atlas Research Network, T. *et al.* The Molecular Taxonomy of Primary Prostate Cancer. *Cancer Genome Atlas Res. Netw. Cell* **163**, 1011–1025 (2015).

59. Lalonde, E. *et al.* Tumour genomic and microenvironmental heterogeneity for integrated prediction of 5-year biochemical recurrence of prostate cancer: a retrospective cohort study. (2014). doi:10.1016/S1470-2045(14)71021-6

60.   Price, T. J. *et al.* Prognostic impact and the relevance of PTEN copy number alterations in patients with advanced colorectal cancer (CRC) receiving bevacizumab. *Cancer Med.* **2**, 277–285 (2013).

61.   Boutros, P. C. *et al.* Spatial genomic heterogeneity within localized, multifocal prostate cancer. *Nat. Genet.* **47**, 1–14 (2015).

62.   Abbosh, C. *et al.* Phylogenetic ctDNA analysis depicts early stage lung cancer evolution. *Nature* **22364**, 1–25 (2017).

63.   Phallen, J. *et al.* Direct detection of early-stage cancers using circulating tumor DNA. **2415**, (2017).

64.   Hennigan, S. T., Trostel, S. Y., Terrigino, N. T., Voznesensky, O. S. & Schaefer, R. J. Low Abundance of Circulating Tumor DNA in Localized Prostate Cancer. (2019). doi:10.1200/PO.19

65.   Kinde, I., Wu, J., Papadopoulos, N., Kinzler, K. W. & Vogelstein, B. Detection and quantification of rare mutations with massively parallel sequencing. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 9530–9535 (2011).

66.   Cario, C. L. *et al.* A machine learning approach to optimizing cell-free DNA sequencing panels: with an application to prostate cancer. *BMC Cancer* 1–9 (2020). doi:10.1101/2020.04.30.069658

67.   Pérez-Barrios, C. *et al.* Comparison of methods for circulating cell-free DNA isolation using blood from cancer patients: Impact on biomarker testing. *Transl. Lung Cancer Res.* **5**, 665–672 (2016).

68.   Benjamin, D. *et al.* Calling Somatic SNVs and Indels with Mutect2. *bioRxiv* 1–8 (2019). doi:10.1101/861054

69. Ramos, A. H. *et al.* Oncotator: Cancer variant annotation tool. *Hum. Mutat.* **36**, E2423–E2429 (2015).

70. Adalsteinsson, V. A. *et al.* Abstract LB-136: High concordance of whole-exome sequencing of cell-free DNA and matched biopsies enables genomic discovery in metastatic cancer. *Cancer Res.* **76**, (2016).

71. Lang, S. H. *et al.* A systematic review of the prevalence of DNA damage response gene mutations in prostate cancer. *Int. J. Oncol.* **55**, 597–616 (2019).

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

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

74. Ashouri, A., Sayin, V., Van den Eynden, J., Singh, S. & Larsson, E. Pan-cancer transcriptomic analysis associates long non- coding RNAs with key mutational driver events. *Nat. Commun.* **7**, (2016).

75. Lanzós, A., Carlevaro-Fita, J., Mularoni, L. & Johnson, R. Discovery of Cancer Driver Long Noncoding RNAs across 1112 Tumour Genomes: New Candidates and Distinguishing Features. *Nat. Sci. Reports* **7**, (2017).

76. Fredriksson, N. J., Ny, L., Nilsson, J. A. & Larsson, E. Systematic analysis of noncoding somatic mutations and gene expression alterations across 14 tumor types. *Nat. Genet.* **46**, 1258–1263 (2014).

77. He, F. *et al.* Integrative Analysis of Somatic Mutations in Non-coding Regions Altering RNA Secondary Structures in Cancer Genomes. *Sci. Rep.* **9**, 1–12 (2019).

78. Abbosh, C., Swanton, C. & Birkbak, N. J. Clonal haematopoiesis: A source of biological

noise in cell-free DNA analyses. *Annals of Oncology* **30**, 358–359 (2019).

79.   Liu, J. *et al.* Biological background of the genomic variations of cf-DNA in healthy

individuals. *Ann. Oncol.* **30**, 464–470 (2019).

80.   Melijah, S. *et al.* The Evolutionary Landscape of Localized Prostate Cancers Drives

Clinical Aggression. *Cell* **173**, (2018).

Funding and Support

Publishing Agreement

It is the policy of the University to encourage open access and broad distribution of all theses, dissertations, and manuscripts. The Graduate Division will facilitate the distribution of UCSF theses, dissertations, and manuscripts to the UCSF Library for open access and distribution. UCSF will make such theses, dissertations, and manuscripts accessible to the public and will take reasonable steps to preserve these works in perpetuity.

I hereby grant the non-exclusive, perpetual right to The Regents of the University of California to reproduce, publicly display, distribute, preserve, and publish copies of my thesis, dissertation, or manuscript in any form or media, now existing or later derived, including access online for teaching, research, and public service purposes.

Author Signature

12/11/2020

Date