Precision Medicine Approaches to Hormone-Driven Cancer

by

Andi Cani

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Molecular and Cellular Pathology)
in The University of Michigan
2020

Doctoral Committee:

Professor Arul M. Chinnaiyan, Co-Chair
Adjunct Associate Professor Scott A. Tomlins, Co-Chair
Professor Gary D. Hammer
Professor Daniel F. Hayes
Associate Professor Zaneta Nikolovska-Coleska
Assistant Professor Rajesh C. Rao

Andi Cani

acani@med.umich.edu

ORCID iD: 0000-0003-4691-656X

DEDICATION

This doctoral dissertation is dedicated to my parents Zhaneta and Kleomeni Cani and sister Xheni Cani, for their unwavering love and support.

ACKNOWLEDGMENTS

The work in this dissertation could not have been completed without the contributions and support of the Tomlins group throughout the years, namely: Scott Tomlins, Chia-Jen “Albert” Liu, Kevin Hu, Lorena Lazo de la Vega, Daniel Hovelson, Sumin Han, Komal Kunder, Moloy Goswamy, Lei “Lucy” Wang, Kelly VanDenBerg, Kei Omata and Mary Greene; contributions and support from the Chinnaiyan group: Arul Chinnaiyan, Javed Siddiqui, Xuhong Cao, Sunita Shankar, Chandan Kumar-Sinha, Xiaoju “George” Wang, Mohan Dhanasekaran, Daniel Robinson, Yi-Mi Wu, Paul Harms, Alexey Nesvizhskii, Abhijit Parolia, Rahul Manan, Steve Kregel, Ron Siebenaler, Palak Shah, Seema Chugh, Yajia Zhang, Lanbo Xiao, Heng Zheng, Fengyun Su, Rui Wang, Vijaiya Lakshmi Dommeti, Phyllis Schooler, Melissa Dunn, Ingrid Apel, Janice Griggs, Alga Abraham, Grace Huang, Jeanmarie Mishler, Lisa McMurry, Mandy Trierweiler, Christine Betts, Nikki Cortis, Sisi Gao; contributions and support from the Salami group: Simpa Salami, Trinh Pham; contributions and support from Aaron Udager, John Wei, Rajesh Rao, Todd Morgan, Ganesh Palapattu and Alexander Zaslavsky; contributions and support from the Hayes group: Dan Hayes, Costanza Paoletti, Emily Dolce, Elizabeth Darga, Marty Brown; contributions and support from the Rae group: Jimmy Rae, Christina Gersch, Jose Larios. The MCP graduate program and the Pathology Department have been a second home and its students, faculty and staff have been akin to

family, including (in addition to already mentioned above): Zaneta Nikolovska-Coleska, David Lombard, Gregory Dressler, Nicholas Lukacs, Andrew Muntean, Andrew Lieberman, Sriram Venetti, David Ferguson, Celina Kleer, Elizabeth Lawlor, Kathleen Cho, Eric Fearon, Charles Parkos, Jean-Francois Rual, Russell Ryan, Thomas Wilson, Laura Labutt, Carrie-Anne Malinczak, Angela Guo, Xiaofang Shi, Sabra Djomehri, Allison Johnson, David Rogawski, Mary Morgan, Siva Natarajan, Samantha Saylor, Anna Ting, Sierrah Grigsby, Carl Engelke, James Ropa, Talha Anwar, Justin Serio, Ashwin Iyer, Derek Dang, David Hu, Thaddeus Kunkel, Sahiti Marella, Michael Pitter, Jessica McAnulty, Alec Monovich. Thesis committee members already mentioned above and Gary Hammer have been instrumental in steering this work and I am humbled to have had such guidance. Administrative assistance from their administrative support staff (in addition to already mentioned above): Chelsea Long, Bridgette Robinson, Wendy Mrdjenovich, Lisa Byrd.

This work would not have been possible without our labs' funding sources acknowledged elsewhere as well as my funding: UM Rackham Merit Fellowship, MCP Training Program in Translational Research NIH T32 grant, the UM Pathology Department and the UM Precision Health Scholar Award.

TABLE OF CONTENTS

LIST OF TABLES

TABLE PAGE

LIST OF FIGURES

FIGURE PAGE

LIST OF APPENDICES

APPENDIX PAGE

ABSTRACT

Cancers of the breast and prostate are primarily driven by sex hormone signaling which has been targeted clinically with considerable success. The genomic and transcriptomic landscape of these tumors has been thoroughly elucidated and major advances have been made in characterization from non-invasive "liquid biopsy", however, the full potential of these advances is yet to be realized in the clinic. Two areas of our focus, with such opportunities to improve patient outcomes in hormone-driven cancers are: 1) Non-invasive tracking of metastatic breast cancer (mBC) evolution for precision medicine; 2) Non-invasive early detection of prostate cancer (PCa).

mBC is an ultimately treatment-resistant, lethal disease characterized by intra-patient molecular heterogeneity including genomic alterations driving resistance to therapy directed at estrogen receptor (ER; *ESR1*) signaling (endocrine therapy). Tumor biopsies are not routinely available in this setting. We sought to determine whether circulating tumor cells (CTCs) recapitulate the genomic landscape of bulk tumor tissue. In project 1, we isolated and genomically profiled individual CTCs from 12 patients with mBC who had concurrent whole exome sequencing of their metastatic biopsy bulk tissue. In 76 individual and pooled informative CTCs, we observed 85% concordance in at least one driver somatic mutation/copy number alteration (CNA) between CTCs and matched tissue metastases, with CTC profiling identifying

diverse intra- and inter-patient molecular mechanisms. For example, in one patient, we observed CTCs that were either wild type for *ESR1* (n = 5/32), harbored the known activating *ESR1* p.Y537S mutation (n = 26/32) also present in tissue, or harbored a novel *ESR1* p.A569S mutation not observed in tissue (n = 1/32), which was demonstrated to be modestly activating in vitro. Our results demonstrate the feasibility and potential clinical utility of comprehensive profiling of archived fixed CTCs.

Despite advances in biomarker development, early detection of aggressive PCa remains challenging. We previously developed a clinical-grade laboratory-developed test—MiProstate Score (MiPS)—for individualized aggressive PCa risk prediction. MiPS combines serum PSA with transcription-mediated amplification (TMA)-quantified expression of the gene-fusion *TMPRSS2:ERG* (*T2:ERG*) and the lncRNA *PCA3* from whole urine obtained after a digital rectal exam (DRE). To improve upon MiPS, in our second project, we describe the pre-clinical development and validation of a post-DRE whole urine targeted RNA NGS assay (NGS-MiPS) assessing ~90 PCa candidate transcriptomic biomarkers, including: *T2:ERG.T1E4* and *PCA3*, additional common PCa gene fusion isoforms, mRNAs, lncRNAs, and expressed mutations. NGS-MiPS showed high analytic validity and was able to detect expressed germline risk *HOXB13* and somatic driver *SPOP* mutations. In an extreme design cohort (benign or Grade Group (GG) 1 vs. GG 3-5 cancer on biopsy) NGS-MiPS showed expected differences in the levels of *T2:ERG.T1E4* and *PCA3*, as well as additional biomarkers, between benign/GG 1 vs. GG 3-5 PCa. A machine learning approach trained on a subset of the extreme design cohort (n=73) generated a 15-transcript model that outperformed derived MiPS and serum PSA models in predicting biopsy outcome in two validation cohorts: 1. A held-out set from the extreme design cohort (n=36); and 2. A separate PCa active surveillance cohort (n=45). These results

support the potential utility and continued development of our novel urine-based targeted RNA NGS assay to improve aggressive PCa early detection.

Leveraging recent technological advances—CTC isolation, NGS, liquid biopsy techniques—as well as knowledge of the genomic landscape of hormone-driven cancers, we demonstrate feasibility of non-invasive precision medicine in mBC and PCa with potential clinical utility.

CHAPTER I

Introduction

1.1 Cancer statistics

Cancer is a relatively common disease, responsible for an estimated 1.75 million new cases in the US in 2019, with over 58,000 cases in the state of Michigan alone[1]. Cancer remains among the leading causes of death with mortality due to all cancer types estimated at over 440,000 patients in 2019, second only to cardiovascular disease[2]. In addition, national cost of direct medical care and loss of economic productivity due to cancer are projected to cost the US economy $158 and $150 billion, respectively, in 2020[3-5]. Importantly, cancer death rates have been on a sustained decrease over the past few decades (~20% in the past 20 years) likely due, at least in part, to advances in early detection, treatment and monitoring[6]. This supports continued efforts on many fronts, ranging from better understanding basic the mechanisms of disease to translating those findings into improved clinical outcomes.

1.2 Breast and prostate normal and tumor tissue growth is largely driven by sex hormone signaling

Cancers of the breast and prostate are among the most commonly diagnosed malignancies responsible for nearly one third of new cancer cases and 17% of cancer deaths in the US[1, 2]. Many molecular similarities exist in the normal function and tumors between these two

secondary sex organs. Both are largely controlled by the signaling activity of sex hormones: estrogens (mainly estradiol, estrone and estriol) and progesterone in breast, and androgens (mainly testosterone, dihydrotestosterone and androstenedione) in the prostate. These steroid hormone family members are derivatives of cholesterol (as are aldosterone and glucocorticoids), and estrogens are derived directly from enzymatic conversion of androgens by the enzyme aromatase (**Figure 1**). Epithelial cells of the breast and prostate are exposed to these growth inducing hormones over decades[7, 8].

Fat-soluble estrogens and androgens, produced primarily in ovaries and testes, respectively, (with more minor contributions from the adrenal cortex) diffuse through the cell membrane of target cells to bind to their respective intracellular receptors: the estrogen and progesterone receptors (ER [*ESR1*], PR) in the breast, and androgen receptors (*AR*) in the prostate. Interaction of these hormones with their receptors promotes transport into the nucleus, where the hormone-bound receptors dimerize (some extranuclear activity of ER has been reported but its function is less well understood)[9]. Receptor homodimers bind to conserved specific DNA sequences: estrogen response elements (EREs) and androgen response elements (AREs)[10, 11]. Upon DNA binding in combination with other co-activators (or co-repressors), they serve as transcription factors to induce expression of downstream genes and implement a transcriptional program that drives cellular growth, among other effects[12, 13]. In cancer, this same signaling promotes tumor cell growth, a notion supported by the fact that therapies targeted at inhibiting ER and AR signaling have effective antitumor activity as described below[14]. It should be mentioned that not all breast and prostate cancers share dependence on sex hormone molecular signaling. The membrane receptor tyrosine kinase Erbb2/Her-2 (not ER/PR signaling) is responsible for driving a subset of hormone receptor-negative breast cancers. Similarly, a

subset of prostate cancers of small cell/neuroendocrine character are independent of AR signaling.

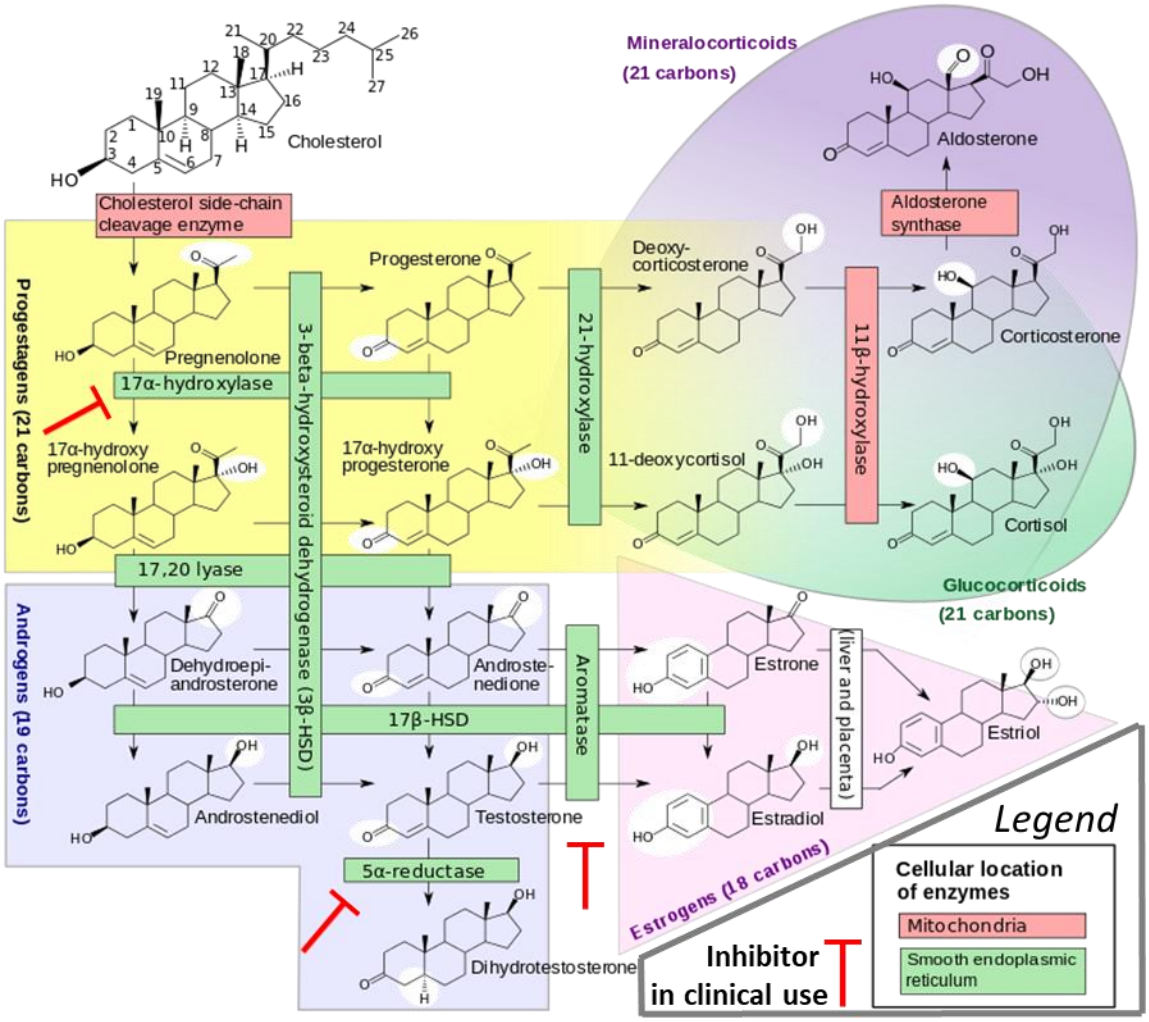


Figure 1. Steroid hormone biosynthesis and pharmacological inhibition. Cholesterol molecules undergo several oxidation/reduction steps via hydroxylation, carbonylation and hydrogenation catalyzed by specific localized enzymes to yield mineralocorticoids, glucocorticoids and sex hormones (modified from Haggstrom et al. [15]).

1.3 Pharmacological endocrine therapy for breast and prostate cancers

As mentioned above, treatment approaches that inhibit hormone signaling in a targeted manner have led to marked improvements in outcomes of hormone driven cancers in both organs. Specifically, endocrine therapy for ER/PR+ breast cancer (with or without chemotherapy

and radiation) is a mainstay treatment strategy used in the adjuvant (i.e. post-surgical resection of a primary breast tumor in order to eradicate any local or distant residual cancer cells and prevent cancer recurrence) and metastatic settings[16-19]. Several drug classes have emerged that inhibit estrogen receptor signaling. Selective estrogen receptor modulators (SERMs), with tamoxifen being a canonical example, have both pro- and anti-estrogenic function depending on tissue type, with anti-estrogenic activity predominating in the breast, although instances of tamoxifen ER agonist activity in breast cancer are known[20]. Tamoxifen is among the most widely-prescribed treatments for hormone sensitive breast cancer and it has been shown to act by binding to ER and inhibiting its transcription factor activity[21]. Selective estrogen receptor degraders (SERDs), exemplified by fulvestrant, on the other hand, are pure anti-estrogens that lack off- (or on-) site pro-estrogenic effects, and act by binding to ER and preventing its dimerization leading to eventual degradation[22]. The anti-cancer activity of fulvestrant is comparable to tamoxifen, however it requires intra-muscular injection of a large-volume formulation which has led to the search for oral SERDs[23, 24]. Other endocrine therapies include gonadotropin releasing hormone (GnRH) agonists such as leuprolide and buserelin, which act by triggering a negative feedback loop on hypothalamic production of GnRH leading to decreased pituitary gonadotropin hormones: follicle stimulating hormone (FSH) and luteinizing hormone (LH), whose normal function is to induce sex hormone production. This decrease, in turn, leads to a shutdown of estrogen production in the ovaries and adrenal cortex[25]. Lastly, aromatase inhibitors (AIs) such as anastrozole and exemestane, are another class of anti ER+ breast cancer drugs that act by inhibiting a key enzyme, aromatase, which converts testosterone to estrogen in extra-ovarian tissues such as the adrenal cortex (**Figure 1**). As such, it is used effectively as an endocrine therapy drug only in post-menopausal women.

Similarities between breast and prostate normal and cancerous tissues extend to the drugs used in treating their respective cancers as well. Thus, anti-androgen therapy, also called androgen deprivation therapy (ADT), is a mainstay in hormone sensitive primary prostate cancer in the adjuvant and especially metastatic setting, alone or in combination with chemotherapy and/or radiation[26]. GnRH agonists, including some of the same compounds mentioned above in breast cancer, such as leuprolide and buserelin, are commonly used ADT drugs for so-called chemical castration (as opposed to orchiectomy - surgical removal of the testicles). Another prostate cancer ADT drug class, GnRH antagonists, are opposite in their immediate function but similar in their ultimate effect to GnRH agonists. Agonists cause an initial increase in gonadotropins such as LH and consequently androgen hormones. This is followed by a sharp decrease due to eventual onset of feedback signaling as described above in breast cancer. GnRH antagonists, on the other hand, directly bind to GnRH receptors on pituitary cells and competitively antagonize binding of the native GnRH hormone, thus having a more immediate effect[27, 28]. GnRH antagonist drug examples include degarelix.

Similarities in breast and prostate cancer pharmacological treatment extend to direct inhibitors of the sex hormone nuclear receptors as well. Thus prostate cancer counterparts to breast cancer SERMs and SERDs are direct AR inhibitors which are used in combination with, or after suppressors of pituitary hormones above (GnRH agonists/antagonists). Part of the benefit of their concomitant use with GnRH agonists at the start of therapy is thought to lie in mitigating the effects of the initial androgen boost by GnRH agonists before negative feedback inhibition sets in[29]. These drugs include the older generation bicalutamide and its other class members[30] as well as the newer drugs, enzalutamide and apalutamide[31]. The latter two have been effective in the clinic for prostate cancer that becomes eventually resistant to older

generation ADT (metastatic castration resistant prostate cancer, mCRPC), an almost universal occurrence[32]. Recently, enzalutamide and the same-class drug apalutamide were shown to also provide benefit in combination with ADT in patients with metastatic castration-sensitive prostate cancer (mCSPC), in progression-free and importantly, in overall survival as well[33, 34]. Finally, inhibitors of steroidogenic enzymes act in a similar way to breast cancer aromatase inhibitors above to block biosynthesis of testosterone and its derivatives in prostate cancer (**Figure 1**). Specifically, dutasteride, finasteride and other class members are inhibitors of 5α-reductase, the enzyme that catalyzes the reduction of testosterone to the much more potent androgen dihydrotestosterone (DHT). They are more commonly used to treat benign prostatic hyperplasia (BPH). A newer drug, abiraterone, inhibits the 17α-hydroxylase enzyme (Cytochrome P450 17A1, CYP17A1), an upstream enzyme in steroidogenesis pathway and has been used successfully in mCRPC[35]. Similarly to enzalutamide and apalutamide, it was recently shown to be improve overall survival in combination with ADT in mCSPC patients as well[36, 37].

Thus overall pharmacological treatment strategy in hormone receptor-positive breast and prostate cancers shows substantial similarities reflecting mirroring molecular biology and endocrinology characteristics of these two organs and by extension, their tumors.

1.4 The molecular landscape of breast and prostate cancer endocrine therapy resistance

The somatic molecular landscapes of breast and prostate primary tumors have been thoroughly elucidated as part of The Cancer Genome Atlas (TCGA) and other efforts[38-42]. Furthermore, large cohorts of metastatic samples of these two tumor types have also been

sequenced to determine recurrent alterations, thought to underlie therapy resistance mechanisms[43-47]. Several patterns have emerged from these studies. Although, as described above, the sex hormone nuclear receptors are key drivers of disease biology of these cancers, recurrent genomic alterations in ER/PR or AR and their fellow signaling pathway protein members in primary tumors are notably absent in both tumor types. In fact, it is not until these cancers become resistant to endocrine therapy or ADT (usually in the metastatic setting) that ER and AR mutations or gene amplifications emerge in dominant tumor cell subclonal populations (whether pre-existing or acquired *de novo*). Evolution of these alterations is thought to happen due to selective evolutionary pressure from endocrine therapy. Specifically, advanced ER+ breast cancers with acquired resistance to SERMs/SERDs and aromatase inhibitors often have mutations in *ESR1*, the gene encoding estrogen receptor alpha (ERα)[48]. These missense amino acid substitutions are concentrated in a hotspot region of the *ESR1* ligand binding domain (LBD) and have been shown to constitutively activate ER[48]. This activation obviates the need for estrogen ligand binding and thus defeats the effects of anti-estrogen endocrine therapy. This cluster of *ESR1* LBD hotspot amino acids was shown by Robinson et al. to include p.Y537, p.D538 and less commonly, p.L536[48]. In some patients, these initially heterozygous mutant cell subclones undergo deletion of the wild type *ESR1* allele (loss of heterozygosity, LOH, with copy loss or copy-neutral) which results in all-mutant ER homodimers. These are hypothesized to be superior transcriptional activators over mutant:wild-type dimers by being entirely free from the inhibitory effects of wild type ER inhibitors[49]. Furthermore, *ESR1* copy number amplifications have also been observed in treated advanced breast cancers making this another apparent endocrine therapy resistance mechanism. Gene amplification presumably causes *ESR1* overexpression due to a greater number of gene copies being transcribed and translated to

generate additional ERα[43].

As may be expected from the discussion so far, the molecular landscape of prostate cancer resistant to androgen deprivation therapy largely parallels that in endocrine therapy-resistant ER+ breast cancer. Thus, the selective pressure that earlier generation ADT drugs of diverse mechanisms exert on AR generates dominant subclones (whether pre-existing or *de novo*) containing genomic alterations of this nuclear receptor gene[50]. Indeed, *AR* gene amplifications are present in ~30% of mCRPC patients and these alterations are not observed in the patient-matched primary tumors, at least not at the sensitivity level of the sequencing methods used[51]. A greater number of *AR* gene copies presumably increases its expression which can overcome AR inhibition in an equivalent manner to *ESR1* gene amplifications in endocrine therapy resistant breast cancer. Interestingly, *AR* gene amplification is the dominant endocrine therapy resistance mechanism in prostate cancer as opposed to activating mutations. This is in contrast with endocrine therapy resistant breast cancer where LBD mutations predominate over gene amplifications. However, mutations in AR are present in approximately 15% of mCRPC patients and, as expected, are found in the AR LBD domain. They are primarily centered on 3 hotspots somewhat distant from each other: p.L702, p.W742 and p.H875/p.T878[50]. As in breast cancer, they cause constitutive activation of the receptor which enables ligand-independent transcriptional activity. Interestingly, in some cell lines/xenografts, mutant AR at positions p.H874 and p.F876 exhibit activation, as opposed to inhibition, by direct AR inhibitors flutamide and enzalutamide, respectively. This indicates a potential resistance strategy by prostate cancer cells by turning an AR antagonist into an agonist for their benefit[52, 53]. In line with this observation, ER+ breast cancer endocrine therapy drugs such as SERMs mentioned above, (namely tamoxifen), are also well-known to paradoxically convert from ER

antagonists to agonists. Tumor regression in these patients upon tamoxifen withdrawal is sometimes observed[54-56]. In these tumors, a molecular "reversal" is observed whereby tamoxifen-bound ER drives robust paradoxical ER signaling whereas the estrogen-bound receptor may impede it. This mechanism is supported by observations in *in vitro* models of such tamoxifen-resistant tumors where tamoxifen promotes tumor cell growth, whereas subsequent estrogen treatment has the opposite effect[55].

Another interesting molecular commonality between subsets of advanced hormone-positive breast and prostate cancers are gene rearrangements of their hormone receptors. Specifically, *ESR1* gene fusions in advanced endocrine therapy-resistant breast cancer have been shown to induce overexpression and estrogen independent activity in vitro[57-60]. Likewise, genomic rearrangements at the *AR* locus have also been observed in patients and, as may be expected, these are found in the context of mCRPC disease[61, 62].

Additional AR-related ADT resistance mechanisms have been described in prostate cancer that have less well-documented (yet not inexistent) equivalents in endocrine resistant breast cancer. These include AR splice variants such as AR-V7 which lacks an LBD (although there are reports of breast cancer ER variants in patient circulating cell-free DNA[63]) and upregulation of glucocorticoids (CG) that can bind to mutant AR or drive AR-controlled gene expression via GC receptor signaling[50].

1.5 Other genomic alterations in advanced breast and prostate cancers show similarities

Sustained anti-androgen therapies with newer drugs that overcome initial prostate cancer resistance to prior generation ADT treatments have been shown in some cases to eventually lead

mCRPC into androgen-independent small-cell/neuroendocrine differentiation[50, 64, 65]. This has been hypothesized to be as a result of either selection from pre-existing ADT-refractory cells or from *de novo* transdifferentiated subclones derived from earlier AR+ cells. Recent evidence seems to support the latter model since *TMPRSS2:ERG* fusions (AR-driven) are present in the genomes of these cells at the same frequency as in primary disease (~50%)[64]. By the time of neuroendocrine disease development, these fusions are likely genomic records of a previously active fusion protein, now a non-functional, passenger event. This is supported by the fact that this stage of disease molecular evolution is confirmed not only by loss of ADT effectiveness, but marked downregulation of androgen signaling in these tumors (which is required for *TMPRSS2* gene expression). This latter point is in contrast to earlier stages where the tumor becomes resistant to traditional ADT (mCRPC), yet maintains robust or even elevated AR function which can be inhibited by next generation drugs such as enzalutamide or apalutamide. Other alterations such as deletion of the tumor suppressors *TP53* and *RB1* (these two typically occurring in neuroendocrine tumors), PI3 kinase negative regulator *PTEN* as well as amplification of the *MYCN* transcriptional activator and deleterious alterations in DNA damage repair genes (*BRCA1*, *BRCA2* and *ATM*) appear to drive tumor growth at this later stage[45, 47, 64, 65].

Some of the same genes are altered in advanced treated breast cancers although similarities are not as striking as for the hormone receptor altered cancers under endocrine therapy[43, 66]. Thus alterations specific to metastatic disease in breast cancer include *TP53*, *RB1* and *PTEN* but also other genes not recurrently altered in prostate cancer. Among these are HER-2 (*ERBB2*) mutations and amplifications, usually as primary Her-2+ breast cancers developing resistance under anti-Her-2 therapy. However, in vitro studies also show ER+ breast cancer cell lines upregulating Her-2 under sustained tamoxifen treatment, thus supporting

evolution from endocrine therapy-treated ER+ breast cancer into more hormone-independent Her-2+ disease as a potential endocrine therapy resistance mechanism[54]. Other genomic alterations in advanced breast cancer include deleterious changes in *NF1* (a GTPase activating protein, GAP), which is a negative regulator of the MAP kinase pathway that normally induces RAS GTPase activity to self-catalyze hydrolysis of Ras-bound GTP into GDP thus inactivating it[66]. Other genes in growth factor receptor tyrosine kinase (e.g. EGFR) – MAP kinase (e.g. KRAS) axis are also recurrently mutated in advanced breast cancers[66].

1.6 The genomic landscapes of primary breast and prostate cancers show fewer similarities

Despite notable parallels in sex hormone receptor signaling driving a considerable portion of pro-cancer transcriptional programing in hormone receptor-positive breast and prostate cancers, the greatest dissimilarity between the two cancer types lies in the genomic alterations driving their primary tumors (which have not been exposed to selective pressures from various anti-cancer therapies). While sex hormone receptor driven transcription is essential in most breast and prostate tumors, as mentioned above, no genomic alterations appear to be involved in this effect in early untreated primary tumors. Studies have shown differing sets of genomic alterations responsible for early cancer development in breast and prostate[38, 39, 42]. Thus hotspot mutations in primary breast cancers center around *TP53*, the PI3 kinase active subunit gene *PIK3CA*, MAP kinase genes, the transcription factor *GATA3* etc. An exception are breast cancers of lobular subtype, which have highly recurrent inactivating mutations (with or without LOH events) in *CDH1*, the gene encoding E-cadherin[38, 42].

By contrast, early genomic events in most primary prostate cancers include highly-

recurrent gene fusions between AR-responsive genes such as *TMPRSS2* or *SLC45A3* and oncogenic Ets transcriptions factors *ERG, ETV1, -4, -5, FLI1* etc. (~65% of cases) with the *TMPRSS2:ERG* fusion being the most common (~50% of all prostate tumors)[39, 67, 68]. In these genomic rearrangements, the 5' AR-controlled fusion partners serve to drive expression of the fusion transcript whereas the 3' Ets genes are the oncogenic end of the fusion (inducing an invasion-associated transcriptional program in the case of *ERG* [69]). The second most common genomically altered gene in primary prostate cancer is the E3 ubiquitin ligase *SPOP*. These account for an additional 5-15% of cases compared to fusion-positive cancers since the presence of hotspot activating mutations in this gene is mutually exclusive with Ets gene fusions[39]. Various less recurrent alterations and other yet unknown molecular mechanisms account for the rest of primary prostate cancer molecular driver alterations.

An interesting subset of prostate cancers include genomic alterations in *FOXA1,* a pioneer transcription factor that precedes binding of nuclear receptors such as AR (and ER) to their genomic response elements[70]. In prostate cancer these alterations fall under distinct subclasses based on the protein domains they affect[71] and like *SPOP* mutations, they are also mutually exclusive with Ets fusions and to some extent with *SPOP* mutations as well[39]. One subclass of *FOXA1* alterations (Class-1 mutations), located in the DNA binding forkhead domain, occurs specifically in early prostate cancer as opposed to mutations in the C-terminal domain which appear to arise in metastatic disease[71]. Interestingly, *FOXA1* mutations are also recurrent in primary breast cancer, particularly of the lobular subtype[38, 42]. Intriguingly these are also limited to the same forkhead domain[42] as Class-1 mutations in early prostate cancer described above[71]. This highlights some rather fundamental similarities in genomic alterations of early breast and prostate cancers, at least in a subset of patients, despite other differences in

their genomic makeup.

1.7 Biomarker potential of tumor molecular alterations for advanced breast cancer precision medicine and prostate cancer early detection

As discussed above, the genomic and transcriptomic alterations emerging at various stages of breast and prostate malignant disease are rich in information. They involve many recurrent DNA point mutations, short insertions/deletions (indels), copy number changes or chromosomal rearrangements/gene fusions. These alterations at the genetic level, as well as other non-genetic modifications, lead to transcriptional changes reflected at the expression level or amino acid sequence of the proteins encoded. This in turn switches the molecular signatures and mechanisms inside the cell from normal to tumorigenic. Importantly, ongoing efforts to make use of this wealth of genomic and transcriptomic information promises to impact clinical care in the form of tumor biomarkers for early disease detection, prognostication, monitoring, evolution and therapy.

The vast majority of the genomic and transcriptomic profiling above has been performed on bulk patient tumor tissue obtained from biopsies or surgical resections. This relatively recently elucidated wealth of candidate molecular biomarkers with potential clinical utility can now be coupled with major advances in “liquid biopsy” techniques. These center around extracting useful information, usually in the form of biomarkers, by sampling relatively (or entirely) non-invasively, various biofluids such as blood, urine, cerebrospinal fluid (CSF), feces, saliva etc. This rests on the principle, backed by abundant observations, that tumors passively or actively release some of their contents into the surrounding extra-tumoral space, which eventually finds its way in the circulation/stream of these biofluids[72]. These materials include

circulating cell-free DNA and RNA (naked, protein-bound or as part of extracellular membrane-bounded vesicles such as exosomes) which can be released by necrotic (dead) or live tumor tissue. They also include intact, usually viable individual tumor cells or cell-clusters which are more commonly thought of in the context of metastatic seeding[72]. These cells obviously contain the entire genomic and transcriptomic makeup of a tumor cell, although they can be limited to the tumor subclonal population they are derived from, which may be important in cases characterized by intra-tumoral molecular heterogeneity. Despite this progress in assaying and analyzing cancer molecular features in individual patients, the full potential of these advances is only beginning to be translated in the clinic.

Two areas of our focus, with such opportunities to improve patient outcomes in hormone-driven cancers are: 1) non-invasive tracking of metastatic breast cancer (mBC) evolution for precision medicine approaches; and 2) non-invasive detection of prostate cancer (PCa) in the early setting.

mBC is an ultimately treatment-resistant, lethal disease characterized by intra-patient molecular heterogeneity including genomic alterations driving resistance to endocrine therapy in hormone receptor-positive tumors, as described earlier. Multiple temporally-distributed tumor biopsies are not readily available in this setting as part of routine clinical care due to cost, complications, metastasis anatomical location etc. This fact limits investigation of the tumor molecular landscape for determining precision medicine strategies (individualized treatment options based on tumors' individual molecular profiles) and tracking of potential molecular resistance mechanisms in real time. A considerable portion of women with mBC have tumor cells circulating in their bloodstream (circulating tumor cells, CTCs) and CTC counts are prognostic markers for disease outcome[73]. CTCs' potential to track the tumor genomic

landscape however, has not been fully utilized. Specifically, it is not clear to what extent the genomic profiles of individual CTCs recapitulate the genomic landscape of bulk tumor tissue. In project 1, we enriched and isolated individual blood circulating tumor cells (CTCs) and performed genomic profiling using targeted next generation sequencing (NGS) of ~135 cancer-related genes in 12 mBC patients who had concurrent whole exome sequencing of their metastatic biopsy bulk tissue. We addressed questions regarding the extent to which tumor tissue and CTC genomic landscapes differ especially with regard to endocrine therapy resistance mechanisms, whether all epithelial cells purified from blood are true tumor cells, and in general the feasibility of CTC comprehensive genomic profiling to contribute to mBC precision medicine. Our results demonstrate the feasibility and potential clinical utility of comprehensive profiling of archived fixed CTCs as a complementary approach to tissue biopsy sequencing for precision medicine in advanced breast cancer patients.

Despite advances in biomarker development, early detection of aggressive prostate cancer remains challenging. Biomarkers currently in use are somewhat limited in number and clinical utility. Serum prostate specific antigen (PSA) has revolutionized prostate cancer early detection largely by improvements in sensitivity. However, despite looking at various serum PSA protein isoforms and combining them with clinical/demographical/medical history parameters, the specificity of cancer detection by this approach remains unsatisfactory. Thus, ~40% of prostate biopsies triggered by a serum PSA level above the clinically used threshold of 4.0 ng/ml at the initial detection setting show no presence of cancer on pathology. Other prostate cancer biomarker tests take advantage of prostate's direct route to the urinary tract by detecting prostate cancer-associated transcriptomic biomarkers in urine RNA. We previously developed one such clinical-grade laboratory-developed urine-based test and predictive algorithm—

MiProstate Score (MiPS)—for individualized prediction of risk of having aggressive prostate cancer on biopsy. MiPS combines serum PSA with transcription-mediated amplification (TMA)-quantified expression of the gene-fusion *TMPRSS2:ERG* (*T2:ERG*; T1E4 splice isoform) and the lncRNA *PCA3* from whole urine obtained after a digital rectal exam (DRE)[74]. To improve upon MiPS, in our second project we describe the pre-clinical development and validation of a post-DRE whole-urine targeted RNA NGS assay (NGS-MiPS) assessing ~90 PCa transcriptomic biomarkers, including: *T2:ERG.T1E4* and *PCA3*, additional common PCa gene fusion isoforms, mRNAs, lncRNAs, and expressed mutations. Our assay showed high technical validity, outperformed currently used clinical models/biomarkers and demonstrated novel features such as detection of expressed mutations. These results support the potential utility and continued development of our novel urine-based targeted RNA NGS assay to supplement serum PSA to help guide initial biopsy and other clinical decisions for improved aggressive prostate cancer early detection.

Our overall approach consists in leveraging recent technological advances (liquid biopsies, CTC isolation, next generation sequencing etc.) as well as knowledge of the genomic landscape of sex hormone-driven cancers to show the potential of non-invasively enabling precision medicine in mBC and accurately detecting early prostate cancer. This work suggests that our approach merits continued development and has potential for clinical utility in order to improve care for breast and prostate cancer patients.

CHAPTER II

Comprehensive mutation and copy number profiling in archived circulating breast cancer tumor cells documents heterogeneous resistance mechanisms[1]

2.1 Abstract

Metastatic breast cancer (mBC) is an ultimately treatment-resistant, lethal disease characterized by intra-patient molecular heterogeneity including genomic alterations driving resistance to therapy directed at estrogen receptor (ER; *ESR1*) signaling (endocrine therapy). Tumor biopsies are not routinely available in this setting. We sought to determine whether circulating tumor cells (CTCs) recapitulate the genomic landscape of bulk tumor tissue. In project 1, we isolated and genomically profiled individual CTCs from 12 patients with mBC who had concurrent whole exome sequencing of their metastatic biopsy bulk tissue. In 76 individual and pooled informative CTCs, we observed 85% concordance in at least one driver somatic mutation/copy number alteration (CNA) between CTCs and matched tissue metastases, with CTC profiling identifying diverse intra- and inter-patient molecular mechanisms. For example, in one patient, we observed CTCs that were either wild type for *ESR1* (n = 5/32), harbored the known activating *ESR1* p.Y537S mutation (n = 26/32) also present in tissue, or harbored a novel *ESR1* p.A569S mutation not observed in tissue (n = 1/32), which was demonstrated to be

[1] This study has previously appeared in Paoletti, C.[#], Cani, A.K.[#], et al. Cancer Research 2018. 49. Paoletti, C., et al., *Comprehensive Mutation and Copy Number Profiling in Archived Circulating Breast Cancer Tumor Cells Documents Heterogeneous Resistance Mechanisms.* Cancer Res, 2018. **78**(4): p. 1110-1122. (# Co-first authors)

modestly activating in vitro. Our results demonstrate the feasibility and potential clinical utility of comprehensive profiling of archived fixed CTCs.

2.2 Introduction

The promise of precision oncology implies that patients with advanced cancer might undergo somatic genomic profiling to identify molecular alterations that match with potentially active molecularly targeted therapies. Most of the strategies of precision medicine to date have used archival or prospectively biopsied tissues for somatic profiling. The use of liquid biopsies, to obtain circulating tumor cells (CTC) or circulating cell free DNA (cfDNA) for phenotypic or molecular analyses, has the potential to overcome tissue availability as a major barrier to precision oncology [75-78]. Such approaches may be particularly valuable in the context of patients progressing after targeted therapy, in whom a single tissue biopsy may be unable to capture the diverse resistance mechanisms driving individual clonal populations of progressing metastases [79]. Importantly, although numerous cfDNA and CTC platforms have been profiled, the only Food and Drug Administration (FDA)-cleared CTC platform is the CellSearch® system (Menarini, Silicon Biosystems), which has regulatory clearance in several cancers for prognostic utility of CTC counts [80, 81]. CTC profiling in treatment-refractory cancers holds particular promise as a translational research tool to capture the global set of resistant cell populations in individual patients, as deconvolution of individual resistant populations in cfDNA requires massive sequencing depth or breadth (when cfDNA tumor content is low) and tissue based profiling only provides information on the biopsied metastasis [48, 82-84].

Nearly all patients with hormone receptor positive (HR+) metastatic breast cancer (MBC) initially respond to anti-estrogen treatments (endocrine therapy, ET), but ultimately progression

is nearly universal. Several ET resistance mechanisms in estrogen receptor (ER)-positive MBC have been identified, including ER down-regulation (through deletion or suppression), alterations in ER-signaling pathway genes, deregulation of growth pathways, down-regulation of apoptosis pathways, and unbalanced ER-co-regulator activity [85]. More recently, we and others have reported mutations in the ligand binding domain (LBD) of *ESR1*, the gene that encodes for ER alpha, in patients with MBC after ET [48, 86-89]. These mutations appear to confer absolute resistance to estrogen depletion, and relative resistance to selective ER modulators, such as tamoxifen, and selective ER down-regulators, such as fulvestrant [48].

We and others have investigated concordance between tissue, cfDNA and CTC-based assessment of *ESR1* mutations in patients with MBC [90-92]. Importantly, the majority of CTC-based genomic profiling approaches in MBC—and other cancers—either rely on platforms that do not fix CTC, utilize low pass whole genome sequencing to identify broad copy number alterations (CNAs) and targeted NGS/Sanger Sequencing approaches that only assess mutations in a very focused set of genes, or rely only on pools of CTC as opposed to single cells [93-97]. In order to interrogate CTC-based ET resistance mechanisms and to determine concordance with tissue biopsy, we performed comprehensive mutation and copy number profiling in 130 genes from archived CTC captured from patients with ET-resistant MBC enrolled on a prospective clinical trial. Importantly, these patients were enrolled on a concurrent clinical trial in which they underwent metastatic tissue biopsy (for whole exome sequencing) enabling comparison of comprehensive CTC and tissue metastasis genomic profiles.

2.3 Materials and Methods

2.3.1 Preclinical evaluation of cultured breast cancer cells spiked into normal human blood

Cultured human breast cancer cell line BT-474 was obtained from the laboratory of Dr. Stephen Ethier and routinely maintained in DMEM +10% FBS at 37°C in 5% CO_2. The identity of the cells was confirmed by standard Short Tandem Repeat profiling (February 2011). BT-474 (*n*=300) were freshly harvested and spiked into pooled de-identified healthy human whole blood (WB) collected in a CellSave Preservative tube (Janssen Diagnostics, LLC). Tumor cell isolation and enumeration was performed using the CellSearch® system with the CellSearch® CXC kit, as previously described [98]. The elapsed time between CTC enrichment and purification varied from 2 years (cartridges were stored at RT), 6 months (cartridge stored at 4°C), and 3 days (cartridge stored at 4°C). This evaluation was also performed to see the effects of long term storage on the quality of the cells. As described below for patient specimens, BT-474 cells were flushed from the CellSearch® cartridges and purified using DEPArray™ technology. Isolated single or groups of CTC were subjected to whole genome amplification (WGA) before sequencing analysis.

2.3.2 Patient population

Twenty-eight eligible patients with MBC, who were enrolled in the MI-ONCOSEQ protocol at the University of Michigan Comprehensive Cancer Center [84], signed a separate informed consent approved by the University of Michigan IRB to be enrolled in a companion protocol designated Mi-CTC-ONCOSEQ that allowed blood collection for cfDNA (data partially reported elsewhere) [90] and CTC analyses. All patients enrolled into this study had WES previously performed on their metastatic research tissue biopsy using the Illumina HiSeq 2500 platform. Thirteen of these 28 patients who had ≥5 CTC/7.5 ml WB by CellSearch® enumeration, and who had at least one DEPArray™-purified CTC with high quality DNA

determined by the Ampli1™ quality control (QC) kit (as described in methods below), were enrolled in the present study.

Blood specimens were collected into fixative-containing 10 ml vacutainer tubes (CellSave) and processed within 96 hours. These specimens were collected from a range of 7 days prior, to 344 days after the research biopsy. For two patients (#4, who had <5 CTC/7.5 ml at time of enrollment and #24, who had ≥5/7.5 ml CTC at time of enrollment), blood for CTC-DNA was also collected at the time of disease progression after 690 days and 465 days from the baseline samples, respectively.

2.3.3 CTC enrichment and enumeration

CTC were enriched from WB using ferrofluid particles coated with EpCAM antibody and enumerated using the CellSearch® System, according to manufacturer's instructions (Janssen Diagnostics, LLC Raritan, NJ, United States), and as previously described [98]. Briefly, cells were stained with anti-cytokeratin (CK) 8, 18, and 19- Fluorescein isothiocyanate (FITC) labeled antibodies, the nuclear dye 4',6'-diamino-2-phenylindole (DAPI), and CD45 antibody labeled with allophycocyanin (APC). CTC were defined by visual inspection as CK positive/DAPI positive/CD45 negative. After processing, CellSearch® cartridges were stored protected from light at 4°C for an average of 322 days (range 37-827 days) before CTC purification (**Table 1**).

Patient #	Timepoint	CTC Enumeration	Unassigned events	Processed by CellSearch	Processed by DEPArray	Time Between Assays (Days)	DEPArray High Quality CTC	DEPArray Low Quality CTC	DEPArray Total CTC	# of single cells sequenced	#of cells sequenced as pool(s)	Sequenced	% of purified CTC	% of high quality CTC/purified CTC	% of high quality CTC/total CTC
CTC-002 A	BL	78	63	1/22/2014	12/18/2014	330	14	4	18	7	6	Yes	23%	78%	18%
CTC-002 B	BL	57	40	1/22/2014	4/28/2016	827	10	12	22	10	12	Yes	39%	45%	18%
CTC-002 C	BL	57	61	1/22/2014	4/28/2016	827	9	13	22	10	10	Yes	39%	41%	16%
CTC-003	BL	6	55	1/23/2014	6/12/2015	505	0	2	2	0	0	No	33%	0%	0%
CTC-007 A	BL	15	1311	2/21/2014	4/15/2015	418	1	2	3	0	3	Yes	20%	33%	7%
CTC-007 B	BL	18	1513	2/20/2014	12/22/2014	305	2	5	7	1		Yes	39%	29%	11%
CTC-008 A	BL	40	379	2/21/2014	3/31/2015	403	2	16	18	0	3	Yes	45%	11%	5%
CTC-008 B	BL	41	361	2/21/2014	12/22/2014	304	0	17	17		0	No	41%	0%	0%
CTC-009	BL	12	68	2/24/2014	4/14/2015	414	0	10	10	1	0	Yes	83%	0%	0%
CTC-010	BL	14	147	2/26/2014	4/14/2015	412	0	16	16	0	0	No	114%	0%	0%
CTC-012	BL	13	31	3/12/2014	4/2/2015	386	1	5	6	1	0	Yes	46%	17%	8%
CTC-014	BL	16	45	4/17/2014	1/5/2015	263	8	1	9	0	3	Yes	56%	89%	50%
CTC-015	BL	5	39	5/2/2014	6/12/2015	406	0	0	0	0	0	No	0%	0%	0%
CTC-017	BL	21	242	7/16/2014	4/1/2015	259	4	16	20	0	4	Yes	95%	20%	19%
CTC-019 A	BL	317	6250	7/16/2014	12/29/2014	166	1	6	7	0	3	Yes	2%	14%	0%
CTC-019 B	BL	344	5404	7/16/2014	3/30/2015	257	2	14	16			Yes	5%	13%	1%
CTC-024	BL	275	71	11/11/2014	12/18/2014	37	14	8	22	0	13	Yes	8%	64%	5%
CTC-025	BL	8	75	7/14/2015	4/5/2016	266	5	8	13	5	5	Yes	163%	38%	63%
CTC-028*	BL	142	329	9/25/2015	4/5/2016	193	3	13	16	2	2	Yes	23%	19%	2%
CTC-029*	BL	81	58	10/9/2015	4/5/2016	179	15	1	16	4	8	Yes	40%	94%	19%
CTC-030*	BL	962	441	10/7/2015	5/23/2016	229	34	0	34	5	4	Yes	7%	100%	4%
CTC-004	P	9	38	12/16/2015	4/5/2016	111	6	7	13	3	3	Yes	144%	46%	67%
CTC-024*	P	126	291	2/19/2016	5/23/2016	94	5	5	10	2	3	Yes	16%	50%	4%
Mean													**47%**	**35%**	**14%**

Table 1. Results of CTC purification by DEPArray. CellSearch®-enriched cells were purified into individual cells by DEPArray™. Patient # and CellSearch® cartridge (by letter) are shown in left-most column. CTC enumeration, processing dates, DNA quality and sequencing status are shown. BL = baseline; P = progression; *For these samples, only 1/2 cartridge was purified by DEPArray, % of purified CTC was adjusted accordingly.

2.3.4 Single cell purification and DNA isolation

Each CellSearch® cartridge was processed to recover single CTC using the DEPArray™ system (Menarini Silicon Biosystems, S.p.A., Bologna, Italy) per the manufacturer's instructions. Briefly, after being flushed from the cartridges, cells were placed into the DEPArray™ 300k chip which separates individual cells based on immunofluorescent staining criteria and cell morphology (CTC: CK-FITC positive, DAPI positive, and CD45-APC negative; WBC: CK-FITC negative, DAPI positive, and CD45-APC positive). After imaging, individual selected cells were routed for isolation and recovery [99]. DNA from individual CTC or WBC was isolated and subjected to WGA using *Ampli*1™ WGA kit (Menarini Silicon Biosystems, S.p.A.) with MseI digestion per the manufacturer's instructions [100]. Subsequent DNA quality control was performed using *Ampli*1™ QC kit [100] and low quality DNA cells were excluded

from down-stream sequencing Because CellSearch® technology enriches, but does not purify CTC from WBC, each CellSearch® cartridge was subsequently processed to recover single and pooled populations of pure CTC using the DEPArray™ system (Di-Electro-Phoretic Array; Menarini Silicon Biosystems, S.p.A., Bologna, Italy) according to the manufacturer's instructions. In detail, after the cells were flushed from the cartridges using the manipulation buffer wash provided by the manufacturer, the cell pellet was re-suspended and placed into the DEPArray™ 300k chip which contains electronically controlled electrodes used to generate the electrophoretic cages. The DEPArray™ single cell sorting system separates individual cells into separate virtual magnetic nests/cages, based on immunofluorescent staining criteria and cell morphology (CTC: CK-FITC positive, DAPI positive, and CD45-APC negative; WBC: CK-FITC negative, DAPI positive, and CD45-APC positive). After imaging of the cell within each virtual magnetic nest/cage generated by DEPArray™, individual cells of interest were routed for isolation and recovery into 0.2 ml tubes [99]. DNA from individual CTC or WBC from the same specimen, was isolated and subjected to whole genome amplification (WGA) using *Ampli*1™ WGA kit (Menarini Silicon Biosystems, S.p.A.) which includes digestion with a restriction enzyme (MseI) according to the manufacturer's instructions [100]. Subsequent DNA quality control was performed using *Ampli*1™ QC kit as described by Polzer *et al* [100]. Only cells with high quality DNA were further investigated for down-stream sequence analysis. Normal control DNA derived from matched WBC was processed and tested for QC following the same process and criteria.

2.3.5 CTC genomic profiling and data analysis

NGS was performed essentially as previously described using 20 ng of WGA DNA for

each CTC/WBC sample for targeted, multiplexed PCR-based NGS (Ampliseq, Ion Torrent)[101, 102]. Selected mutations were confirmed by Sanger Sequencing. Targeted next-generation sequencing (NGS) was performed essentially as previously described [101-103]. Briefly, 20 ng of WGA DNA (quantified with the Qubit Fluorometer 3.0, ThermoFisher, Waltham, MA) for each single, or equally pooled CTC, or WBC DNA was used for targeted, multiplexed PCR-based NGS (Ampliseq, Ion Torrent). Libraries with barcode incorporation were constructed using the DNA component of the Oncomine Comprehensive Assay (OCP), a panel comprised of 2,531 amplicons targeting 130 genes covering 260,717 bases (ThermoFisher, Waltham, MA). Genes were selected based on pan-solid tumor NGS and copy number profiling data analysis that prioritized somatic, recurrently altered oncogenes, tumor suppressor genes, genes present in high level copy gains/losses and known/investigational therapeutic targets [102]. Libraries for spiked BT-474 cells were constructed using the Cancer Hotspot v2 Assay (CHPv2), a panel comprised of 207 amplicons targeting regions containing recurrent alterations in 50 onco- and tumor suppressor genes (ThermoFisher, Waltham, MA). Because the WGA method we used digests the DNA with the MseI restriction enzyme, some amplicons in each panel were adversely affected based on the presence of the MseI restriction site (for OCP see **Table 2;** for CHPv2, as previously described [104]).

Unaffected Genes	*ARAF, BTK, FOXL2, GNA11, IDH2, IFITM1, IFITM3, KNSTRN, MAP2K2, MAPK1, MLH1, MPL, RHOA, STAT3, HNF1A, HRAS, JAK1, MAX, PPP2R1A, SRC, CSF1R, ESR1, MED12, NKX2 -8, NRAS, SMO*
Partially Affected Genes	*ABL1, ACVRL1, AKT1, ALK, APC, APEX1, AR, ATM, ATP11B, BAP1, BCL2L1, BCL9, BIRC2, BIRC3, BRAF, BRCA1, BRCA2, CBL, CCND1, CCNE1, CD274, CD44, CDH1, CDK4, CDK6, CDKN2A, CSNK2A1, CTNNB1, DCUN1D1, DDR2, DNMT3A, EGFR, ERBB2, ERBB3, ERBB4, FBXW7, FGFR1, FGFR2, FGFR3, FGFR4, FLT3, GAS6, GATA2, GATA3, GNAS, IDH1, IGF1R, IL6, JAK3, KDR, KIT, KRAS, MAP2K1, MCL1, MDM2, MDM4, MET, MSH2, MTOR, MYC, MYCL, MYCN, MYD88, MYO18A, NF1, NF2, NKX2 -1, NOTCH1, PDCD1LG2, PDGFRA, PIK3CA, PIK3R1, PNP, PPARG, PTCH1, PTEN, PTPN11, RB1, RET, RPS6KB1, SMAD4, SMARCB1, SOX2, STK11, TERT, TET2, TIAF1, TP53, TSC1, TSC2, VHL, WT1, ZNF217*
Completely Affected Genes	*CHEK2, EZH2, GNAQ, JAK2, MAGOH, NFE2L2, NPM1, PAX5, RAC1, RAF1, RHEB, SF3B1, SPOP, U2AF1, XPO1*

Table 2. Oncomine Comprehensive Panel (OCP) target genes by presence of the MseI restriction site. Genes included in the Oncomine Comprehensive Panel (OCP) divided in three categories based on the presence of the MseI restriction site in at least one amplicon.

Library preparation, template, and sequencing using the Ion Torrent Proton were performed according to the manufacturer's instructions. Data analysis was performed using Torrent Suite 4.0.2 – 5.0.2, with alignment by TMAP using default parameters, and variant calling with the Ion Torrent Variant Caller plugin (version v4.0-r76860 – v5.0.2.1) using default low-stringency somatic variant settings. Variant annotation, filtering and prioritization was performed essentially as previously described using validated in-house pipelines [101-103]. Briefly, called variants were filtered to remove synonymous or non-coding variants, those with flow corrected read depths (FDP) less than 10, flow corrected variant allele-containing reads (FAO) less than 10, variant allele frequencies (FAO/FDP) less than 0.10, extreme skewing of forward/reverse flow corrected reads calling the variant (FSAF/FSAR <0.2 or >5), or indels within homopolymer runs >4. Called variants were filtered using a panel-specific, in house blacklist. Variants reported at population allele frequencies >0.5% in ESP6500 or 1000 Genomes

were considered germ line variants unless occurring at a known cancer hot-spot. Variants located at the last mapped base (or outside) of amplicon target regions, those with the majority of supporting reads harboring additional mismatches or indels (likely sequencing error), those in repeat–rich regions (likely mapping artifacts), and those occurring exclusively in one amplicon if overlapping amplicons cover the variant, were excluded. High confidence somatic variants passing the above criteria were then visually confirmed in Integrated Genome Viewer (IGV) as were the same regions in samples from the same patient where the variant was not called to confirm both coverage and absence of the variant. We have previously confirmed that these filtering criteria identify variants that pass Sanger Sequencing validation with >95% accuracy [101, 103, 105]. We prioritized potential driving alterations using the Oncomine annotation in the Oncomine Knowledge Base, which uses pan-cancer NGS data to identify genes as oncogenes or tumor suppressors, based on overrepresentation of hot-spot or deleterious mutations, respectively. Variants in oncogenes are then considered gain-of-function (GoF) if at a hot-spot, and variants in tumor suppressors are considered loss-of-function (LoF) if deleterious or at a hot-spot [101, 102, 105].

Copy number analysis from total amplicon read counts provided by the coverageAnalysis plugin (version v4.4.2.2 - v5.0.2.0) was performed essentially as previously described using a validated approach [101-103, 106] with adaptations for single cell sequencing. Specifically, for copy number analysis, we retained only well-performing amplicons (with >100 reads in all individual or pooled WBC samples,795 total amplicons) in order to exclude amplicons lost due to their containing the MseI restriction site. To identify CNAs, normalized read counts per amplicon for each sample were divided by the normalized mean read count of the corresponding amplicons from the WBC samples, yielding a copy-number ratio for each amplicon. Gene-level

copy-number estimates were determined by taking the coverage-weighted mean of the per amplicon ratios, with expected error determined by the amplicon-to-amplicon variance [106]. Genes with a log2 copy-number estimate < -2 or > 2 were considered to have high-level loss or gain, respectively.

2.3.6 Confirmation of *ESR1*, *CDH1* and *BRCA2* mutations by Sanger Sequencing

Bidirectional Sanger Sequencing to validate patient #2 *CDH1* (p.Q641X) and *ESR1* (p.Y537S, p.D538G, p.A569S) and patient #17 *BRCA2* (p.Q1931X) somatic variants was performed on selected samples. WGA CTC/WBC Genomic DNA (10 ng) was used as template in PCR amplifications with Invitrogen Platinum PCR Supermix Hi-Fi (ThermoFisher, Waltham, MA) with the suggested denaturation and cycling conditions. Primer sequences were designed with the PrimerQuest tool from IDT (Coralville, IA) (*CDH1* Q641 forward: CAGACCTTCCTCCCAATACATC, reverse: AAGGGAAGCATGGCAGTT; *ESR1* Y537/D538 forward: CTTTCTGTGTCTTCCCACCTAC, reverse: TTTGGTCCGTCTCCTCCA; *ESR1* A569 forward: CTGCTGGAGATGCTGGAC, reverse: AGGGAAACCCTCTGCCT; *BRCA2* Q1931 forward: TTGTTACGAGGCATTGGATGA, reverse: ACTGACTTATGAAGCTTCCCTATAC) (**Table 3**). PCR products were subjected to bidirectional Sanger Sequencing by the University of Michigan DNA Sequencing Core after purification with the QIAquick PCR Purification Kit (Qiagen, Hilden, Germany) and sequencing traces were analyzed using SeqMan Pro software (DNASTAR) and visually inspected.

Pre-amplification Primers Sequences	Forward	Reverse
ESR1 Y537S/D538G	TCCCAGCTCCCATCCTAAAGTG	CACTGCGGGCTCTACTTCAT
ESR1 A569S	TGCTGCTGGAGATGCTGGAC	CCTCCCCCGTGATGTAATACT
ddPCR Primers Sequences	**Forward**	**Reverse**
ESR1 Y537S/D538G	GCATGAAGTGCAAGAACGTG	AGACGGACCAAAGCCACTT
ESR1 A569S	TACATGCGCCCACTAGCCGT	GCAAGGAATGCGATGAAGTA
ddPCR Probes Sequences	**Mutant**	**Wild type**
ESR1 Y537S	TCTcTGACCTGCTGCTGGAGATGCT	TCTATGACCTGCTGCTGGAGATGCT
ESR1 D538G	TCTATGgCCTGCTGCTGGAGATGCT	TCTATGACCTGCTGCTGGAGATGCT
ESR1 A569S	ACCAAAGCCACTTGtCCACTGCGG	ACCAAAGCCACTTGGCCACTGCGG

Table 3. Primer/probe sequences used in Sanger sequencing and ddPCR

2.3.7 Tissue biopsy whole exome sequencing

Sequencing of clinical samples was performed as previously described [48]. Genomic DNA from frozen needle biopsies was used to generate exome libraries of matched tumor/normal pairs using the Illumina TruSeq DNA Sample Prep kit. WES was performed on Illumina HiSeq 2000 or HiSeq 2500 (paired-end) and analyzed as previously described[48], see below.

2.3.8 Tissue Analysis: Specimen collection and processing

Sequencing of clinical samples was performed under IRB-approved studies at the University of Michigan. Patients were enrolled and consented for integrative tumor sequencing in MI-ONCOSEQ Protocol, HUM00046018, as previously described [48]. Needle biopsies were snap frozen in Optimal Cutting Temperature compound, and a longitudinal section was cut. Frozen sections stained with hematoxylin and eosin were reviewed by pathologists to identify cores with the highest tumor content. Remaining portions of each needle biopsy core were retained for nucleic acid extraction.

2.3.9 DNA extraction and library preparation for exome sequencing.

Genomic DNA from frozen needle biopsies was isolated using the Qiagen DNeasy Blood and Tissue kit, according to the manufacturer's instructions. Exome libraries of matched tumor/normal paired DNA were generated using the Illumina TruSeq DNA Sample Prep kit, per manufacturer's instructions as previously described [48]. Briefly, 1–3 μg of DNA was sheared to a 250 bp peak target size, and end repaired followed by A-base addition and ligation of Illumina indexed adaptors. Agarose gel purified 300–350 bp fragments were amplified using Illumina index primers for eight cycles and purified. One μg of library was hybridized to the Agilent SureSelect Human All Exon v4 chip (Agilent Technologies) and hybridized exon fragments were captured and amplified with the Illumina index primers for nine additional PCR cycles. Purified PCR products were analyzed for quality and quantity with the Agilent 2100 Bioanalyzer (DNA 1000 kit). WES was performed on Illumina HiSeq 2000 or HiSeq 2500 (paired-end). Sequencing quality was assessed with FastQC. For each sequencing lane, we examined per-base quality scores across the length of the reads with lanes deemed passing if the per-base quality is >Q20 in >85% of the reads for bases 1–100. We also assessed alignment quality (Picard package) by monitoring duplication rates and potential chimeric reads from ligation artifacts.

2.3.10 Tissue mutation analysis.

Analysis was performed as previously described [48]. Briefly, FASTQ converted sequence files were processed through an in-house pipeline for WES analyses of paired cancer and normal genomes. Sequencing reads were aligned to reference genome hg19 with Novoalign multithreaded (v2.08.02, Novocraft) and converted into BAM files using SAMtools (v0.1.18)

[107]. Mutation analysis was performed using VarScan2 (v2.3.2) [108] with the pileup files created by SAMtools mpileup for tumor and matched normal samples, simultaneously performing pairwise comparisons of base call and normalized sequence depth at each position. For SNV detection, filtering parameters including coverage, variant read support, variant frequency, P value, base quality, the presence of homopolymers and strandedness were applied. For indel analysis, Pindel (v0.2.4) was used and indels common to both tumor and matched normal samples were classified as germline, whereas indels present in tumor but not in normal samples were classified as somatic. Finally, a list of candidate indels and somatic mutations was generated by excluding synonymous SNVs. Detected variants were functionally annotated with ANNOVAR [109].

Tumor content for each tumor exome library was estimated from the sequence data by fitting a binomial mixture model with two components to the set of most likely SNV candidates from two-copy genomic regions as determined by copy number analysis, as previously described [48]. Copy number aberrations were quantified for each gene as the segmented, normalized, GC content-corrected, log2-transformed exon coverage ratio between each tumor sample and its matched normal sample [110].

2.3.11 ddPCR analysis of tissue and cfDNA: DNA extraction from FFPE blocks

Formalin fixed paraffin embedded (FFPE) blocks and slides were obtained from University of Michigan Department of Pathology archives. A representative block from each specimen was chosen and DNA extracted using a commercial kit (Qiagen QiAMP FFPE tissue kit) according to the manufacturer's instructions.

2.3.12 Isolation of cfDNA for ddPCR

Blood sample and plasma preparation were performed as previously described [111]. Blood was collected in DNA BC tube (Streck) and centrifuged within 7 days. Plasma was obtained by a double-spin centrifuge protocol of whole blood to remove cellular concomitants and DNA was extracted using the Qiagen Circulating Nucleic Acid Kit (Qiagen) per the manufacture's protocol as previously described [90].

2.3.13 ddPCR for FFPE DNA and cfDNA

Genomic and plasma DNA samples were subjected to short cycle PCR amplification for the *ESR1* p.Y537S/p.D538G and p.A569S locus with the primers shown in **Table 3**. Amplified DNA was purified using the QIAquick PCR Purification kit (Qiagen). The Bio-Rad QX200 platform was then used for ddPCR per the manufacturer's protocols with the ddPCR primers and probes listed in **Table 3**. The results are reported as fractional abundance of mutant DNA alleles to total (mutant plus wild-type) DNA alleles for each sample assayed.

2.3.14 In vitro functional studies of ESR1 A569S mutation

pCDH-ESR1 plasmid was mutated at alanine 569 of ER to serine with Quick Change Lightning Kit (Agilent Technologies). Cells were obtained from the Tissue Culture Shared Resource at the Lombardi Comprehensive Cancer Center (Georgetown University, Washington, DC) in 2001. Cells were never passaged more than 15 times, were free of mycoplasma contamination (most recent testing, May, 2015) and identity was confirmed by short tandem repeat (most recent testing, July, 2014). Lentivirus containing the p.A569S *ESR1* transgene was packaged in 293T cells. 24 hours after plating, cells were transfected with 8ug of pCDH-*ESR1*-

A569S plasmid, 5ug psPAX2 and 2ug pMD2.G plasmids. MCF-7 cells were virally transduced with 1ml of viral supernatant supplemented with 4ug/ml polybrene for 8 hours. Steroid-depleted parental and *ESR1* p.A569S MCF-7 cells were seeded into 96-well plates and treated with 17β-estradiol (Sigma Aldrich) or ethanol control alone or in combination with tamoxifen, 4-hydroxytamoxifen, endoxifen or fulvestrant. Cell number was assessed by crystal violet stain five days after hormone treatment as previously described [112].

2.3.15 Steroids and drugs

17β-estradiol, tamoxifen, fulvestrant, polybrene and caffeine were obtained from Sigma Aldrich. 4-hydroxytamoxifen and endoxifen were obtained from Toronto Research Chemical.

2.3.16 Plasmids and mutagenesis

The lentiviral vector for wild-type *ESR1* coding sequence, pCDH-ESR1, was obtained from stock used in [48]. The viral packaging plasmids psPAX2 and pMD2.G were gifts from Didier Trono (Addgene plasmids # 12260 and 12259, respectively). The previously constructed plasmid pCDH-ESR1 was mutated with Quick Change Lightning Kit (Agilent Technologies). The alanine residue at position 569 of ER was mutated to a serine residue according the manufacturer's protocol. The sequences for the sense and anti-sense primers for site-directed mutagenesis were as follows, sense: 5'-GAGCCCGCA GTGGAC<u>C</u>AAGTGGCTTTGG-3'; anti-sense: 5'-CCAAAGCCACTTG<u>T</u>CCACTGCGGGCTC-3' (the underlined nucleotides are mutated). Mutagenesis was confirmed by Sanger Sequencing.

2.3.17 Cell Culture

MCF-7 and 293T cells were obtained from the Lombardi Comprehensive Cancer Center and routinely maintained in DMEM (ThermoFisher) supplemented with 5% fetal bovine serum (Valley Biomedical). Lentivirus containing the p.A569S *ESR1* transgene was packaged in 293T cells. 293T cells were plated at a density of 2 x 10^6 cells/10 cm dish. 24 hours after plating, cells were transfected with 8ug of pCDH-*ESR1*-A569S plasmid, 5ug psPAX2 and 2ug pMD2.G plasmids via calcium chloride. Transfection medium was allowed to remain on the cells for 8 hours and then replaced with fresh DMEM/5% FBS medium and cells were incubated overnight. On the following day, 12 ml of DMEM/5% FBS supplemented with 0.5mM caffeine was added to the dish for virus production. 48 hours later, the supernatant was carefully removed from the dish, filtered through a 0.45 um filter and frozen down in 1 ml aliquots.

MCF-7 cells were seeded into a 6-well plate at a density of 0.25 x 10^6 cells per well and allowed to attach overnight. 24 hours after plating, cells were virally transduced with 1ml of viral supernatant supplemented with 4ug/ml polybrene for 8 hours and then restored in fresh DMEM/5% FBS. 48 hours after viral transduction, cells were stably selected in DMEM/5% FBS medium supplemented with 1ug/ml puromycin for one week. p.A569S *ESR1* transgene expression was confirmed by Sanger Sequencing of genomic DNA and cDNA reverse transcribed from 1ug of DNase-treated total RNA. For growth assays in defined hormone conditions, cells were depleted of exogenous steroids as previously described [112]. Steroid-depleted parental and *ESR1* p.A569S MCF-7 cells were seeded into 96-well plates at a density of 1,000 cells/well in 0.1 ml and allowed to attach overnight. Cells were treated with 17β-estradiol (Sigma Aldrich) or ethanol control alone or in combination with tamoxifen, 4-hydroxytamoxifen, endoxifen or fulvestrant. Cell number was assessed by crystal violet stain five days after hormone treatment as previously published [112].

2.4 Results

2.4.1 Preclinical, proof-of-concept validation of targeted NGS of CTCs from archived CellSearch® cartridges

We performed a pilot, pre-clinical study to determine if cells stored for long periods of time in CellSearch® cartridges after enrichment from whole blood (WB) could be purified using the DEPArray™ technology for subsequent high quality genomic profiling. Cultured BT-474 human breast cancer cells (containing a *TP53* p.E285K hotspot mutation) spiked into normal human blood and pre-enriched by CellSearch® were archived in the collection cartridges for several months-years. After separation from leucocytes to purity, single cell DNA derived from spiked cells underwent whole genome amplification (WGA) using Ampli1™ WGA kit and NGS with an ~300-amplicon, 50-gene multiplexed PCR based panel (CHPv2). This approach successfully identified the expected *TP53* p.E285K mutation in cells from all seven samples processed, with an elapsed time between archiving of the CellSearch® cartridge (in 4°C) and cell isolation using DEPArray™ of three days to six months; cartridges archived for two years in RT did not yield assessable genomic DNA (**Table 4**).

Cartridge Number	Number of CTC Single (S) Pooled (P)	Elapsed time between CellSearch® and DEPArray	Storage condition	Genome integrity index (GII)	*TP53* (p.E285K) Variant frequency
1	1 (S)	3 days	4°C	4	1
	1 (S)	3 days		4	0.8
	1 (S)	3 days		4	1
2	1 (S)	6 months	4°C	3	1
	1 (S)	6 months		2	1
	1 (S)	6 months		3	1
	9 (P)	6 months		3	1
3	1 (S)	2 years	RT	1	*
	1 (S)	2 years		1	*

Table 4. NGS of BT-474 cells spiked into blood and purified identifies known TP53 mutation. NGS of CellSearch®-enriched and DEPArray™-purified spiked in cells from the BT-474 breast cancer cell line with the CHPv2 panel. Number of cells per sample, storage time and conditions, DNA quality (4=high) and the detected variant frequency of the expected mutation are shown. RT= room temperature; * Low library quality and uniformity of coverage over the target regions.

2.4.2 Trial cohort for CTC assessment and comparison to matched tissue metastases

Genomic ET resistance mechanisms in individual CTC and concordance with biopsy obtained fresh tissue was examined in thirty patients with MBC. These patients participated in the Michigan Oncology Sequencing Center (MI-ONCOSEQ) trial, in which a biopsy of metastatic tissue was subjected to genomic profiling. They were also enrolled in a companion trial (MI-CTC-ONCOSEQ) to collect WB for CTC enrichment and purification using the CellSearch® and DEPArray™ systems. Two patients were deemed ineligible for regulatory reasons. Of the 28 remaining patients, 16 (57%) had ≥5 CTC/7.5 ml whole blood (WB) by CellSearch® at baseline (**Figure 2**).

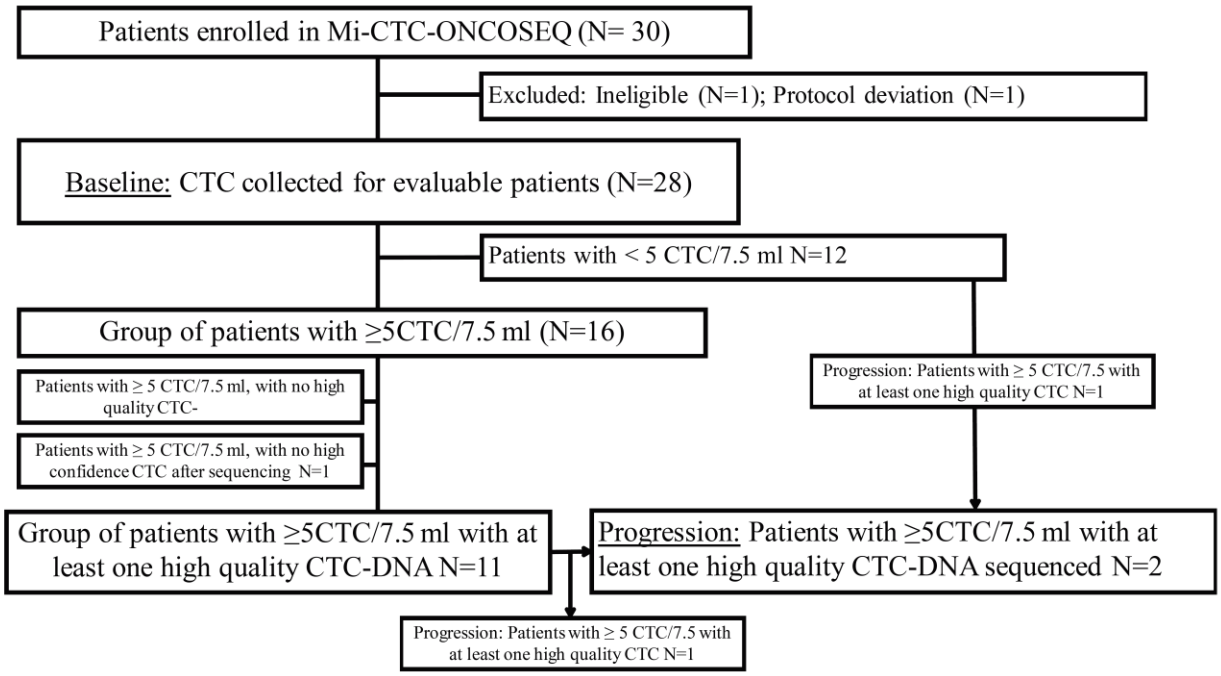


Figure 2. REMARK diagram for patient enrollment and distribution. Of 30 enrolled patients, 12 were protocol-conforming, had ≥5 CTC/7.5 ml WB by CellSearch® and had at least one CTC with high quality DNA at baseline or progression.

Approximately 15% of the enriched CTC were purified using DEPArray™ and had high quality DNA (**Table 1,** referenced earlier). Eleven patients had at least one CTC with sufficiently high quality DNA for genomic analysis.

Two patients had a blood sample drawn containing ≥5 CTC/7.5 ml WB at the time of disease progression, so that a total of 12 patients had a sufficient number of CTC for genomic analysis at either baseline or progression. Eleven of the 12 patients had been diagnosed with HR-positive breast cancer (either from primary or metastatic tissue); the remaining patient (#19) was diagnosed with triple negative (ER/progesterone receptor [PR]$^-$/HER2 [*ERBB2*]$^-$) breast cancer in both primary and metastatic tissue. Clinical histories are shown in **Figure 3**.

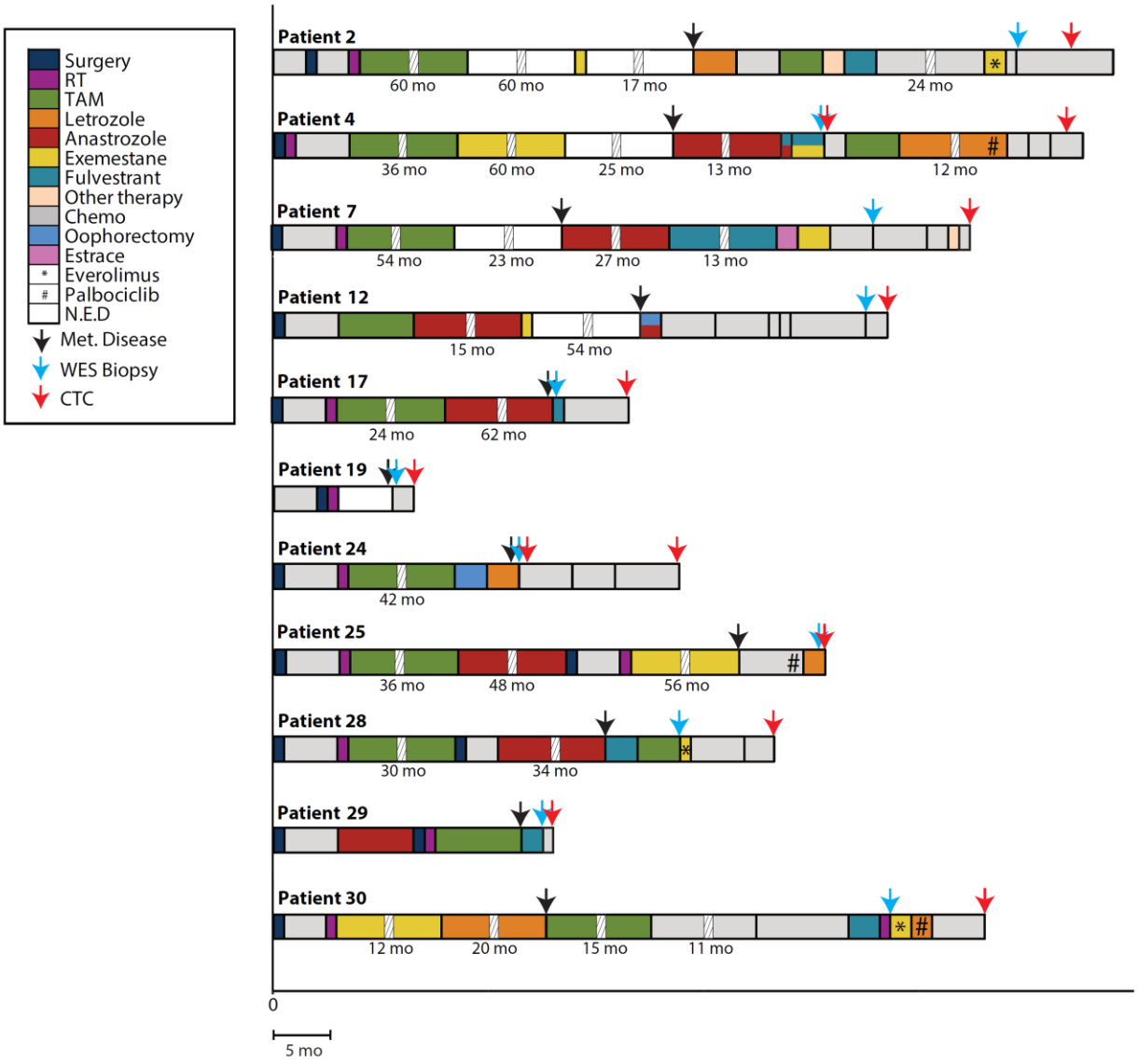


Figure 3. Clinical timelines. Clinical timelines and treatment from first diagnosis until enrollment into MI-CTC-ONCOSEQ for the 11 metastatic breast cancer patients (No available timeline for patient #14). Each bar represents the timeframe of treatment. Colors represent different therapies: Surgery=■; RT=■; NED= ; Tamoxifen=■; Letrozole =■; Anastrazole=■; Exemestane=■; Fulvestrant =■; Therapy other than chemo-or endocrine-therapy =■; chemotherapy =■; oophorectomy = ■; Estrace = ■; Everolimus = *; Palbociclib = #.

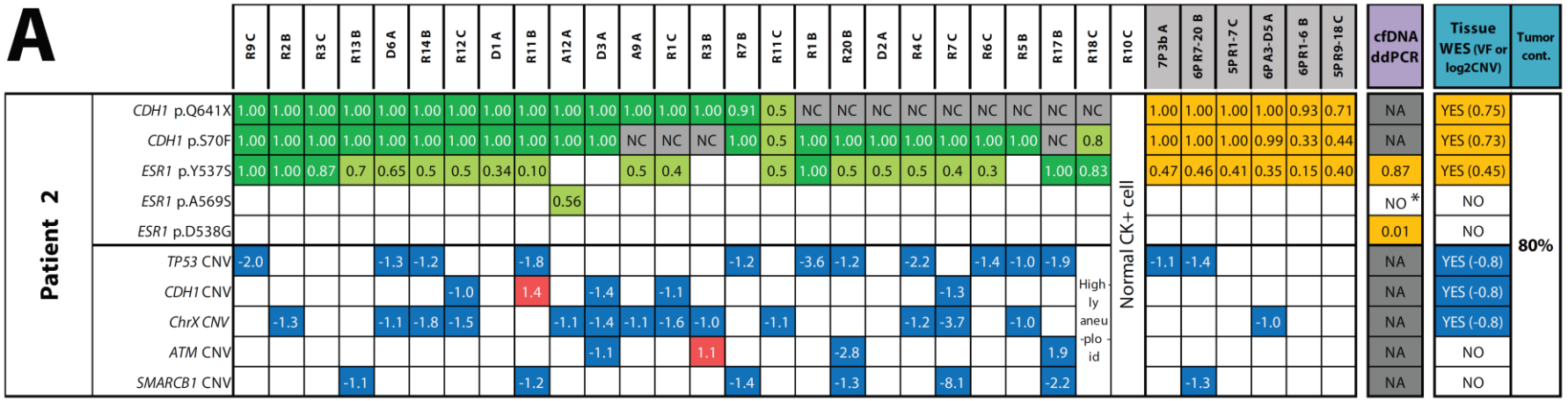
2.4.3 Simultaneous assessment of somatic mutations and copy number alterations in archived CTC from patients with metastatic breast cancer

From the 12 patients with evaluable CTC, individual CTC were isolated from archived CellSearch® cartridges by DEPArray™, followed by genomic DNA extraction and WGA. DNA from 53 individual CTC, 23 pooled CTC samples (containing equal DNA amounts from 2 - 7 individual CTC), and 16 individual or pooled white blood cells (WBC) was subjected to comprehensive multiplexed-PCR based NGS using the DNA component of the Oncomine Cancer Assay (OCP) (**Table 5**). This assay, which is also being used in the National Cancer Institute MATCH trial [113], interrogates activating and deleterious mutations and CNAs in 130 genes [102]. All patients had WES performed on near synchronous metastatic tissue biopsies as part of the MI-ONCOSEQ platform. After OCP profiling, CTC doubly positive ($CD45^+/CK^+$) or pooled CTC samples lacking all molecular alterations (any somatic CNAs or prioritized mutations) were excluded, resulting in 67 evaluable CTC samples (both single and pooled cells) from 12 patients.

Mapped Reads	Reads On Target	Mean Sequencing Depth	Sequencing Uniformity	Variants
3,683,759	96.93%	1,508	41.05%	151

Table 5: Mean CTC Next Generation Sequencing parameters. Mean sequencing parameters for 91 samples including mapped reads (total number of targeted reads per sample), reads on target (percentage of total reads that mapp to regions targeted by the sequencing panel), mean sequencing depth (mean number of reads covering each targeted base), sequencing uniformity (percentage of targeted bases that have at least 0.2X of the mean sequencing depth), variants (mean of total number of sequencing variants detected per sample by the somatic low stringency setting).

In these 67 evaluable CTC samples, we identified a total of 23 high-confidence, prioritized, somatic, nonsynonymous point mutations and short insertions/deletions, (median = 1.5, range = 0–6 per patient) and 31 high-confidence high-level CNAs (median = 2.5, range = 0–6 per patient), as shown in an integrative heat map (**Figure 4A; B**).


A
Patient 2
R9 C
R2 B
R3 C
R13 B
D6 A
R14 B
R12 C
D1 A
R11 B
A12 A
D3 A
A9 A
R1 C
R3 B
R7 B
R11 C
R1 B
R20 B
D2 A
R4 C
R7 C
R6 C
R5 B
R17 B
R18 C
R10 C
7P3b A
6P R7-20 B
5P R1-7 C
6P A3-D5 A
6P R1-6 B
5P R9-18 C
cfDNA ddPCR
Tissue WES (VF or log2CNV)
Tumor cont.
CDH1 p.Q641X
CDH1 p.S70F
ESR1 p.Y537S
ESR1 p.A569S
ESR1 p.D538G
TP53 CNV
CDH1 CNV
ChrX CNV
ATM CNV
SMARCB1 CNV
Highly aneu-plo-id
Normal CK+ cell
NO *
80%

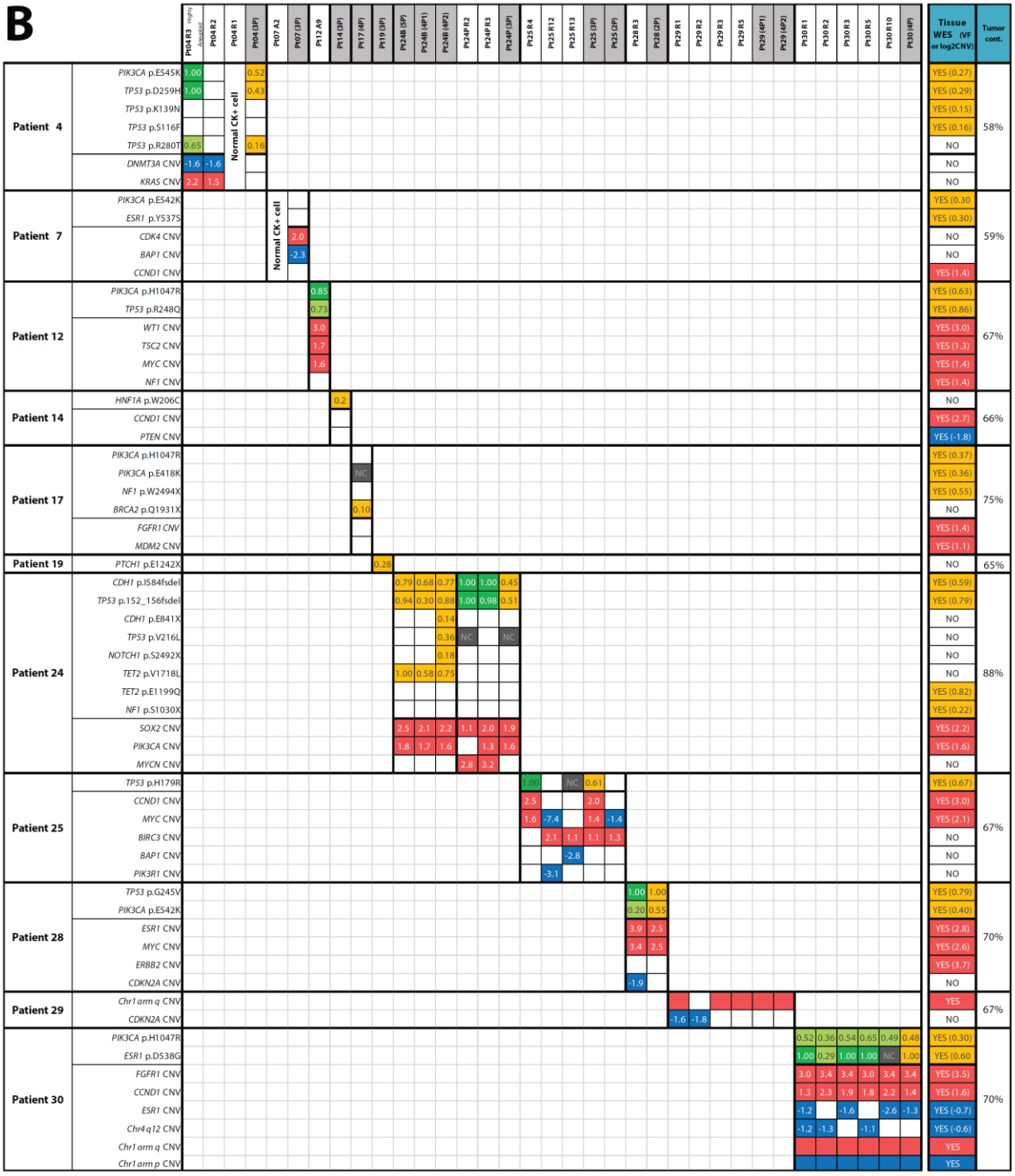
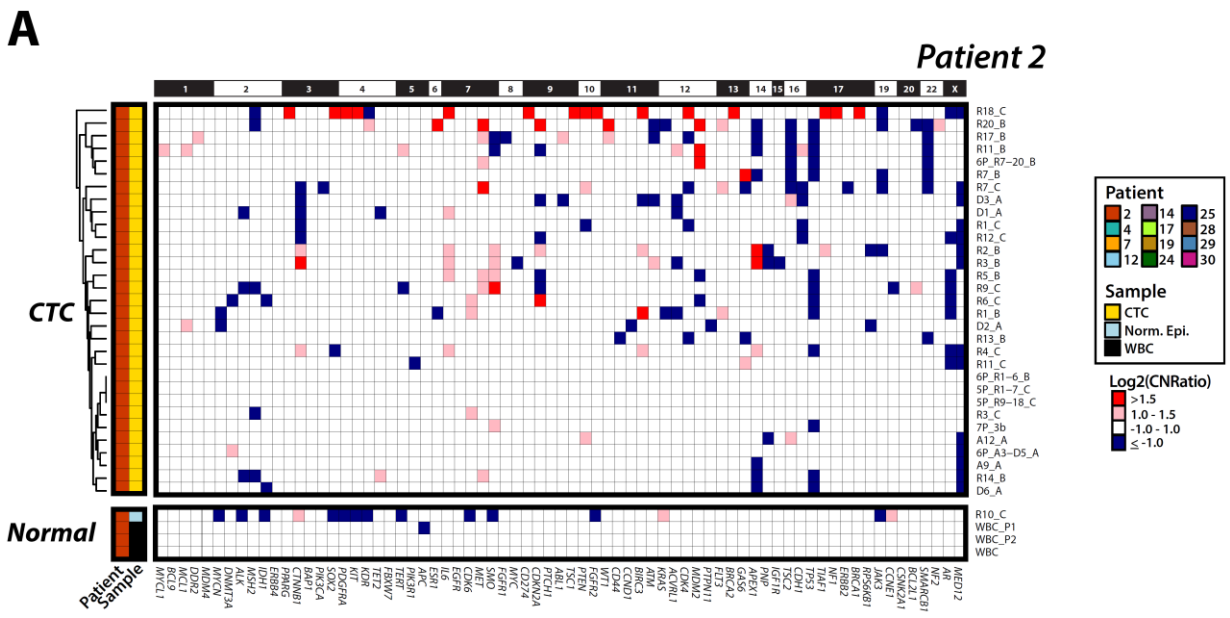


Figure 4. Integrative heat map of somatic molecular alterations identified in archived circulating tumor cells (CTC) and comparison with metastatic tissue in endocrine therapy-resistant metastatic breast cancer patients. Next generation targeted sequencing for circulating tumor cells (CTC), whole exome sequencing (WES) for tissue biopsy, and digital droplet PCR (ddPCR) for cell free tumor DNA (cfDNA) identified high confidence mutations (top half of each patient table) and copy number alterations (CNAs) (bottom half) in **A)** Patient #2 and **B)** Patients #4

– 30. Colored boxes indicate presence of alteration, empty boxes represent absence of alteration (despite adequate sequencing coverage for the position). NC = no adequate sequencing coverage to evaluate mutation presence; NA = not assayed. Numbers inside mutation boxes show variant frequency (VF), with dark and light green boxes indicating homozygous (>0.8 VF) and heterozygous/complex (<0.8 VF) mutations, respectively. Orange boxes indicate presence of mutation and corresponding VF for CTC pools, tissue WES, and cfDNA. Numbers inside CNA boxes show the $\log_2$ copy number ratio (CNR), with gains and losses shown in red and blue, respectively. CTC IDs are shown at the top of heat map with pooled sample IDs shaded. Note: tumor cont. = tumor content; for patient #24, "B" and "P" in sample ID represent CTC samples at baseline and progression, respectively; *ddPCR droplets for this mutation were detected, but below the predetermined threshold.

Importantly, no high-confidence mutations were found in the matched WBC in any of the 12 patients, and high-confidence CNAs in WBC were exceedingly rare and were limited to occasional copy losses, consistent with high fidelity purification, WGA and NGS (**Figure 5A; B**).



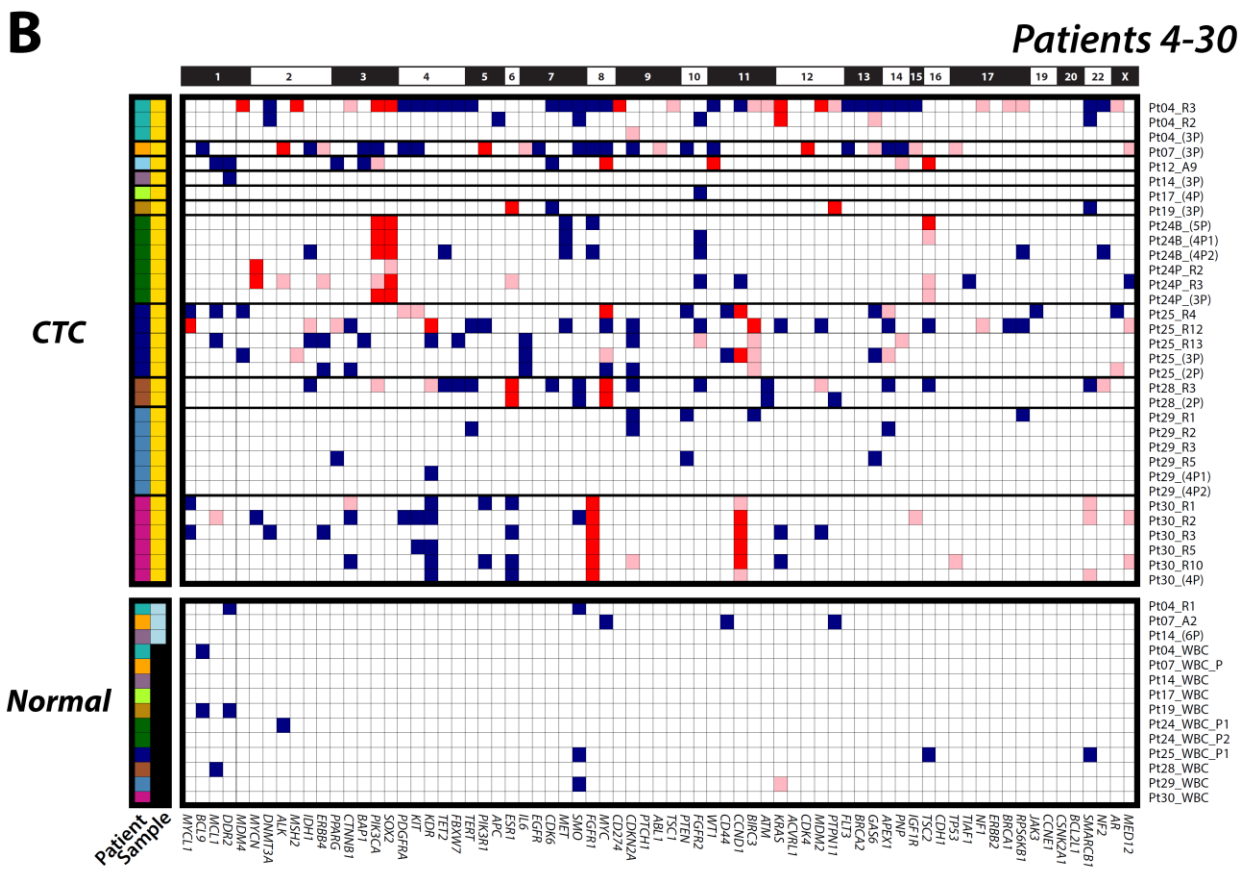


Figure 5. Heterogeneity of copy number alterations (CNAs) detected in circulating tumor cells (CTC) across patients with metastatic breast cancer. Next generation targeted sequencing generated gene-level $\log_2$ copy number ratios (CNR), with copy number alteration (CNA) losses in blue (negative $\log_2$CNR) and gains in red (positive CNR). **A.** Patient #2; and **B.** Patients #4 – 30. Greater absolute value of $\log_2$CNR indicates a more pronounced CNA. A dendrogram from unsupervised hierarchical clustering of CNAs in Patient #2 CTC is shown on the left.

2.4.4 Comparison of somatic mutations and copy number alterations in matched CTC and tissue metastases

Our study design provided the opportunity to compare prioritized somatic mutations and CNAs in CTC vs. synchronous/near-synchronous metastatic tissue biopsies. Critically, we observed highly concordant somatic alterations from CTC subjected to targeted NGS and matched WES of fresh metastatic tissue biopsy. Specifically, 57 of 67 CTC samples (85%) and

CTC in 8 of 12 patients (67%) showed at least one, but usually multiple, prioritized genomic alterations detected in the corresponding tissue biopsy (**Figure 4A; B; Table 6**).

SAMPLE	Gene	Location	REFERENCE	ALTERNATE	Aminoacid	RefSeq	Exonic Function	Vartaiont Fraction	Varaint Reads	Total Reads
Pt02 5P R1-R7 C	CDH1	chr16:68856113	C	T	p.Q641X	NM_004360	stopgainSNV	1.00	246	246
Pt02 5P R9-R18 C	CDH1	chr16:68856113	C	T	p.Q641X	NM_004360	stopgainSNV	0.71	550	771
Pt02 6P A3-D5 A	CDH1	chr16:68856113	C	T	p.Q641X	NM_004360	stopgainSNV	1.00	1995	1995
Pt02 6P R1-R6 B	CDH1	chr16:68856113	C	T	p.Q641X	NM_004360	stopgainSNV	0.93	569	611
Pt02 6P R7-R20 B	CDH1	chr16:68856113	C	T	p.Q641X	NM_004360	stopgainSNV	1.00	21	21
Pt02 7P 3b A	CDH1	chr16:68856113	C	T	p.Q641X	NM_004360	stopgainSNV	1.00	1652	1652
Pt02 A12 A	CDH1	chr16:68856113	C	T	p.Q641X	NM_004360	stopgainSNV	1.00	1997	1997
Pt02 A9 A	CDH1	chr16:68856113	C	T	p.Q641X	NM_004360	stopgainSNV	1.00	1993	1993
Pt02 D1 A	CDH1	chr16:68856113	C	T	p.Q641X	NM_004360	stopgainSNV	1.00	1993	1993
Pt02 D3 A	CDH1	chr16:68856113	C	T	p.Q641X	NM_004360	stopgainSNV	1.00	1994	1994
Pt02 D6 A	CDH1	chr16:68856113	C	T	p.Q641X	NM_004360	stopgainSNV	1.00	1996	1996
Pt02 R1 C	CDH1	chr16:68856113	C	T	p.Q641X	NM_004360	stopgainSNV	1.00	441	441
Pt02 R11 B	CDH1	chr16:68856113	C	T	p.Q641X	NM_004360	stopgainSNV	1.00	12	12
Pt02 R11 C	CDH1	chr16:68856113	C	T	p.Q641X	NM_004360	stopgainSNV	0.52	438	846
Pt02 R12 C	CDH1	chr16:68856113	C	T	p.Q641X	NM_004360	stopgainSNV	1.00	1676	1676
Pt02 R13 B	CDH1	chr16:68856113	C	T	p.Q641X	NM_004360	stopgainSNV	1.00	120	120
Pt02 R14 B	CDH1	chr16:68856113	C	T	p.Q641X	NM_004360	stopgainSNV	1.00	37	37
Pt02 R2 B	CDH1	chr16:68856113	C	T	p.Q641X	NM_004360	stopgainSNV	1.00	262	262
Pt02 R3 B	CDH1	chr16:68856113	C	T	p.Q641X	NM_004360	stopgainSNV	1.00	1584	1584
Pt02 R7 B	CDH1	chr16:68856113	C	T	p.Q641X	NM_004360	stopgainSNV	0.91	21	23
Pt02 R3 C	CDH1	chr16:68856113	C	T	p.Q641X	NM_004360	stopgainSNV	1.00	489	489
Pt02 R9 C	CDH1	chr16:68856113	C	T	p.Q641X	NM_004360	stopgainSNV	1.00	116	116
Pt02 R17 B	CDH1	chr16:68863698	G	T	p.E813X	NM_004360	stopgainSNV	0.83	149	180
Pt02 5P R1-R7 C	CDH1	chr16:68835618	C	T	p.S70F	NM_004360	nonsynonymousSNV	1.00	1980	1980
Pt02 5P R9-R18 C	CDH1	chr16:68835618	C	T	p.S70F	NM_004360	nonsynonymousSNV	0.44	867	1976
Pt02 6P A3-D5 A	CDH1	chr16:68835618	C	T	p.S70F	NM_004360	nonsynonymousSNV	0.98	1964	1997
Pt02 6P R1-R6 B	CDH1	chr16:68835618	C	T	p.S70F	NM_004360	nonsynonymousSNV	0.33	665	1997
Pt02 6P R7-R20 B	CDH1	chr16:68835618	C	T	p.S70F	NM_004360	nonsynonymousSNV	1.00	1990	1990
Pt02 7P 3b A	CDH1	chr16:68835618	C	T	p.S70F	NM_004360	nonsynonymousSNV	0.99	1981	1998
Pt02 A12 A	CDH1	chr16:68835618	C	T	p.S70F	NM_004360	nonsynonymousSNV	1.00	1990	1990
Pt02 D1 A	CDH1	chr16:68835618	C	T	p.S70F	NM_004360	nonsynonymousSNV	1.00	1991	1991
Pt02 D2 A	CDH1	chr16:68835618	C	T	p.S70F	NM_004360	nonsynonymousSNV	1.00	1989	1989
Pt02 D3 A	CDH1	chr16:68835618	C	T	p.S70F	NM_004360	nonsynonymousSNV	1.00	1990	1990
Pt02 D6 A	CDH1	chr16:68835618	C	T	p.S70F	NM_004360	nonsynonymousSNV	1.00	1989	1989
Pt02 R1 B	CDH1	chr16:68835618	C	T	p.S70F	NM_004360	nonsynonymousSNV	1.00	1984	1984
Pt02 R11 B	CDH1	chr16:68835618	C	T	p.S70F	NM_004360	nonsynonymousSNV	1.00	1980	1980
Pt02 R11 C	CDH1	chr16:68835618	C	T	p.S70F	NM_004360	nonsynonymousSNV	0.52	1029	1986
Pt02 R12 C	CDH1	chr16:68835618	C	T	p.S70F	NM_004360	nonsynonymousSNV	1.00	1983	1983
Pt02 R13 B	CDH1	chr16:68835618	C	T	p.S70F	NM_004360	nonsynonymousSNV	1.00	1981	1981
Pt02 R14 B	CDH1	chr16:68835618	C	T	p.S70F	NM_004360	nonsynonymousSNV	1.00	1991	1991
Pt02 R2 B	CDH1	chr16:68835618	C	T	p.S70F	NM_004360	nonsynonymousSNV	1.00	1974	1974
Pt02 R20 B	CDH1	chr16:68835618	C	T	p.S70F	NM_004360	nonsynonymousSNV	1.00	1989	1989
Pt02 R3 C	CDH1	chr16:68835618	C	T	p.S70F	NM_004360	nonsynonymousSNV	1.00	1980	1980
Pt02 R4 C	CDH1	chr16:68835618	C	T	p.S70F	NM_004360	nonsynonymousSNV	1.00	1989	1989
Pt02 R5 B	CDH1	chr16:68835618	C	T	p.S70F	NM_004360	nonsynonymousSNV	1.00	1977	1977
Pt02 R6 C	CDH1	chr16:68835618	C	T	p.S70F	NM_004360	nonsynonymousSNV	1.00	1981	1981
Pt02 R7 B	CDH1	chr16:68835618	C	T	p.S70F	NM_004360	nonsynonymousSNV	1.00	1985	1985
Pt02 R7 C	CDH1	chr16:68835618	C	T	p.S70F	NM_004360	nonsynonymousSNV	1.00	1976	1976
Pt02 R9 C	CDH1	chr16:68835618	C	T	p.S70F	NM_004360	nonsynonymousSNV	1.00	1991	1991
Pt02 R18 C	CDH1	chr16:68835618	C	T	p.S70F	NM_004360	nonsynonymousSNV	0.79	11	14
Pt02 A12 A	ESR1	chr6:152420018	G	T	p.A569S	NM_000125	nonsynonymousSNV	0.56	1117	1994
Pt02 5P R1-R7 C	ESR1	chr6:152419923	A	C	p.Y537S	NM_000125	nonsynonymousSNV	0.41	819	1998
Pt02 5P R9-R18 C	ESR1	chr6:152419923	A	C	p.Y537S	NM_000125	nonsynonymousSNV	0.40	797	2000
Pt02 6P A3-D5 A	ESR1	chr6:152419923	A	C	p.Y537S	NM_000125	nonsynonymousSNV	0.35	693	1997
Pt02 6P R1-R6 B	ESR1	chr6:152419923	A	C	p.Y537S	NM_000125	nonsynonymousSNV	0.15	299	2000
Pt02 6P R7-R20 B	ESR1	chr6:152419923	A	C	p.Y537S	NM_000125	nonsynonymousSNV	0.46	911	1998
Pt02 7P 3b A	ESR1	chr6:152419923	A	C	p.Y537S	NM_000125	nonsynonymousSNV	0.47	947	1997
Pt02 A9 A	ESR1	chr6:152419923	A	C	p.Y537S	NM_000125	nonsynonymousSNV	0.52	1046	1995
Pt02 D1 A	ESR1	chr6:152419923	A	C	p.Y537S	NM_000125	nonsynonymousSNV	0.34	676	1998
Pt02 D2 A	ESR1	chr6:152419923	A	C	p.Y537S	NM_000125	nonsynonymousSNV	0.46	920	1995

Table continued below

SAMPLE	Gene	Location	REFERENCE	ALTERNATE	Aminoacid	RefSeq	Exonic Function	Vartaiont Fraction	Varaint Reads	Total Reads
Pt02 D6 A	ESR1	chr6:152419923	A	C	p.Y537S	NM_000125	nonsynonymousSNV	0.65	1288	1994
Pt02 R1 B	ESR1	chr6:152419923	A	C	p.Y537S	NM_000125	nonsynonymousSNV	1.00	715	715
Pt02 R1 C	ESR1	chr6:152419923	A	C	p.Y537S	NM_000125	nonsynonymousSNV	0.37	730	1999
Pt02 R11 B	ESR1	chr6:152419923	A	C	p.Y537S	NM_000125	nonsynonymousSNV	0.10	631	5645
Pt02 R11 C	ESR1	chr6:152419923	A	C	p.Y537S	NM_000125	nonsynonymousSNV	0.49	988	1999
Pt02 R12 C	ESR1	chr6:152419923	A	C	p.Y537S	NM_000125	nonsynonymousSNV	0.48	956	2000
Pt02 R13 B	ESR1	chr6:152419923	A	C	p.Y537S	NM_000125	nonsynonymousSNV	0.68	1357	1999
Pt02 R14 B	ESR1	chr6:152419923	A	C	p.Y537S	NM_000125	nonsynonymousSNV	0.52	1030	2000
Pt02 R17 B	ESR1	chr6:152419923	A	C	p.Y537S	NM_000125	nonsynonymousSNV	1.00	1961	1961
Pt02 R2 B	ESR1	chr6:152419923	A	C	p.Y537S	NM_000125	nonsynonymousSNV	1.00	2000	2000
Pt02 R20 B	ESR1	chr6:152419923	A	C	p.Y537S	NM_000125	nonsynonymousSNV	0.51	1021	2000
Pt02 R3 C	ESR1	chr6:152419923	A	C	p.Y537S	NM_000125	nonsynonymousSNV	0.87	965	1110
Pt02 R4 C	ESR1	chr6:152419923	A	C	p.Y537S	NM_000125	nonsynonymousSNV	0.45	909	1999
Pt02 R6 C	ESR1	chr6:152419923	A	C	p.Y537S	NM_000125	nonsynonymousSNV	0.31	623	1996
Pt02 R7 C	ESR1	chr6:152419923	A	C	p.Y537S	NM_000125	nonsynonymousSNV	0.39	782	2000
Pt02 R9 C	ESR1	chr6:152419923	A	C	p.Y537S	NM_000125	nonsynonymousSNV	1.00	1998	1998
Pt02 R18 C	ESR1	chr6:152419923	A	C	p.Y537S	NM_000125	nonsynonymousSNV	0.83	10	12
Pt04 (3P)	PIK3CA	chr3:178936091	G	A	p.E545K	NM_006218	nonsynonymousSNV	0.53	190	361
Pt04 R3	PIK3CA	chr3:178936091	G	A	p.E545K	NM_006218	nonsynonymousSNV	1.00	1138	1996
Pt04 (3P)	TP53	chr17:7577506	C	G	p.D259H	NM_000546	nonsynonymousSNV	0.43	857	1821
Pt04 R3	TP53	chr17:7577506	C	G	p.D259H	NM_000546	nonsynonymousSNV	1.00	1986	1138
Pt04 (3P)	TP53	chr17:7577099	C	G	p.R280T	NM_000546	nonsynonymousSNV	0.16	292	1986
Pt04 R3	TP53	chr17:7577099	C	G	p.R280T	NM_000546	nonsynonymousSNV	0.65	1294	1998
Pt12 A9	PIK3CA	chr3:178952085	A	G	p.H1047R	NM_006218	nonsynonymousSNV	0.85	1703	1997
Pt12 A9	TP53	chr17:7577538	C	T	p.R248Q	NM_000546	nonsynonymousSNV	0.73	1457	1999
Pt14 (3P)	HNF1A	chr12:121431414	G	T	p.W206C	NM_000545	nonsynonymousSNV	0.17	270	1569
Pt17 (4P)	BRCA2	chr13:32914283	C	T	p.Q1931X	NM_000059	stopgainSNV	0.10	88	878
Pt19 (3P)	PTCH1	chr9:98211431	C	A	p.E1242X	NM_000264	stopgainSNV	0.28	390	1378
Pt24B (4P2)	CDH1	chr16:68867274	G	T	p.E841X	NM_004360	stopgainSNV	0.14	135	952
Pt24 R2	CDH1	chr16:68855944	TC	T	p.I584fs	NM_004360	frameshiftdeletion	1.00	1007	1007
Pt24 R3	CDH1	chr16:68855944	TC	T	p.I584fs	NM_004360	frameshiftdeletion	1.00	1978	1978
Pt24B (4P1)	CDH1	chr16:68855944	C	-	p.I584fs	NM_004360	frameshiftdeletion	0.69	1361	1986
Pt24B (4P2)	CDH1	chr16:68855944	C	-	p.I584fs	NM_004360	frameshiftdeletion	0.77	1536	1992
Pt24B (5P)	CDH1	chr16:68855944	C	-	p.I584fs	NM_004360	frameshiftdeletion	0.79	1568	1977
Pt24P (3P)	CDH1	chr16:68855944	TC	T	p.I584fs	NM_004360	frameshiftdeletion	0.45	749	1648
Pt24B (4P2)	NOTCH1	chr9:139390716	G	T	p.S2492X	NM_017617	stopgainSNV	0.18	350	1998
Pt24 R2	TP53	chr17:7578464	CGGGTGCCGGGCGG	C	p.152_156del	NM_000546	frameshiftdeletion	1.00	1971	1971
Pt24 R3	TP53	chr17:7578464	CGGGTGCCGGGCGG	C	p.152_156del	NM_000546	frameshiftdeletion	0.98	1932	1971
Pt24B (4P1)	TP53	chr17:7578464	GGGTGCCGGGCGG	-	p.152_156del	NM_000546	frameshiftdeletion	0.30	578	1938
Pt24B (4P2)	TP53	chr17:7578464	GGGTGCCGGGCGG	-	p.152_156del	NM_000546	frameshiftdeletion	0.88	1702	1941
Pt24B (5P)	TP53	chr17:7578464	GGGTGCCGGGCGG	-	p.152_156del	NM_000546	frameshiftdeletion	0.94	1831	1952
Pt24P (3P)	TP53	chr17:7578464	CGGGTGCCGGGCGG	C	p.152_156del	NM_000546	frameshiftdeletion	0.51	1003	1958
Pt24B (4P2)	TP53	chr17:7578203	C	A	p.V216L	NM_000546	nonsynonymousSNV	0.36	286	784
Pt24B (5P)	TET2	chr4:106196820	G	T	p.V1718L	NM_001127208	nonsynonymousSNV	0.90	1207	1334
Pt24B (4P1)	TET2	chr4:106196820	G	T	p.V1718L	NM_001127208	nonsynonymousSNV	0.53	1303	2449
Pt24B (4P2)	TET2	chr4:106196820	G	T	p.V1718L	NM_001127208	nonsynonymousSNV	0.70	346	497
Pt25 R4	TP53	chr17:7578395	T	C	p.H179R	NM_000546	nonsynonymousSNV	0.99	3084	3113
Pt25 (3P)	TP53	chr17:7578395	T	C	p.H179R	NM_000546	nonsynonymousSNV	0.61	1057	1723
Pt28 (2P)	TP53	chr17:7577547	C	A	p.G245V	NM_000546	nonsynonymousSNV	1.00	1987	1987
Pt28 R3	PIK3CA	chr3:178936083	G	A	p.E542K	NM_006218	nonsynonymousSNV	0.20	86	437
Pt28 (2P)	PIK3CA	chr3:178936083	G	A	p.E542K	NM_006218	nonsynonymousSNV	0.55	296	539
Pt28 R3	TP53	chr17:7577547	C	A	p.G245V	NM_000546	nonsynonymousSNV	1.00	1987	1987
Pt30 R1	PIK3CA	chr3:178952085	A	G	p.H1047R	NM_006218	nonsynonymousSNV	0.52	330	636
Pt30 R2	PIK3CA	chr3:178952085	A	G	p.H1047R	NM_006218	nonsynonymousSNV	0.37	212	567
Pt30 R3	PIK3CA	chr3:178952085	A	G	p.H1047R	NM_006218	nonsynonymousSNV	0.54	360	661
Pt30 R5	PIK3CA	chr3:178952085	A	G	p.H1047R	NM_006218	nonsynonymousSNV	0.65	589	900
Pt30 R10	PIK3CA	chr3:178952085	A	G	p.H1047R	NM_006218	nonsynonymousSNV	0.49	311	635
Pt30 (4P)	PIK3CA	chr3:178952085	A	G	p.H1047R	NM_006218	nonsynonymousSNV	0.48	260	536
Pt30 R1	ESR1	chr6:152419926	A	G	p.D538G	NM_000125	nonsynonymousSNV	1.00	2829	2843
Pt30 R2	ESR1	chr6:152419926	A	G	p.D538G	NM_000125	nonsynonymousSNV	0.29	1050	3634
Pt30 R3	ESR1	chr6:152419926	A	G	p.D538G	NM_000125	nonsynonymousSNV	1.00	2589	2599
Pt30 R5	ESR1	chr6:152419926	A	G	p.D538G	NM_000125	nonsynonymousSNV	1.00	3137	3151
Pt30 (4P)	ESR1	chr6:152419926	A	G	p.D538G	NM_000125	nonsynonymousSNV	1.00	2550	2562

Table 6. **NGS-identified prioritized mutations in individual and pooled CTC samples.** All high confidence, non-synonymous variants identified in CTC samples are shown. For each variant, the gene, location (hg19), reference and variant (Alternate) alleles, amino acid (AA) change, RefSeq ID, variant type, variantiant read frequency (flow corrected variant over total reads) and flow-corrected variant and total reads separately are shown. See supplementary methods for variant filtering and prioritization. Each variant location was visually checked in IGV (Integrated Genome Viewer) in all samples from each patient who harbored at least one sample with that variant in order to confirm adequate coverage and quality.

Of 23 point mutations and short indels detected in CTC across all patients, 14 (61%) were also found in the WES of corresponding tissue biopsies, which additionally harbored another 9

mutations/indels that were assayed, but not detected by the targeted panel in any of the corresponding CTC (**Table 7A**).

A. CTC vs. tissue mutation concordance

		Tissue			
		Y	N	Total	
CTC	Y	14	9	23	61%
	N	9	0	9	
	Total	23	9	32	
		61%			

B. CTC vs. tissue CNA concordance

		Tissue			
		Y	N	Total	
CTC	Y	19	12	31	61%
	N	7	0	7	
	Total	26	12	38	
		73%			

Table 7. 2 x 2 contingency concordance tables for alterations in CTC vs tissue. Legend: Cohort-wide numbers of alterations for **A.** Mutations and **B.** Copy number alterations (CNA) that were detected by either method, showing numbers and percentages of alterations concordant in either direction. Y= Yes (detected), N= No (not detected).

Of note, the fraction of sequencing reads containing the variant (variant frequency, VF) in individual CTC was in the vast majority of cases either 1.0 or approximately 0.5, consistent with homo- or heterozygous status of mutations in individual cells, and was highly concordant with tumor content-corrected VFs in tissue samples (**Figure 4A; B**).

Although previous studies have assessed mutations *or* CNAs from CellSearch® isolated CTC and other fixative based CTC platforms, we are unaware of simultaneous assessment of both categories of somatic alterations in fixed CTC. We have previously validated the ability of our multiplexed PCR based NGS approach to assess gene-level CNAs in fresh tissue, FFPE tissues and cfDNA [101, 102, 106, 114]. Herein, we adapted our approach to include only high performance amplicons in WGA CTC genomic DNA allowing us to assess CNAs (both high-level amplifications and deletions) in 71 robustly assessed OCP target genes (See **Materials and Methods**). Critically, we observed high concordance in prioritized CNAs detected in CTC and matched tissue biopsies across our cohort. Specifically, 19 of 31 (61%) total CNAs detected in CTC were also present in fresh tissues subjected to WES, while WES identified an additional 7 alterations assessable but not detected in matched CTC (**Table 7B**). Furthermore, similarly to mutations, the approximate copy number of concordant CNAs was highly consistent between CTC and matched tissue biopsies, as shown in **Figure 4A; B** and **Figure 5A; B**. For example, the tissue metastasis from patient #12 harbored prioritized *PIK3CA* p.H1047R and *TP53* p.R248Q mutations, as well as *WT1, TSC2, MYC* and *NF1* amplifications. The single high quality CTC available from this patient similarly harbored both mutations, as well as the *WT1, TSC2* and *MYC* amplifications, with *WT1* showing the greatest estimated copy number in both the tissue and CTC samples (**Figure 4B; 5B**). Further supporting our technique, patient #2 CTC CNA data showed no batch effects by unsupervised clustering despite being processed from three different CellSearch® cartridges (A; B; C) stored for varying amounts of time before single cell isolation (**Figure 5A; Table 1**).

Discordant genomic alterations between CTC and corresponding tissue metastases were found in several patients, even if they only had a few high quality CTC. For example, patients

#7, 14, 17, and 19, had complete discordance in genomic alterations between CTC and tumor tissues (**Figure 4B; 5B**). Of note, potentially targetable *BRCA2* p.Q1931X and *PTCH1* p.E1242X mutations were found exclusively in CTC (and not tissues) from patients #17 and 19, respectively. Whether such events represent bona-fide mutations or WGA/NGS technical artifacts cannot be reliably determined. To further investigate these possibilities, we performed Sanger Sequencing for the *BRCA2* p.Q1931X mutation, which confirmed its presence at low level in that pooled sample made up of four CTCs, consistent with the subclonal VF (0.10) in that pooled NGS sample (**Figure 6A**). Sanger sequencing of these four individual CTCs confirmed the mutation at heterozygous level in one of these four CTCs pooled in the NGS sample.

A

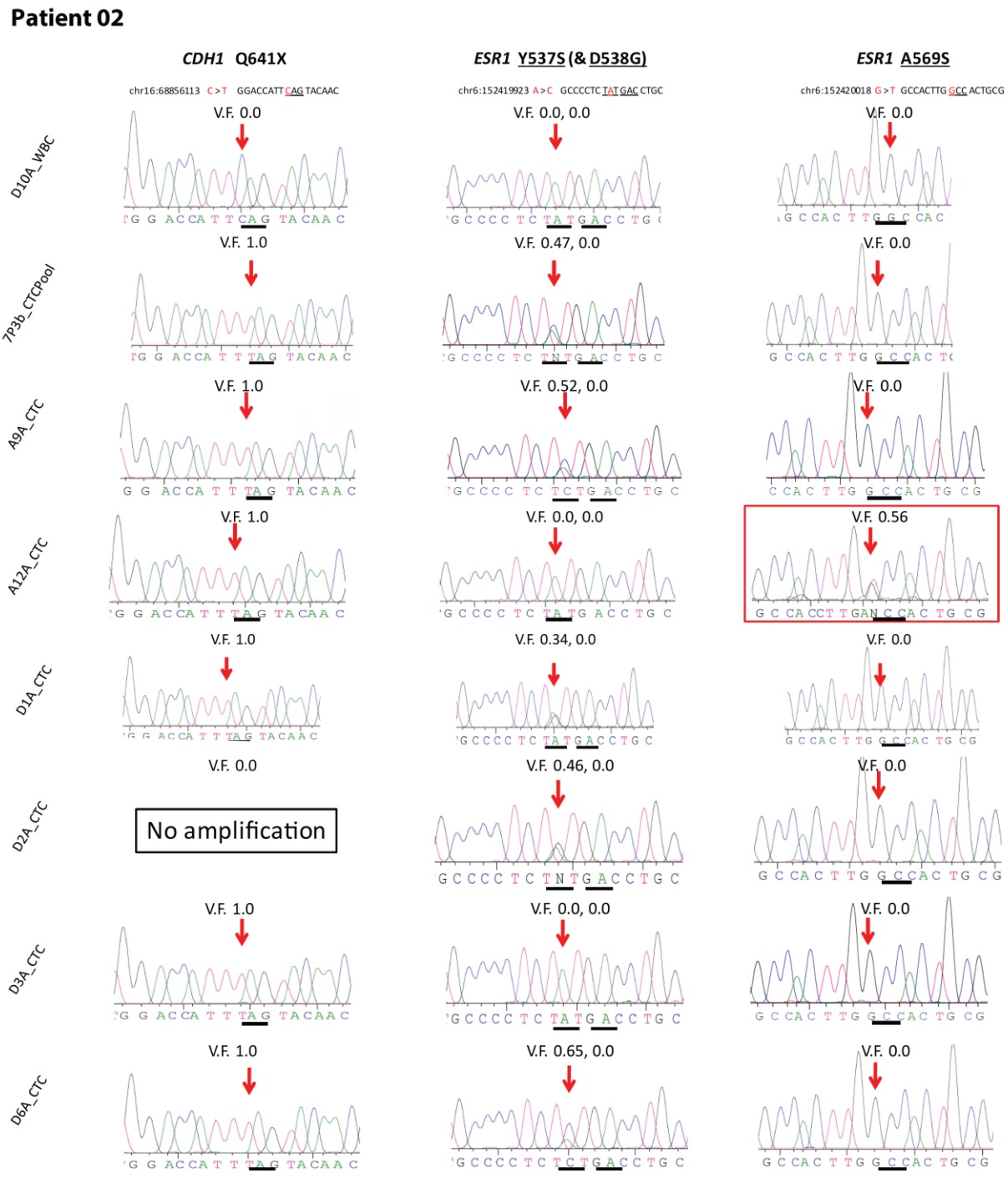
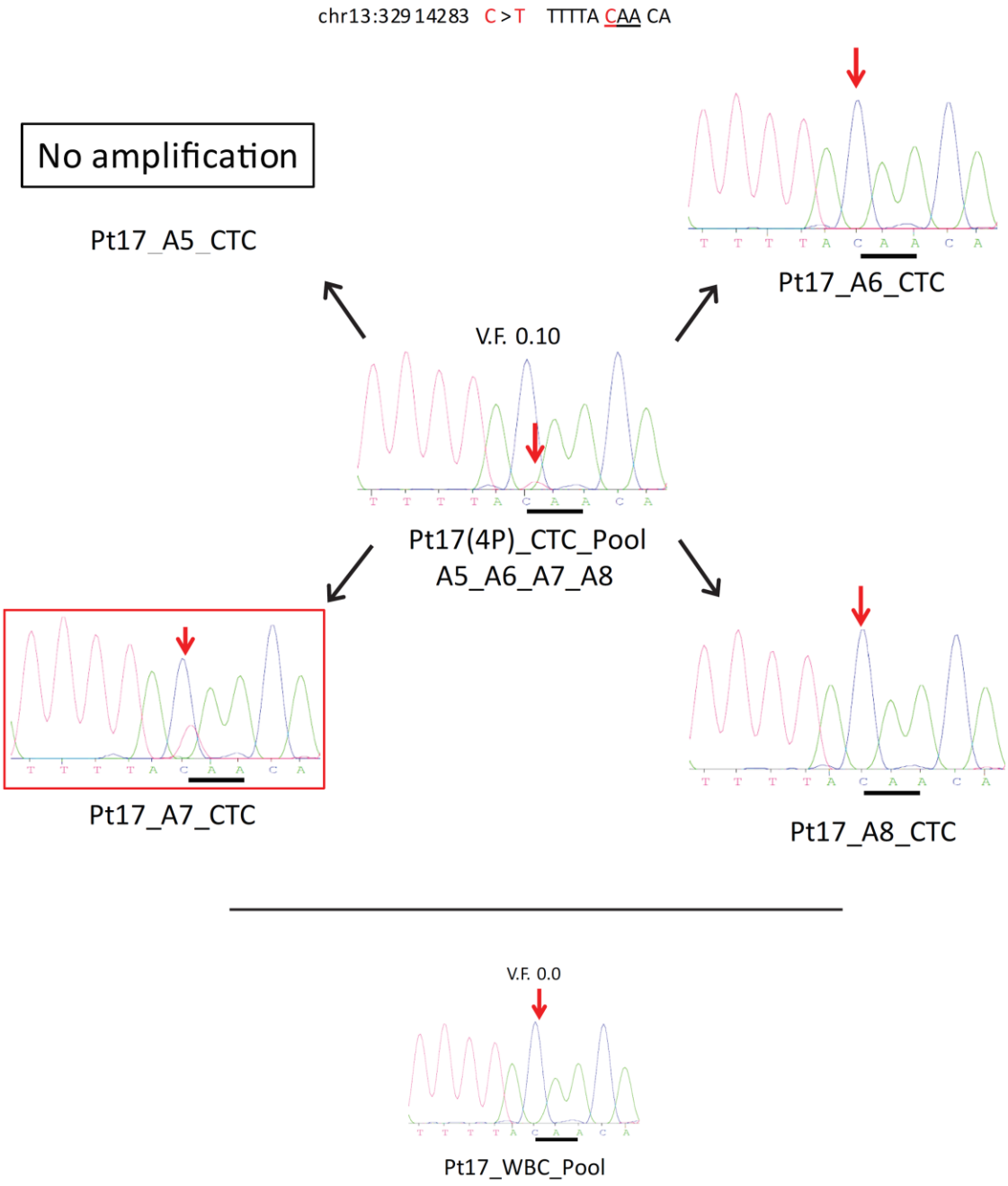


Figure continued below

Patient 17

BRCA2 Q1931X



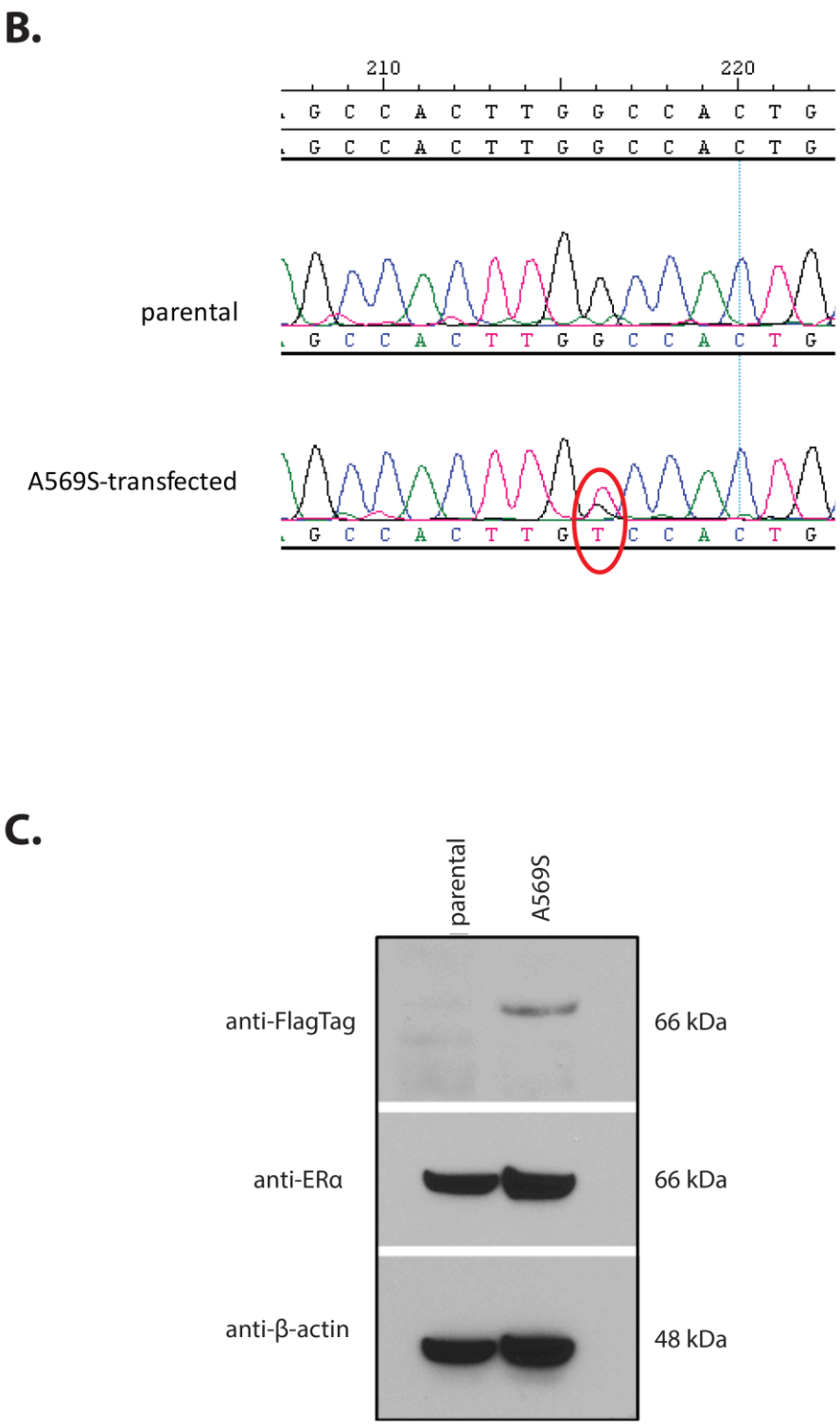


Figure 6. Sanger sequencing of WGA CTC DNA was fully concordant with NGS sequencing. A. Sanger sequencing of for CDH1 (Q641X), ESR1 (Y537S, D538G, A569S) and

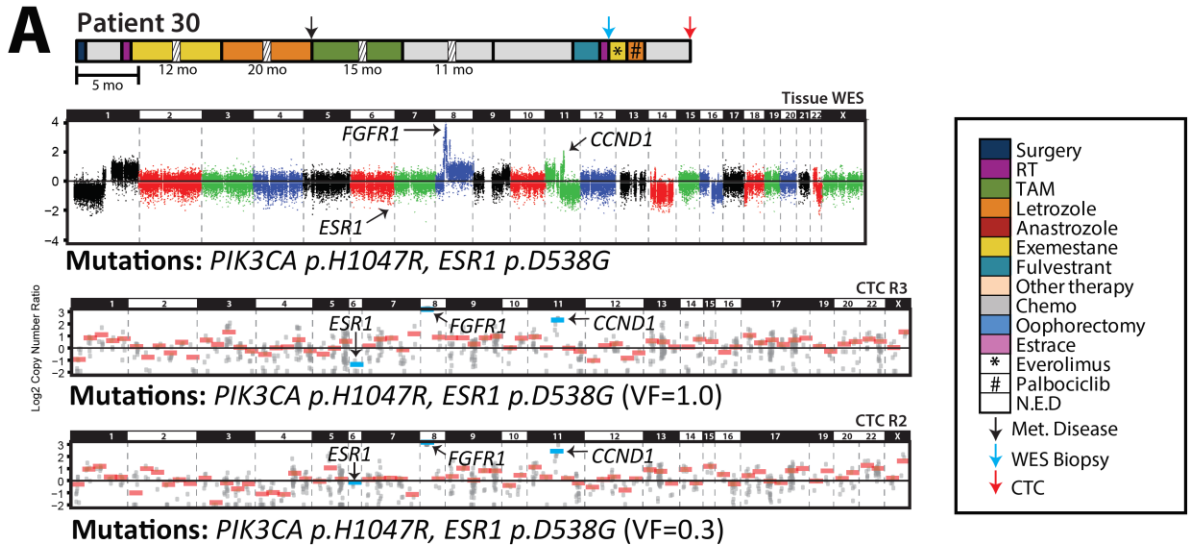
BRCA2 (Q1931X). NGS variant fraction (V.F.) shown when performed. Codon of affected aminoacid is underlined. Red rectangles denote Pt #2 single CTC with novel ESR1 mutation and Pt #17 single CTC with BRCA2 mutation. Only forward Sanger traces are shown, reverse traces were all in concordance (not shown); **B.** Sanger sequencing results of cDNA from parental and A569S over-expressing MCF-7 cells; **C.** Western blot confirming the presence of A569S mutation in overexpressed cells compared to Wild Type.

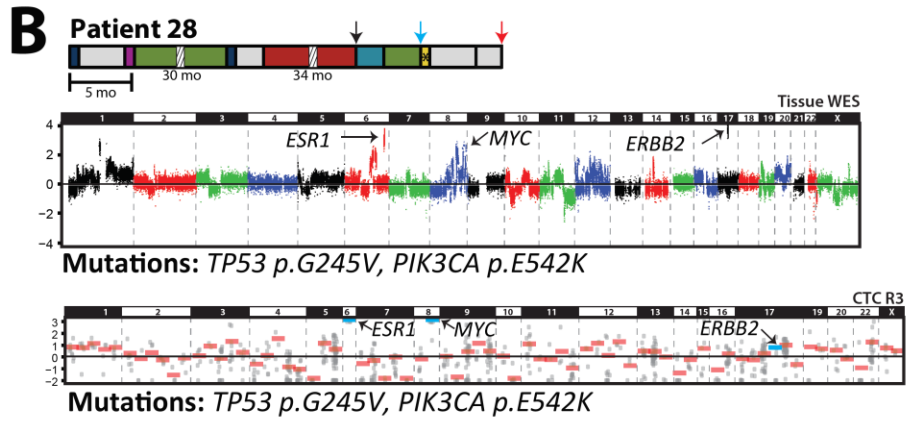
Importantly, however, NGS of this CTC pool did not show the *PIK3CA* (p.H1047R) or the *NF1* (p.W2494X) mutations clearly present in this patient's tissue, suggesting that these cells are not of the same subclonal origin as the biopsied tumor tissue region that underwent WES. Similarly, none of the CTC from patient #17 harbored the *PIK3CA* (p.E542K) and *ESR1* (p.Y537S) hotspot mutations present in tissue (**Figure 4B**). In addition, CTC and tissue from these patients harbored several discordant copy number changes, again suggesting that these CTC were from subclones that were entirely different from that of the biopsied tissue (**Figure 4B and 5B**).

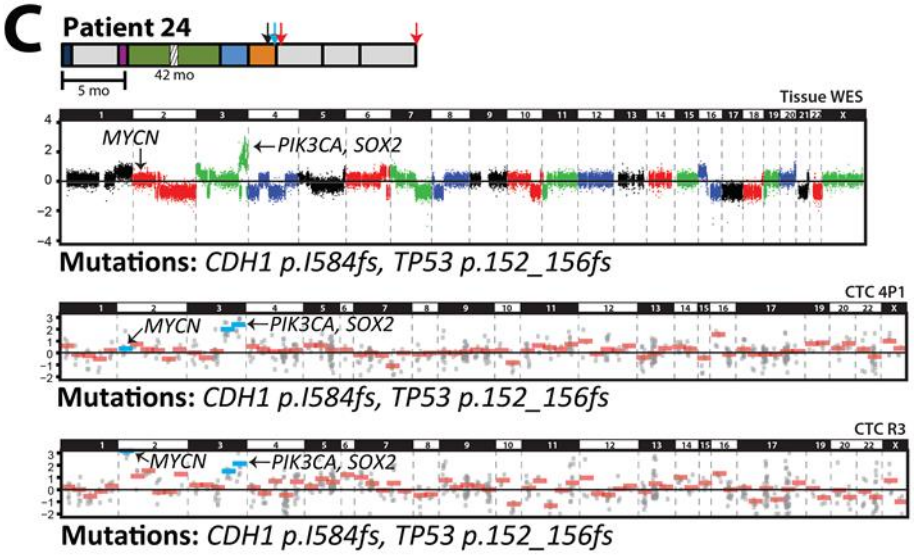
2.4.5 Integrative mutation and CNA assessment of resistance mechanisms and clinically relevant intratumoral heterogeneity in CTC from individual patients

Our ability to assess mutations and CNAs in individual CTC from patients with matched tissue metastases profiled by WES provides a unique cohort to assess intra- and inter-patient diversity in ET resistance/progression mechanisms—including *ESR1* mutations and CNAs—as well as comprehensive genomic profiles. For example, patient #30 had HR+ MBC and developed ET-resistance between her tissue biopsy and CTC collection. In the tissue biopsy subjected to WES, we detected a heterozygous *PIK3CA* p.H1047R hotspot mutation, a homozygous *ESR1* p.D538G hotspot mutation (with LOH due to single copy *ESR1* loss), and

high level *FGFR1* and *CCND1* amplifications. Consistent with the presumed contribution of *ESR1* hotspot mutations in resistance to some, but not all, types of ET, she progressed on ET after her tissue biopsy (**Figure 3**). Critically, all informative individual (*n*=5) and pooled (*n*=1) CTC samples harbored the same heterozygous *PIK3CA* p.H1047R mutations that were detected in her tissue biopsy. Similarly, while all informative individual (*n*=4) and pooled (*n*=1) CTC samples also harbored the *ESR1* p.D538G mutation, our integrative mutation and copy number profiling allowed us to clearly distinguish CTC harboring homozygous *ESR1* mutations with LOH (*n*=3 individual and *n*=1 pooled CTC samples; VF of 1.0 and one copy *ESR1* loss) from CTC harboring *ESR1* mutation with no CNA (*n*=1 individual CTC; VF of 0.29 CTC and no *ESR1* copy alteration). Lastly, both the *FGFR1* and *CCND*1 amplifications were detected in all individual and pooled CTC at estimated copy numbers that were concordant with those in the tissue biopsy (*FGFR1>CCND1*) (**Figure 4B, Figure 5B, Figure 7A**). Taken together, these results support the ability of our single (and pooled) CTC profiling approach to identify known clinically relevant mutations and CNAs from tissue samples, as well as integrate the CTC sequencing results to identify LOH mechanisms.


A
Patient 30
5 mo
12 mo
20 mo
15 mo
11 mo
Tissue WES
FGFR1
CCND1
ESR1
Mutations: PIK3CA p.H1047R, ESR1 p.D538G
CTC R3
Log2 Copy Number Ratio
ESR1
FGFR1
CCND1
Mutations: PIK3CA p.H1047R, ESR1 p.D538G (VF=1.0)
CTC R2
ESR1
FGFR1
CCND1
Mutations: PIK3CA p.H1047R, ESR1 p.D538G (VF=0.3)
Surgery
RT
TAM
Letrozole
Anastrozole
Exemestane
Fulvestrant
Other therapy
Chemo
Oophorectomy
Estrace
* Everolimus
# Palbociclib
N.E.D
Met. Disease
WES Biopsy
CTC


B
Patient 28
5 mo
30 mo
34 mo
Tissue WES
ESR1
MYC
ERBB2
Mutations: TP53 p.G245V, PIK3CA p.E542K
CTC R3
ESR1
MYC
ERBB2
Mutations: TP53 p.G245V, PIK3CA p.E542K


C
Patient 24
5 mo
42 mo
Tissue WES
MYCN
PIK3CA, SOX2
Mutations: CDH1 p.I584fs, TP53 p.152_156fs
CTC 4P1
MYCN
PIK3CA, SOX2
Mutations: CDH1 p.I584fs, TP53 p.152_156fs
CTC R3
MYCN
PIK3CA, SOX2
Mutations: CDH1 p.I584fs, TP53 p.152_156fs

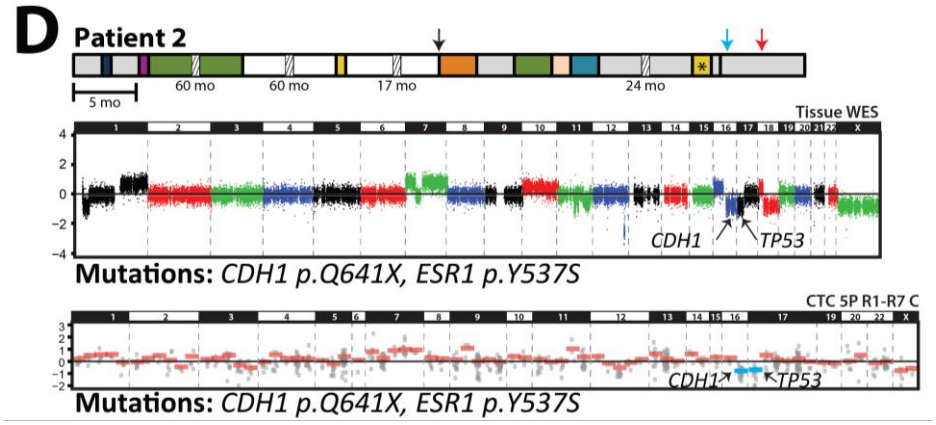


Figure 7. Integrative CNA and mutational profiling of CTCs in comparison to tissue metastases identifies intra- and inter-patient heterogeneity in resistance/progression alterations. A. Clinical time lines, copy number plots and mutations from tissue metastases subjected to whole exome sequencing (WES) and CTCs subjected to targeted NGS from **A**) Patient #30, **B**) Patient #28, **C**) Patient #24 and **D**) Patient #2. For timelines, treatment courses, development of metastasis, and research biopsy (for exome sequencing) and CTC collection time points are indicated according to the legend. Treatment courses shown with a broken box are given in months. For WES, $\log_2$ copy-number ratios per segment are plotted and prioritized mutations are given below the plot. For CTCs, each individual amplicon is represented by a single gray dot and gene-level copy-number estimates are shown in red bars (blue bars for genes of interest). Selected high-level CNAs are indicated.

Likewise, in patient #28, who had ET-resistant HR+ MBC, we detected homozygous *TP53* p.G245V and heterozygous *PIK3CA* p.E542K mutations in both CTC (one individual and one pooled sample) and in a liver tissue biopsy sequenced by WES (**Figure 4B and Figure 7B**). In addition, this patient harbored high-level *ESR1*, *MYC*, and *ERBB2* amplifications in the tissue biopsy (**Figure 5B; Figure 7B**). Importantly, while both the *ESR1* and *MYC* amplifications were clearly detected in the CTC samples, neither of the CTC samples harbored the *ERBB2* amplification (**Figure 7B**). Of note, although this patient was initially considered HER2/ERBB2 negative based on clinical immunohistochemistry (1+ expression in primary and metastatic bone lesion), the *ERBB2* amplification in the liver biopsy prompted repeat bone metastasis IHC that

showed heterogeneous HER2/ERBB2 expression, with 10-15% of cells having 3+ staining. In addition to further validating the utility of our approach, these results highlight the importance of discordant subclonal alterations, as previously reported at the mRNA level[115], that may only be present in individual metastases resulting in low levels/frequencies or absence in the circulation.

2.4.6 Comprehensive profiling of single CTC identifies potential alterations driving progressive disease

A major potential advantage of "liquid" biopsies is the ability to non-invasively monitor driving genomic alterations during disease progression. For patient #24, who had ET-resistant lobular MBC, CTC were isolated from blood specimens concurrently with the tissue biopsy (baseline) as well as at progression (465 days later) after three lines of chemotherapy (**Figure 7C**). As expected for lobular carcinoma, we identified a *CDH1* frameshift mutation (p.I584fsdel) in all informative baseline CTC (*n*=3 pooled CTC, B in **Figure 4B**) as well as the tissue biopsy. We similarly detected a *TP53* frameshift mutation (p.152_156fsdel), as well as *PIK3CA* and *SOX2* amplifications in all baseline CTC samples and the tissue biopsy. Of note, while all four of these alterations were also present in the CTC samples at progression (P in **Figure 4B, Figure 7C**), two of the individual progression CTC also harbored high-level *MYCN* amplifications (estimated at 16 copies) that were not present in any baseline specimen (CTC or tissue), demonstrating the utility of this approach to identify somatic alterations conferring treatment resistance during disease monitoring.

2.4.7 Comprehensive CTC profiling in a single patient identifies multiple ET resistance

mechanisms in circulation

Tissue and cfDNA based studies have demonstrated numerous resistance mechanisms to targeted therapies in the same patient [116], including studies in breast cancer demonstrating multiple *ESR1* mutations detectable in cfDNA from patients with ET resistant MBC [116, 117]. Hence, we performed detailed profiling of numerous CTC from patient #2, who had ET-refractory lobular breast carcinoma at the time of research biopsy (for WES) and CTC collection (189 days later) (**Figure 7D**). Tissue WES detected two E-cadherin (*CDH1*) mutations (p.Q641X and p.S70F) at homozygous VF due to single copy *CDH1* loss **(Figure 2A and Figure 7D**) as well as a heterozygous *ESR1*p.Y537S mutation, which is presumed to be one mechanism contributing to ET-resistance [48]. No high level amplifications or deletions (in genes targeted in CTC) were present, although *TP53* and chromosome (chr) X one copy losses were present by WES.

Across this patient's 32 assessed CTC samples (26 individual and 6 pools of 5-7 total individual CTC), we identified the *CDH1* p.Q641X and p.S70F mutations in 22/23 (96%) and 27/28 (96%) CTC samples informative for those genomic positions, respectively. Of note, 12/13 individual CTC informative for both *CDH1* mutations showed both at homozygous VF, consistent with tissue metastasis WES and the known early role of deleterious alterations in *CDH1* as founder events in lobular breast carcinoma. Although our CNA profiling is not optimal for single copy alterations, we observed the *CDH1*, *TP53*, and chr X losses variably across individual and pooled CTC (n= 13, 14, and 4 of 32 CTC samples, respectively). Lastly, one of this patient's purified $CK^+/CD45^-$ circulating cells was wild-type for all the tested genes, suggesting the presence of a minority of circulating epithelial cells of non-tumor origin (CTC R10C, **Figure 4A**).

As expected, in this patient, greater genomic heterogeneity was observed for the *ESR1* p.Y537S mutation, which presumably arose after ET treatment (whether through selection for a

rare pre-existing or acquired mutated clone). *ESR1* p.Y537S mutations were detected in 26/32 individual and pooled CTC samples. However, five of the six CTC harboring one or more *CDH1* mutations,(strongly suggesting these are true cancer cells), lacked the *ESR1* mutation. Likewise, both heterozygous (*n*=14/20 *ESR1* mutation-harboring individual CTC) and homozygous (*n*=6/20 *ESR1* mutation-harboring individual CTC) *ESR1* p.Y537S mutations were observed in CTC. Of particular interest, one individual CTC from this patient (A12A) harbored both homozygous *CDH1* mutations and lacked the *ESR1* p.Y537S mutation, but instead harbored a unique, previously undescribed *ESR1* p.A569S mutation at heterozygous VF (**Figure 4A**). Although the *ESR1* p.A569S mutation was not observed in any other individual or pooled CTC, or in the patient's tissue metastasis (**Figure 4A**), it, as well as *ESR1* p.Y537S and *CDH1* mutations in other CTC, were confirmed by Sanger Sequencing of WGA DNA (**Figure 6A,** referenced earlier).

Genomic profiling of tissue biopsy material and a finite number of CTC, as in our study, is still likely to underestimate the full repertoire of minor subclonal driving mutations in an individual patient with targeted therapy resistance. To further investigate this concept, droplet digital PCR (ddPCR) was performed on various specimen types at several time points during this patient's course. An additional *ESR1* p.D538G hotspot mutation was detected albeit at extremely low level. Interestingly, the novel *ESR1* p.A569S mutation detected in one CTC was also only present in a few droplets, but below the predetermined detection threshold [90] in post-ET specimens (**Table 8; Table 3,** referenced earlier**)**.

Time of sampling	Initial Diagnosis		1st clin. met. 2009		Later metastasis		Later metastasis			
Date	7/1/97		7/1/09		7/19/13		1/24/14			
Site	Breast Primary		Omental		Pleural effusion		Blood			
Type of specimen	FFPE		FFPE		cytospin		CTC		cfDNA	
Assay	ddPCR	WES	ddPCR	WES	ddPCR	WES	Targeted NGS	Sanger	ddPCR	WES
ESR1										
Y537S	WT		mut (45%)		mut (45%)	mut (44.5%)	mut 26 /32 (81%)	mut 4/6 (67%)	mut (0.87%)	
D538G	mut (0.01%)		mut (0.01%)		mut (0.02%)	WT	mut 0/32 (0%)	mut 0/6 (0%)	mut (0.01%)	
A569S	WT		WT (few droplets +**)		WT (few droplets +**)	WT	mut 1/32 (3%)	mut 1/6 (17%)	WT	

Table 8. Patient #2 analysis of primary, clinical metastatic tissue, research biopsy, pt-DNA by ddPCR. For each timepoint, specimen type and detection technique, presence/absence (and variant frequency) are shown for all three ESR1 mutations. FFPE = Formalin-Fixed Paraffin-Embedded; ddPCR = droplet digital PCR; WES = whole exome sequencing; mut = mutation present; WT = Wild Type; **a few droplets mutated for specific mutation, but below the predetermined threshold. Shaded area = assay not performed. With this approach, we detected, albeit at extremely low levels, an ESR1 p.D538G hotspot mutation that was absent from both CTC and tissue by the respective NGS techniques applied. ddPCR droplets carrying this mutation were present at VF 0.0001 – 0.0002 (0.01 – 0.02%) but consistently appeared in the primary tumor, two consecutive metastases, and later in plasma derived cfDNA collected concurrently with CTC. In concordance with CTC and tissue profiling, ddPCR of tissues and cfDNA detected the ESR1 p.Y537S mutation only in post-ET specimens. Interestingly, the novel ESR1 p.A569S mutation detected in one CTC was also only present in a few droplets, but below the predetermined detection threshold [Chu, Clin Cancer Res 2016; Beaver, Clin Cancer Res 2014] in post-ET specimens.

Taken together, these results are consistent with a large circulating pool of heterogeneous sub-clones, which may each harbor differing mechanisms of resistance to one or more types of ET in patients with initially HR positive MBC. They further support tissue, CTC and cf-DNA as complementary approaches to characterize this diversity of resistance mechanisms.

2.4.8 In vitro functional characterization of the novel ESR1 A569S mutation

As mentioned, the observed *ESR1* p.A569S mutation has not been described previously in ET resistant MBC and was only observed in 1 of 32 total CTC samples (pooled and individual) from patient #2 (**Figure 4A; Figure 8A**). Therefore, we hypothesized that it conferred a modest selective advantage since the cell in which it was detected was clearly a CTC

(it harbored both *CDH1* mutations present in the tissue metastasis and other CTC), it lacked the *ESR1* p.Y537S mutation present in 26 of 32 other CTC samples from the patient, and no other prioritized mutations were observed in this or any other CTC from the patient.

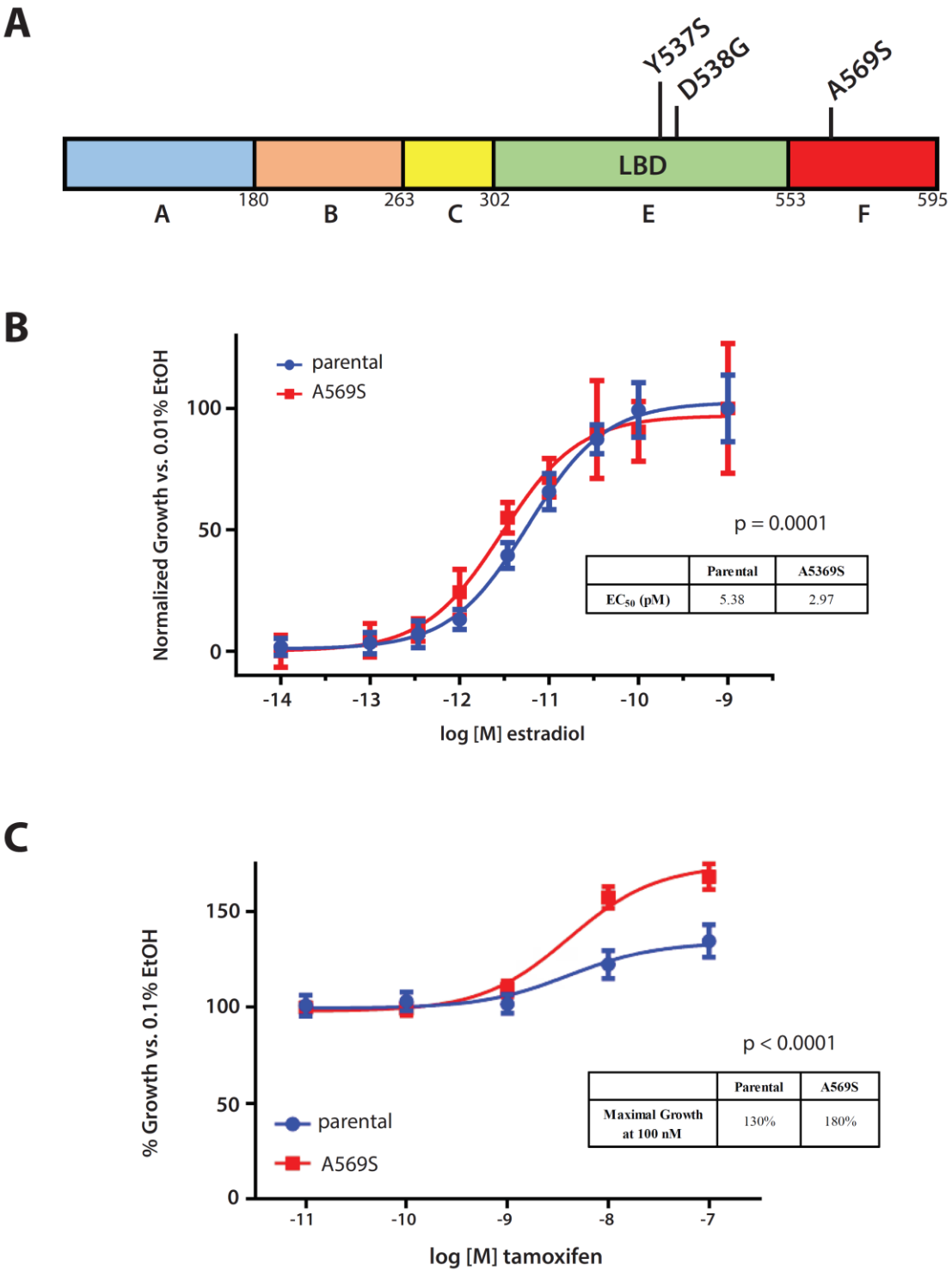


Figure 8. Functional validation of novel *ESR1* p.A569S mutation demonstrates modest estradiol sensitivity and increased tamoxifen agonist activity. A. Summary of *ESR1* mutations detected in Patient #2 from tissue, CTC-DNA, and cfDNA in schematic representation of the encoded ERα with the LBD and F domains indicated; **B.** MCF-7 *ESR1* A569S are more sensitive to

estradiol after 5 days when assessed by crystal violet assay. Inset table indicates EC_{50} of estradiol in pM; **C.** Tamoxifen has increased agonist activity in MCF-7 cells over-expressing ER-A569S. Inset table indicates maximal growth at 100nM tamoxifen. Assays were performed in triplicates with P-values and error bars (standard error) indicated.

Hence, we used lentiviral mediated infection to stably over-express the *ESR1* p.A569S mutation in the ER-positive breast cancer cell line MCF-7. Expression of the p.A569S mutation was confirmed by Sanger Sequencing and western blot analysis (**Figure 6B**; **C)**. *In vitro* growth assays, which compared MCF-7 *ESR1* p.A569S with parental MCF-7, showed that the former was estrogen dependent, unlike cells expressing *ESR1* p.Y537S or p.D538G mutations, which confer estrogen-independent growth (**Figure 9**). Notably, however, MCF-7 *ESR1* p.A569S were more responsive to estradiol induced growth compared to parental MCF-7 (EC_{50} of 2.97 pM vs 5.38 pM; p=0.0001, F-test), (**Figure 8B**).

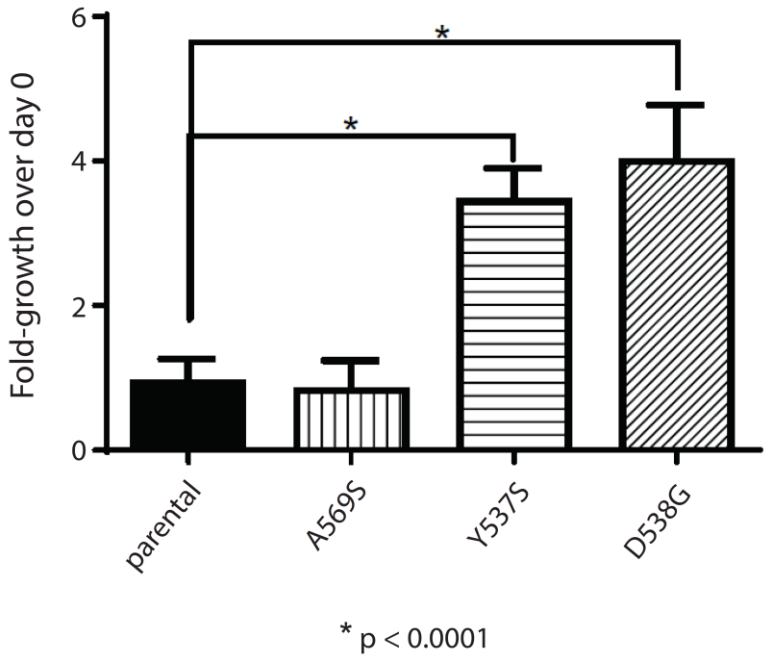


Figure 9. MCF-7 (ESR1 A569S) are not estrogen independent over 5 days in hormone-free conditions when assessed by crystal violet assay. Cells were depleted of exogenous steroids and treated with 17β-estradiol or ethanol control in triplicates. Cell number was assessed by crystal violet stain five

days after treatment.

Tamoxifen is a partial ER agonist. Therefore, we tested its ability to stimulate growth in parental MCF-7 and MCF-7 *ESR1* p.A569S cells in the absence of estradiol. Although tamoxifen stimulated growth in both MCF-7 and MCF-7 *ESR1* p.A569S cells in the absence of estradiol, MCF-7 *ESR1* p.A569S cells showed a significantly greater growth increase than parental MCF-7 cells (130% vs. 180% over vehicle control, p<0.0001, F-test, **Figure 8C**). As expected, the tamoxifen metabolites 4-hydroxytamoxifen and endoxifen (4-Hydroxy-N-desmethyltamoxifen), both potent ER antagonists, did not stimulate growth in either parental MCF-7 or MCF-7 *ESR1* p.A569S cells in the absence of estradiol (**Figure 10A; B**).

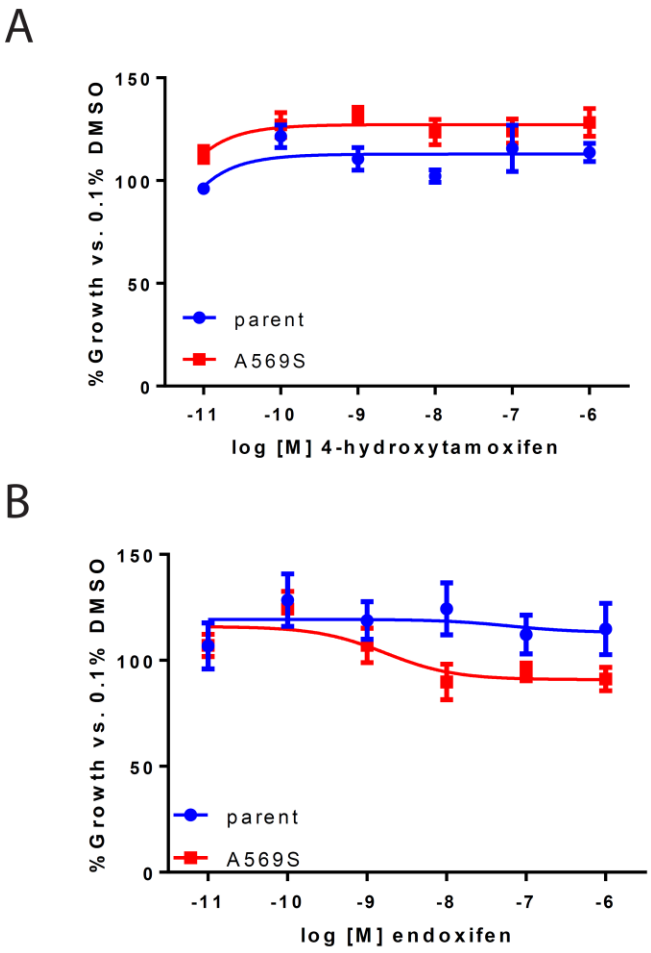


Figure 10. The tamoxifen metabolites 4-hydroxytamoxifen and endoxifen have no increased agonistic effect in MCF-7 cells over-expressing ER-A569S.

Lastly, the antiestrogens tamoxifen, 4-hydroxytamoxifen, endoxifen, and fulvestrant were all able to similarly attenuate growth induced by 50 pM estradiol in both parental MCF-7 and MCF-7 *ESR1* p.A569S (**Figure 11A; B; C; D**, respectively). Taken together, these results demonstrate that this *ESR1* mutation showed increased estradiol and tamoxifen induced growth. Presumably, this mutation provides a modest selective growth advantage, consistent with the rarity of this mutation in circulation (1/32 CTC samples) compared to the *ESR1* p.Y537S mutation known to confer estrogen independent growth and present at higher frequency in the patient's CTC (26/32 samples) and tissue metastasis.

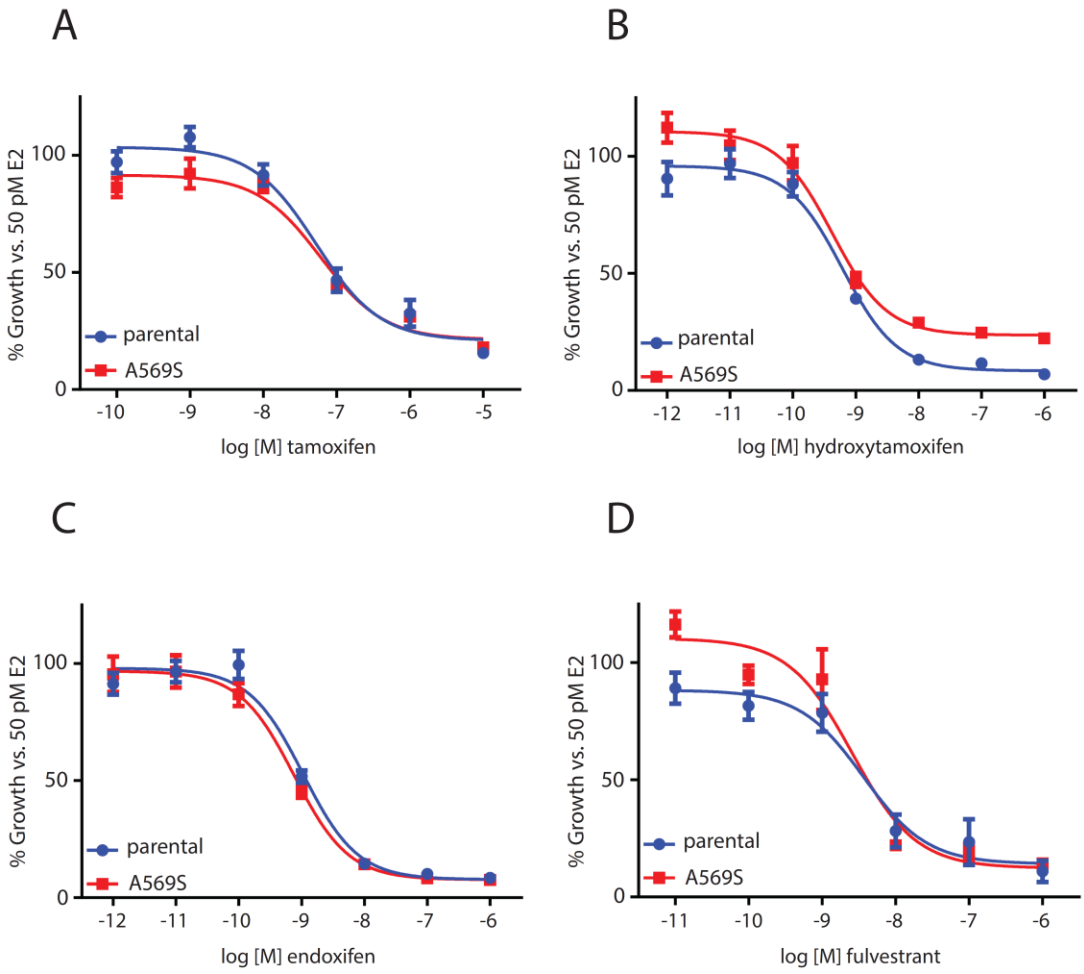


Figure 11. Inhibition of estradiol-stimulated MCF-7 parental or A569Soverexpressing cells. Cells were withdrawn from estrogen and treated with 50pM estradiol in combination with increasing concentrations of **A**. tamoxifen, **B**. 4-hydroxytamoxifen, **C**. endoxifen or **D**. fulvestrant, in triplicates. Cell growth was assessed six days after treatment by crystal violet staining and plotted as %-growth vs. vehicle control.

2.5 Discussion

2.5.1 Overall review of findings

In these proof-of-concept studies, we have demonstrated that individual CTC from archived cartridges of the FDA cleared CellSearch® system can be isolated, whole genome amplified, and comprehensively profiled for somatic mutations and CNAs using multiplexed

PCR-based NGS. Importantly, although comprehensive genomic profiling of single non-fixed CTC has been reported [118], and CTC isolated by CellSearch® (or other fixative based approaches) have been used for assessment of genome wide CNA profiling and mutations in a limited set of genes [92, 97, 119, 120], we are unaware of prior concurrent somatic mutation and precise gene-level CNA assessment in individual fixed CTC from NGS. Leveraging near synchronous matched tissue biopsies subjected to WES as part of the MI-ONCOSEQ clinical trial, we demonstrated high concordance between driving somatic mutations and CNAs identified by our approach in CTC compared to single tissue metastases, including CNAs, mutations, and LOH assessment in both oncogenes (e.g. *ESR1* and *PIK3CA*) and tumor suppressors (e.g. *CDH1* and *TP53*).

In addition to the high concordance between prioritized alterations identified in CTC and matched tissue metastases, several lines of evidence support the validity of our findings. Process-matched WBC assessed in parallel for each of the patients were wild-type for all assessed genomic alterations. Likewise, at the patient level, in addition to high concordance for somatic mutation and CNA presence/absence in CTC vs. matched tissue metastases, VFs (for mutations) and estimated copy number ratios (for CNAs) were similarly concordant. Lastly, Sanger Sequencing of WGA CTC DNA validated the selected mutations tested in two patients, one of which had highly heterogeneous findings among tissue, CTC and cfDNA.

Several patients had somatic alterations present in tissue metastases that were not present in all matched CTC such as *ESR1* p.Y537S mutation for patient #2. Likewise, in patient #28, the high level *ERBB2* amplification present in the liver metastasis was not identified in the CTC. In fact, only ~15% of tumor cells in a femur metastasis had 3+ HER2 expression by IHC, consistent with *ERBB2* amplification being a subclonal event within an individual tissue metastasis as well

as in tissue vs. circulation.

A purported advantage of liquid biopsies is the theoretical ability to capture the global collection of molecular alterations harbored by a patient with metastatic cancer. In our study, we were able to simultaneously assess mutations and copy number in individual cells. Thus, we observed numerous examples of intra-patient somatic alteration variability between individual CTC and tissue metastases, including assessment of heterozygosity/homozygosity supported by copy number state. For example, although all CTC from patient #30 harbored *PIK3CA* p.H1047R mutations, we detected individual CTC that harbored either heterozygous *or* homozygous (with one copy loss) *ESR1* p.D538G mutations. Similarly, we observed several CTC that harbored somatic mutations and CNAs not present in tissue, but also detected alterations in tissue metastases not observed in CTC. Importantly, we identified potentially targetable alterations (in *PTCH1* and *NOTCH1*) in subsets of CTC from two patients (#19 and #24) but not in matched tissue, a finding that supports these as subclonal alterations. In this case, knowledge of the subclonal populations might enable more specific, and even tailored combination targeted therapy, either at initiation of treatment or for emerging subclones during cancer progression short of any deep sequencing of primary/metastatic tissue. The latter has shown that most discordant "private" CTC alterations can be detected in the tissue at low frequencies [97].

As an additional mechanism of discordance between CTC and tissue metastases, patient #4 harbored *PIK3CA* p.E545K and *TP53* p.D259H mutations in both CTC and tissue metastases. Importantly, however, while both these mutations were heterozygous in the tissue metastasis, we observed both as homozygous in a single CTC (without evidence of copy loss), but wild-type in the other two cells. Lastly, in patient #24, while CTC and the concurrent tissue metastasis shared

several alterations, CTC taken subsequently after progression on several lines of chemotherapy not only harbored these alterations, but uniquely had high level *MYCN* amplifications. These findings further suggest that serial "liquid profiling" to monitor molecular alterations mediating resistance might permit specific selection of combination targeted therapies during a patient's clinical course.

Emerging genomic tumor heterogeneity due to accumulated subclones with different mutations, as well as more plastic epigenetic/transcriptomic heterogeneity as a function of cellular stress and environmental changes, are drivers of resistance to specific treatments. Indeed, we and others have demonstrated inter- and intra-patient heterogeneity of CTC-protein expression, including ER, BCL2, HER2, Ki-67, and PI3K from patients with MBC [98, 121-124]. These data, coupled with the results of the present report, demonstrate similar heterogeneity in both somatic mutations and CNAs through simultaneous profiling, and suggest that both genetic and protein data should be monitored to truly tailor precision therapy.

2.5.2 Circulatory non-tumoral epithelial cells

Of note, 3 out of 46 individual CTC included in our cohort matched the immunohistochemical definition of CTC (CK^+, $CD45^-$), but did not contain any high confidence mutations, indels or high-level CNAs. These cells were present in patients where other CTC were concordant with tissue metastases. These observations suggest that benign, non-hematopoietic circulating cells of epithelial origin may be captured by platforms using anti-epithelial enrichment/purification methods. Because the CellSearch® enrichment method is based on epithelial cell capture by EpCAM expression and characterized by the presence of CK

and lack of CD45 expression, it is possible that these cells represent normal circulating epithelial cells (CEC) that were in the patient's blood either as a function of shedding or during the blood draw in patients without cancer [125, 126]. Likewise, we confirmed (by Sanger Sequencing) the presence of a subset of mutations that were detected exclusively in several CTC and not matched tissue samples. However, due to the complete discordance between alterations in these CTC (some of which were pooled) and tissue samples, we cannot exclude that these cells are either of non-tumor origin or that these findings may be technical artifacts. These findings emphasize the need for CTC platforms to document that detected CTC are malignant cells through the identification of pathognomonic molecular alterations or orthogonal tissue based validation. We cannot be certain that CTC captured by the CellSearch assay are the source of tissue metastases. Since CellSearch requires fixation of the captured cells, culture of those cells is not possible. Other investigators have demonstrated presumed CTC (CK+/DAPI+/CD45-) captured using different platforms can, indeed, be cultured *ex vivo* and *in vivo* [127, 128]. Importantly, although numerous CTC platforms report the ability to detect CTC "missed" by capture and/or expression based CTC platforms[77], the CellSearch® platform has been clinically validated for its prognostic role in breast colon, prostate, and lung cancer [80, 81, 129]. Combined with the high concordance of known oncogenic genomic alterations between CellSearch- identified CTCs and tissue metastases, our results support these CTCs as having malignant potential, at least in part. In addition, our results herein support the vast majority of identified CTC as bona-fide tumor cells based on somatic molecular alteration concordance with tissue metastases.

2.5.3 Heterogeneity of endocrine therapy resistance mechanisms in CTCs

In an effort to comprehensively profile the CTC genomic landscape in an ET resistant patient, we profiled 32 individual and pooled CTC samples from patient #2, who had ET resistant metastatic lobular breast cancer. Both CTC and tissue metastases showed concordant *CDH1* mutations, as expected. However, while 20 of 26 single CTC harbored *ESR1* p.Y537S mutations consistent with that detected in the tissue metastasis, a single CTC instead harbored a novel unreported *ESR1* p.A569S mutation confirmed by Sanger Sequencing of the amplified genomic DNA. Intriguingly, p.A569 is localized to the F domain of the ER receptor, which differs from the more common mutations, such as p.Y537S and p.D538G, identified in the LBD (**Figure 8A**). LBD missense mutations lead to ligand-independent constitutive ER activation [48, 85-87, 130] and confer relative resistance to tamoxifen and fulvestrant. In contrast, *ESR1* p.A569S expressing MCF-7 cells were estrogen dependent and not tamoxifen nor fulvestrant resistant. However, the *ESR1* p.A569S mutation conferred a modest, but statistically significant, increased agonist activity to tamoxifen in the absence of estrogen in MCF-7 cells compared to wild-type *ESR1*-expressing cells. Interestingly, a mutation in the adjacent aminoacid (p.T570I) was recently reported (without functional characterization) in a CTC of an ET-treated MBC patient [94]. In addition, a mutation in the same domain in another steroid hormone receptor family member, the androgen receptor, (p.F876L), in prostate cancer cell lines confers agonist activity to the normally androgen antagonist enzalutamide [53]. It is well established that in breast, tamoxifen is not a pure ER antagonist, but rather serves as a selective estrogen receptor modulator, with mixed ER agonism and antagonism, including in estrogen-deprived ER positive cultured human breast cancer cells. Multiple mechanisms of this duality in tamoxifen effect have been proposed, including up-regulation of a variety of other genes such as *NFķB* and *HER-2* [54, 55]. Of note, the F, domain which harbors our novel mutation, was shown to be important in the

agonist and antagonist balance of antiestrogens [131]. We speculate that the *ESR1* p.A569S mutation may also contribute to this paradoxical effect of tamoxifen on ER.

Notably, this modest activity of the *ESR1* p.A569S mutation is consistent with the rarity of this mutation amongst the patient's CTC burden compared to the highly active p.Y537S. However, given the interest in developing therapeutic approaches for *ESR1* LBD mutations, including ongoing clinical trials based on *ESR1* LBD mutation detection (e.g. NCT03079011), our data demonstrate that such patients likely have rare tumor subclones potentially capable of expanding upon selective pressure. Of note, ddPCR profiling of previous tissue samples and cfDNA demonstrated even more complexity in regards to *ESR1* status in this patient. Taken together, these results highlight the diverse repertoire of ET resistance mechanisms in patients with advanced, endocrine treated ER positive breast cancer, analogous to those observed in rapid autopsy series of hormonally treated prostate cancer [132].

2.5.4 Novel *ESR1* mutation is not a technical artifact

We cannot entirely exclude that the observed *ESR1* (p.A569S) mutation may be a technical artifact from CTC fixation, WGA or NGS [133]. However, this unique mutation was found at uniallelic variant fraction of 0.56 in a cell containing the same *CDH1* mutations observed in other CTCs and the tissue metastasis. Likewise detection by Sanger sequencing in pre-NGS WGA DNA strongly argues against this being an NGS error. In addition, the fact that the other most common ESR1 mutation (Y537S) is not present in this CTC underlines that these mutations are mutually exclusive. Furthermore, multiple circulating *ESR1* mutations in the same patient are well described in metastatic breast cancer, with 17% of patients having >2 detectable

ESR1 LBD mutations in cfDNA[134]. Lastly, the modest, but observable functional activity of the mutant support this mutation as a bona-fide mutation present as a minor circulating subclone rather than a technical artifact.

In more detail, this cell appears to be a true CTC as both founder *CDH1* mutations present in the matching tissue were also present in this CTC in 100% of the sequencing reads. In addition, this cell did not harbor the Y537S hotspot mutation which is the most prevalent *ESR1* mutation in this patient (26 of 32 CTC samples), which suggested to us the possibility that this could be a bona fide mutation with some functional activity allowing this cell to be resistant to endocrine therapy.

While it is true that the Ion Torrent sequencing platform can have a slightly more elevated error profile compared to Illumina sequencing, this is mainly limited to the introduction of erroneous indels in homopolymer regions. All our data were filtered to exclude indels in homopolyers of >4 bases long and this mutation is a G > T single nucleotide substitution variant in a stretch of only 2 guanines and thus not a homopolymer indel. And obviously, the fact that we also detected this (and all other mutations tested) by Sanger Sequencing of the post-WGA DNA at trace peak heights consistent with the Ion Torrent variant frequency greatly increases our confidence that this mutation is at least present in the whole genome amplified DNA and not an Ion Torrent artifact.

However, we have reason to believe that this mutation was also present in the original pre-amplified DNA. Firstly, the mutation was present in 56% of the reads covering that position, consistent with heterozygosity (50%) +/- some degree of error, also supported by Sanger sequencing peak heights. If the mutation were to be erroneously introduced during the whole genome amplification step prior to sequencing, it would have a wide distribution of probability

of the final variant frequency (0 – 100%) depending on how early in the PCR cycles it is introduced. It is further reassuring that the Ampli1 PCR-based WGA kit that was used on our CTC DNA, is a WGA platform with the best error profile among the ones tested. Another potential point of acquisition of the DNA error could be in the pre-amplified DNA at the cell preservation and fixation step as the damaging activity of cross-linking fixatives on DNA is well recognized. While it is true that our cell preservation method (CellSave) uses fixative substances that are potentially damage-inducing, we have extensive experience with Ion Torrent sequencing of formalin fixed tissue. We have for example shown 100% concordance of mutation detection in a cohort of over 100 FFPE tumor samples orthogonally tested in a clinical molecular diagnostic lab among other cohorts, (although DNA was obviously extracted by different methods that in the current project). Furthermore, as we report, we did not observe a high level of erroneous or discordant mutations in this cell, patient or other patients compared to tissue.

While we agree that the absence of this mutation in the biopsied matched tissue casts doubt on our claim, this result is perhaps not surprising considering the extensive intrapatient heterogeneity we observed in this patient. After all, tissue analysis is only limited to the material obtained in the biopsy needle and CTC NGS and functional data suggested this mutation had very low prevalence in this patient. Furthermore, as we reported, the mutation was present by ddPCR in both omental and pleural effusions from this patient in a few droplets in each, (although this was below our pre-determined threshold of positivity), suggesting the presence of this mutation in the patient.

2.5.5 Novel *ESR1* mutation functional activity

Finally, the activity of our *ESR1* p.A569S mutation in vitro is modest and unlike p.Y537S or p.D538G mutations, is not consistent with true estrogen independence suggesting that this is at best a passenger mutation if not entirely an artifact. However, this mutation is present in the F domain which is C-terminal to the ligand binding domain that harbors the p.Y537 and p.D538 aminoacids. As we point out, the F domain has been shown to be involved with regulation of agonist/antagonist balance of estrogen receptor modulators such as tamoxifen (used in this patient) perhaps though involvement of receptor co-activators and co-inhibitors. Therefore this mutation would confer to this cell a more nuanced and milder selective advantage which is consistent with its low prevalence in circulation (1 of 32 CTC samples). Importantly, there are recent reports in metastatic breast cancer of one patient with a mutation in the *ESR1* aminoacid adjacent to ours (p.T570I) in only one CTC (although without functional characterization). To our knowledge, biologic/functional characterization of our mutation or of any in the vicinity has not been reported.

In conclusion, we are unable to absolutely exclude the possibility of the artefactual nature of this mutation. However, given the reasoning and evidence presented above, and in absence of evidence to the contrary, the speculation that this mutation was present and had a functional significance in this cell appears to be the most parsimonious explanation.

2.5.6 Limitations

Our study is limited due to relatively small sample size. Indeed, the heterogeneity that we have detected has effectively reduced each patient to an "n of 1." Further, this investigation was performed as a pilot, principally to determine if DNA from CTC that have been fixed and

archived in CellSearch® cartridges for some period of time could be harvested, purified, and sequenced with analytical fidelity, and subsequently, if these data can be reliably compared to genomic analysis of tissue collected in roughly the same time frame. Importantly, our data provide proof-of-concept evidence that this strategy is viable, and studies to determine if our approach can be used to predict patient outcome or guide therapy are ongoing.

2.5.7 Overall conclusions

In conclusion, we have demonstrated the ability to purify and comprehensively sequence archived individual CTCs from an FDA-cleared CTC detection platform (CellSearch®) coupled with an automated technique for single cell purification and analysis (DEPArray™). Further, we have reproducibly performed simultaneous somatic mutation and CNA profiling of these cells. We observed high, but not absolute, concordance between somatic alterations in CTC and matched tissue metastases subjected to WES. The resultant discordance may identify potentially clinically informative/actionable alterations exclusively present either in CTC or tissue metastases, supporting the complementary nature of these approaches. Through sequencing >20 CTC in a single patient with ET resistant lobular breast carcinoma, we identified distinct *ESR1* mutations in individual CTC, including a novel, modestly active *ESR1* mutation present in only a single cell compared to the strongly activating *ESR1* mutation present in the vast majority of the patient's CTC and a tissue metastasis. Taken together, our results support the feasibility of simultaneous somatic mutation and copy number profiling from archived CTC, which may be used to track resistance mechanisms under selective pressure of targeted therapy. We speculate that these findings could lead to identification of both CTC-protein expression and genomic

abnormalities that could serve as potential therapeutic targets, and complement tissue or cfDNA based precision oncology approaches.

2.5.8 Acknowledgments

We are grateful to all the patients who generously volunteered to participate in the study. We thank the research nurses, and study coordinators for their efforts on the behalf of the patients. We would like to thank the MI-ONCOSEQ team (Michigan Oncology Sequencing Center). We would also like to acknowledge thoughtful suggestions from Dr. Benita Katzenellenbogen.

2.5.9 Funding

This work was supported by Veridex/Janssen, LLC, Menarini Silicon Biosystems, Inc., Fashion Footwear Charitable Foundation of New York/QVC Presents Shoes on Sale™ (D.F. Hayes). D.Chu and B.H.Park were supported by the Commonwealth Foundation, NIH CA194024 and the Breast Cancer Research Foundation. A.K.Cani was supported by the NIH Training Program in Translational Research T32-GM113900. S.A.Tomlins was supported by the A. Alfred Taubman Medical Research Institute.

CHAPTER III

Development of a Whole-Urine, Multiplexed, Next Generation RNA-Sequencing Assay for Aggressive Prostate Cancer Early Detection[2]

3.1 Abstract

Despite advances in biomarker development, early detection of aggressive PCa remains challenging. We previously developed a clinical-grade laboratory-developed test—MiProstate Score (MiPS)—for individualized aggressive PCa risk prediction. MiPS combines serum PSA with transcription-mediated amplification (TMA)-quantified expression of the gene-fusion *TMPRSS2:ERG* (*T2:ERG*) and the lncRNA *PCA3* from whole urine obtained after a digital rectal exam (DRE). To improve upon MiPS, in our second project, we describe the pre-clinical development and validation of a post-DRE whole urine targeted RNA NGS assay (NGS-MiPS) assessing ~90 PCa candidate transcriptomic biomarkers, including: *T2:ERG.T1E4* and *PCA3*, additional common PCa gene fusion isoforms, mRNAs, lncRNAs, and expressed mutations. NGS-MiPS showed high analytic validity and was able to detect expressed germline risk *HOXB13* and somatic driver *SPOP* mutations. In an extreme design cohort (benign or Grade Group (GG) 1 vs. GG 3-5 cancer on biopsy) NGS-MiPS showed expected differences in the levels of *T2:ERG.T1E4* and *PCA3*, as well as additional biomarkers, between benign/GG 1 vs. GG 3-5 PCa. A machine learning approach trained on a subset of the extreme design cohort

[2] This study will appear in literature as Cani, A.K., et al. (in submission).

(n=73) generated a 15-transcript model that outperformed derived MiPS and serum PSA models in predicting biopsy outcome in two validation cohorts: 1. A held-out set from the extreme design cohort (n=36); and 2. A separate PCa active surveillance cohort (n=45). These results support the potential utility and continued development of our novel urine-based targeted RNA NGS assay to improve aggressive PCa early detection.

3.2 Introduction

Prostate cancer (PCa) can follow a varied clinical course ranging from indolent to lethal metastatic disease [135], and despite recent advances in biomarker development and imaging, non-invasive early detection of PCa with aggressive potential remains challenging. Prostate biopsy and pathological grading is currently the gold standard[26]. Although serum prostate specific antigen (PSA) has revolutionized PCa early detection in the past few decades, it's performance is limited by poor specificity[136]. Recently, multiparametric magnetic resonance imaging (mpMRI) and subsequent targeted biopsy has improved the ability to sample clinically significant lesions for assessment of tumor aggressiveness[137], however, mpMRI misses up to 35% of aggressive PCa foci in the context of multifocal disease [138-140]. In addition, prostate biopsy remains a costly procedure with potential complications including bleeding and infections[141], as well as other possible effects on the prostate/tumor microenvironment[142, 143]. The potential consequences of this paradigm are: delayed detection of high-grade prostate cancer; unnecessary biopsies and their associated effects; overdetection, overtreatment and the associated treatment-related complications for likely indolent, low grade (Gleason score 6 [Grade Group 1; GG1]) PCa and the consequent addition of unnecessary burden to the health care system[144]. Hence, non-invasive biomarkers that can accurately identify patients at risk for

aggressive PCa, particularly at the first biopsy decision-making point, are urgently needed.

Serum prostate specific antigen (PSA) has revolutionized the PCa early detection biomarker field[136]. Several forms of the protein (total-, free-, complexed-, intact- or [-2]pro-PSA) as well as multiplexed models/tests combining PSA with other clinicopathological parameters or the related kallikrein KLK2 (e.g. PCPT risk calculator, PHI, 4K score[145-147]) have improved on its performance. However, as a prostate epithelial tissue as opposed to a pure PCa protein, PSA remains a low specificity biomarker not recommended for widespread screening by the US Preventative Services Task Force (USPSTF), in large part due to false positive findings and overdetection of indolent PCa[148]. Furthermore, in the active surveillance (AS) setting (a low-grade PCa management strategy of "watchful waiting") PSA shows poor performance in discerning cancers likely to progress, with biopsy remaining the best strategy, albeit an imperfect one[149].

Other "liquid biopsy" approaches have included urine as a biofluid uniquely positioned anatomically to non-invasively assay prostate contents particularly when immediately preceded by a digital rectal exam (DRE) in order to enrich for prostate-derived material[150]. PCa is a multifocal disease with individual co-occurring clonally unrelated foci having heterogeneous inter- and intra-focal genomic and transcriptomic profiles[151-153]. Urine provides the opportunity to sample contents from the entire prostate (which biopsies may under-sample or entirely miss), a particularly important aspect for small or mpMRI-negative cancer foci. The only FDA approved urine PCa biomarker test is the PROGENSA PCA3 assay, which uses transcription mediated amplification (TMA) to measure the PCa-associated lncRNA *PCA3*[150]. Additional urine based laboratory developed tests include SelectMDx which uses (reverse transcription quantitative PCR (RT-qPCR) to measure post-DRE whole urine mRNA for the

genes *DLX1* and *HOXC6* [normalized to urinary *KLK3* (PSA)][154] and the ExoDx Prostate IntelliScore, also an RT-qPCR test measuring lncRNA *PCA3* and the oncogenic Ets family transcription factor *ERG* mRNA (normalized to *SPDEF*) in urine exosomes without a DRE [155]. A third urine-based laboratory developed test, Mi Prostate Score (MiPS), previously-developed by our group, uses TMA to measure *PCA3* and the chimeric RNA product of the fusion between androgen receptor (*AR*) driven gene *TMPRSS2* and *ERG* (*TMPRSS2-ERG*; splice isoform T1E4[68]) normalized to *KLK3* in post-DRE whole urine[74].These available urine tests have been shown to outperform serum PSA alone in predicting the presence of PCa and high-grade PCa (GG > 1) at first biopsy and other settings.

Importantly, the wealth of information generated in the past decade by large-scale next generation sequencing (NGS) profiling studies of the PCa tissue genome and transcriptome remains largely underdeveloped for the purposes of early detection biomarker translation[39, 44, 156, 157]. Prominent PCa molecular features include rearrangements involving Ets transcription factor family members primarily fused to highly expressed AR-controlled genes[67] (e.g. *TMPRSS2-ERG*), *SPOP* somatic mutations, and germline PCa predisposing variants (e.g. the *HOXB13* p.G84E risk SNP) [39, 44, 68]. As part of the Early Detection Research Network (EDRN), our group has discovered, validated and/or utilized a number of tissue-based PCa RNA biomarkers[158], including lncRNAs (such as *SChLAP1* and *ARLNC1*) that have been increasingly associated with PCa development and progression[159-161]. Tissue RNA PCa biomarkers have also been shown to predict various cancer related outcomes and are in clinical use primarily in the three available tissue prognostic assays (Prolaris – CCP[162], Decipher – GC[163] and OncotypeDx – GPS[164, 165]. Importantly, none of the currently available urine (or tissue) assays target gene fusions (beyond *TMPRSS2-ERG*), recurrent SNPs/somatic

mutations or employ NGS technology.

Here we leveraged the wealth of PCa genomic and transcriptomic knowledge, including our previous development and application of a 306 gene targeted multiplexed RNA NGS assay to formalin fixed paraffin embedded (FFPE) PCa tissue samples[158] to develop a whole post-DRE urine, multiplexed, RNA NGS assay for aggressive prostate cancer early detection, named NGS-MiPS. NGS-MiPS includes 84 transcriptomic biomarkers (assessed by amplicons qualified in our previously developed FFPE tissue assay[158]) including those in the three aforementioned urine tests, many additional splicing isoforms of common PCa gene fusions, gene signature mRNAs and established and novel lncRNAs as well as expressed germline SNP variants and somatic mutations. We describe the pre-clinical development of NGS-MiPS showing high analytical performance and concordance with MiPS. We used a machine learning approach to train a preliminary NGS-MiPS model for aggressive PCa detection that outperformed serum PSA and other current models in two separate validation cohorts. Lastly, NGS MiPS detected germline predisposing and somatic PCa driver mutations in patient urine RNA. Taken together our data support continued clinical development of NGS-MiPS for the non-invasive early detection of aggressive prostate cancer.

3.3 Materials and Methods

3.3.1 Patient selection

All samples were collected under Institutional Review Board (IRB)-approved protocols by the University of Michigan or other local IRBs with patient informed consent. To asses MiPS vs. NGS-MiPS concordance, we selected patients from multiple cohorts with available clinical

MiPS (or PROGENSA *PCA3* alone) testing representing a range of sores for *PCA3* and *TMPRSS2-ERG*. For model training and validation in our extreme design cohort, we selected Michigan Medicine patients representing the entire spectrum of disease grade on pathology (Benign to Gleason score 10 [Grade Group 5], but excluding Gleason score 3+4=7 [Grade Group 2]) in order to assess more pronounced transcriptomic differences likely to be found in the more extreme ends of localized disease spectrum. For our active surveillance (AS) cohort we selected Michigan Medicine patients on AS with urine collected prior to an on-AS biopsy where the biopsy pathology showed the range of NCCN guideline category (very low, low, and intermediate) risk or benign prostate tissue. Lastly, pre-DRE work was done on matched pre- and post-DRE urine of a pilot cohort of randomly selected Michigan Medicine PCa patients.

3.3.2 Digital rectal exam and urine collection

For all urine samples, collection was done as for the MiPS assay[74]. DRE was conducted immediately prior to urine collection applying enough pressure to slightly depress the prostate surface from the base to the apex and from the lateral to the median line with 3 strokes for each lobe. 20-30 ml (but no less than 2.5 ml) of first-catch urine (initial part of the urine stream) was collected in a preservative-free cup following DRE and chilled to 2 to 8°C or on ice. Within 4 hours of collection, the chilled urine specimen was inverted 5 times to mix and ~2.5 ml of it was transferred into the urine specimen transport tube (Progensa *PCA3* Urine Specimen Transport Tube containing 2.3 ml of urine transport media) and the tube was inverted 5 times to mix. The mixed samples were processed immediately or stored within five days at 2°C to 8°C for up to 14 days, –35°C to –15°C for up to 11 months or below –65°C for up to 36 months.

Samples were subjected to no more than five freeze–thaw cycles.

3.3.3 RNA isolation

RNA isolation for NGS-MiPS was performed using the ZR Viral RNA Kit (Zymo Research, Irvine, CA). The entire ~5 ml of urine/urine transport media mixture was fully thawed and stored at room temperature for 30 min (vortexing occasionally to dissolve precipitate), then warmed up to 37°C for 5 min to fully dissolve. 3x volume (~15 ml) of Viral RNA Buffer, pre-warmed for 30 min at 37°C, was added and the mix was vortexed. The mix was transferred immediately to a reservoir connected to the spin column in a vacuum manifold filtration system connected to -850 to -900 mbar vacuum pressure generated using a KNF Laboport pump (Qiagen, Hilden, Germany) until the entire sample filtered through or for up to 80 min, whichever came first. 500 ul Viral Wash Buffer was added to the spin columns which were then centrifuged for 2 min x 13,000g. RNA was eluted by centrifuging as in the previous step with 13 ul of RNAse-free water and stored in -80°C. RNA concentration was measured with the High Sensitivity RNA Kit on the Qubit Fluorometer (Thermo Fisher Scientific, Waltham, MA). RNA for the high-throughput testing experiments in Figure 16 was isolated using 0.5 ml of the ~5 ml urine/media mix with a high-throughput method.

3.3.4 Panel selection, NGS and data analysis

We used our previously developed and validated amplicon based multiplexed RNAseq PCa tissue panel[158] as the starting pool of 306 target transcripts. These were filtered down to

84 transcripts targeting prostate-, PCa- or aggressive PCa-specific genes, and this custom targeted RNA NGS-MiPS panel was generated by the ThermoFisher white glove team for the multiplexed PCR Ion Torrent Ampliseq platform (Thermo Fisher Scientific, Waltham, MA). All amplicons are exon-exon junction-spanning except for the one targeting the *HOXB13* p.G84E risk SNP.

Targeted NGS was performed essentially as previously described[49, 102, 158, 166-169]. Briefly, to construct targeted RNA NGS libraries, 10 - 15 ng RNA was treated with DNase (ezDNase Enzyme kit, Thermo Fisher Scientific, Waltham, MA; 5 ul reaction volume) to remove any DNA impurities and subjected to random priming mediated RT with the SuperScript IV VILO Master Mix kit (Thermo Fisher Scientific, Waltham, MA; 10 ul reaction volume). Target amplification was then performed on the resulting cDNA using the custom designed NGS-MiPS panel containing 84 amplicons covering 7,587 bases and the Ion AmpliSeq™ Library Kit 2.0 (Thermo Fisher Scientific, Waltham, MA; 20 ul reaction volume) with 23 PCR amplification cycles (24 or 25 cycles were used for ~5% of samples that only had 5 – 10 or 2 – 5 ng RNA available, respectively). The rest of the library procedure was performed per kit protocol with barcode incorporation. Libraries were quantitated with the Ion Library TaqMan™ Quantitation Kit (Thermo Fisher Scientific, Waltham, MA; 10 ul reaction volume in duplicates). Templating was performed using the Ion PI™ Hi-Q™ OT2 200 Kit on the Ion OneTouch™ 2 System using the Qubit Fluorometer QC method (Thermo Fisher Scientific, Waltham, MA). Sequencing was performed using the Ion PI™ Hi-Q™ Sequencing 200 Kit on the Ion Proton System with the Ion PI chip (Thermo Fisher Scientific, Waltham, MA).

Sequencing data were analyzed using Torrent Suite 5.0.4 with alignment by TMAP and coverage was analyzed with the Ion Torrent coverageAnalysis plugin (version 5.0.4.0), both

using default parameters. Filtering thresholds of mapped reads > 300,000, on-target mapped reads > 60%, and end-to-end mapped reads > 50% were used to select samples for downstream analysis. End-to-end read numbers for each target were normalized to sample-specific *KLK3* end-to-end reads and multiplied by 100,000 yielding normalized read counts. Log2 of these normalized reads + 1 are plotted in figures as indicated. Expression level heatmaps were generated and visualized in Cluster 3.0 and TreeView, respectively. Targets with zero reads in all samples (those targeting rare gene fusions) were removed prior to heatmap generation. To generate NGS-derived traditional MiPS-like scores for MiPS and NGS-MiPS comparison analysis, *KLK3*-normalized *PCA3* reads were multiplied by 1,000 (as opposed to 100,000) for consistency with the clinical MiPS formulas[74]. Traditional MiPS (or the PROGENSA PCA3 assay alone) was performed per protocol as previously described[74]. Expressed mutations were analyzed using aligned read pileups in the Integrated Genome Viewer (IGV, Broad Institute, Boston, MA).

3.3.5 RT-qPCR

KLK3, *PCA3*, *TMPRSS2-ERG.T1E4* and *TMPRSS2-ERG.T2E4* NGS-MiPS expression levels were validated by quantitative RT-qPCR for 8 samples with a range of expression. Pre-designed TaqMan assays were obtained from Thermo Fisher Scientific, Waltham, MA. cDNA generated from 2ng RNA with the SuperScript IV VILO Master Mix kit (Thermo Fisher Scientific, Waltham, MA; 10 ul reaction volume) was subjected to pre-amplification using the TaqMan™ PreAmp Master Mix kit, (Thermo Fisher Scientific, Waltham, MA; 25 ul reaction volume, 6.25 ul cDNA, 12.5 ul Pre-amplification Mix, pooled TaqMan assays to 0.05x final

concentration, 14 PCR cycles). qPCR reactions (20 ul) were performed in triplicate using TaqMan Gene Expression Assay (5' FAM, minor groove binder (MGB) and 3' non-fluorescent quencher NFQ, Thermo Fisher Scientific, Waltham, MA). 0.55 ul of pre-ampified DNA was mixed with 10 ul TaqMan Gene Expression Mastermix and 1 ul Taqman Assay (20x) and water. Reactions were performed in 384 well plates on the QuantStudio 12K Flex (Applied Biosystems). Baseline and Ct thresholds were set using QuantStudio 12K Flex Real-Time PCR System Software. $2^{\Delta Ct}$ values (ΔCt calculated as the difference in Ct of *KLK3* to each of the other 3 transcripts) were scaled up by 30,000, 100,000 and 500,000 for *PCA3*, *TMPRSS2-ERG.T1E4* and *TMPRSS2-ERG.T2E4*, respectively, for ease of visualization.

3.3.6 Expressed mutation confirmation

Urine RNA for Sanger sequencing was purified as for NGS-MiPS above. cDNA was made from 9 ng urine RNA using the SuperScript IV VILO Master Mix (Thermo Fisher Scientific, Waltham, MA). Urine genomic DNA was purified using the ZYMO Quick-DNA Urine Kit (Zymo Research, Irvine, CA) per the manufacturer's protocol to purify total (cell and cell-free) DNA from 3.0 ml of the urine:media mix, as above. Bi-directional Sanger sequencing was performed over the observed *HOXB13* SNP (G84) on the genomic and cDNA using primers described in Storebjerg et al.[170] 1.0 ng genomic DNA or the entire cDNA was used as PCR template with the Invitrogen Platinum PCR Supermix Hi-Fi (Thermo Fisher Scientific, Waltham, MA) using the manufacturer's suggested conditions except a 2-round PCR with 42 and then 50 cycles was used for genomic DNA due to low amounts. PCR products were subjected to bidirectional Sanger sequencing by the University of Michigan DNA Sequencing Core after

treatment with ExoSAP-IT PCR cleanup reagent (GE Healthcare) and sequences were analyzed using SeqMan Pro software (DNASTAR, Madison, WI).

3.3.7 NGS-MiPS modeling

Modeling and statistical analysis was done using R version 3.2.3 (R foundation for Statistical Computing, http://ww.R-project.org). Quality control thresholds of > 300,000 mapped reads, on-target percentage > 60%, and > 50% of the mapped reads being end-to-end were used to select samples for downstream analysis. The extreme design cohort of patient samples described in Patient Selection section above was used for model training and validation (109 of 126 samples passed QC filtering). Prior to selecting informative variables, a third of the data (n=36) proportional to the ratio of classes present in the entire cohort (Benign/Grade Group (GG1) vs. GG3-GG5), was randomly selected and set aside, to be used as a held-out testing set for the model. The remaining training dataset was subjected to a random forest variable selection method (VSURF R package[171]) which reduced the target pool from a total of 84 to 29 transcripts. Briefly, VSURF calculates variable importance (VI) by averaging the mean decrease in the out-of-bag error from permutation of variables across 50 grown forests. Variables are then ranked and uninformative variables are removed based on a threshold for the standard deviation of the VI - computed through a classification and regression tree (CART)[171]. VSURF was run with default parameters (ntree = 2000 and mtry = number of transcripts/3). The resulting variable list of 29 transcripts was used to build a regularized logistic regression model with an R package caret[172]. The caret package was used to train and perform a 5-fold validation process by performing a grid search that maximizes the area under the receiver operating characteristic

(AUROC [AUC]) curve. The final tuned model was then used to predict the classes on the: 1) held-out validation dataset and 2) Active Surveillance cohort dataset and AUC was used to evaluate the model's performance (with 95% confidence intervals). Model predictions with NGS-MiPS data were calculated for: 1) Clinical high-grade MiPS[74]; 2) A retrained MiPS model (using the 3 MiPS variables on the Extreme Design NGS-MiPS training dataset); 3) Serum PSA alone; 4) A retrained SelectMDx model (using the 2 SelectMDx variables (*HOXC6 + DLX1*) on the Extreme Design NGS-MiPS training dataset) and 5) A retrained ExoDx Prostate IntelliScore model (using the 2 ExoDx Prostate IntelliScore variables (*PCA3 + ERG*) on the Extreme Design NGS-MiPS training dataset), were compared with the 15-transcript model.

3.4 Results

3.4.1 NGS-MiPS assay design and workflow

The starting pool of candidate targets for our urine RNA NGS assay (NGS-MiPS), consisted of the 306 transcripts in our previously reported tissue RNA NGS assay[158]. These include transcripts from: the three commercially available PCa tissue prognostic assays (Prolaris – CCP[162], Decipher – GC[163] and OncotypeDx – GPS[164, 165]); relevant transcriptional signatures such as stroma-expressed genes, AR-driven transcriptional module; multiple splice isoforms of the most commonly occurring PCa gene fusions; PCa-associated lncRNAs and expressed germline variants and somatic mutations. Application of that panel on tissue-derived RNA representing the entire spectrum of disease aggressiveness showed the expected expression patterns[158] (**Figure 12,** heatmap) and validated individual amplicons for inclusion in our urine assay.

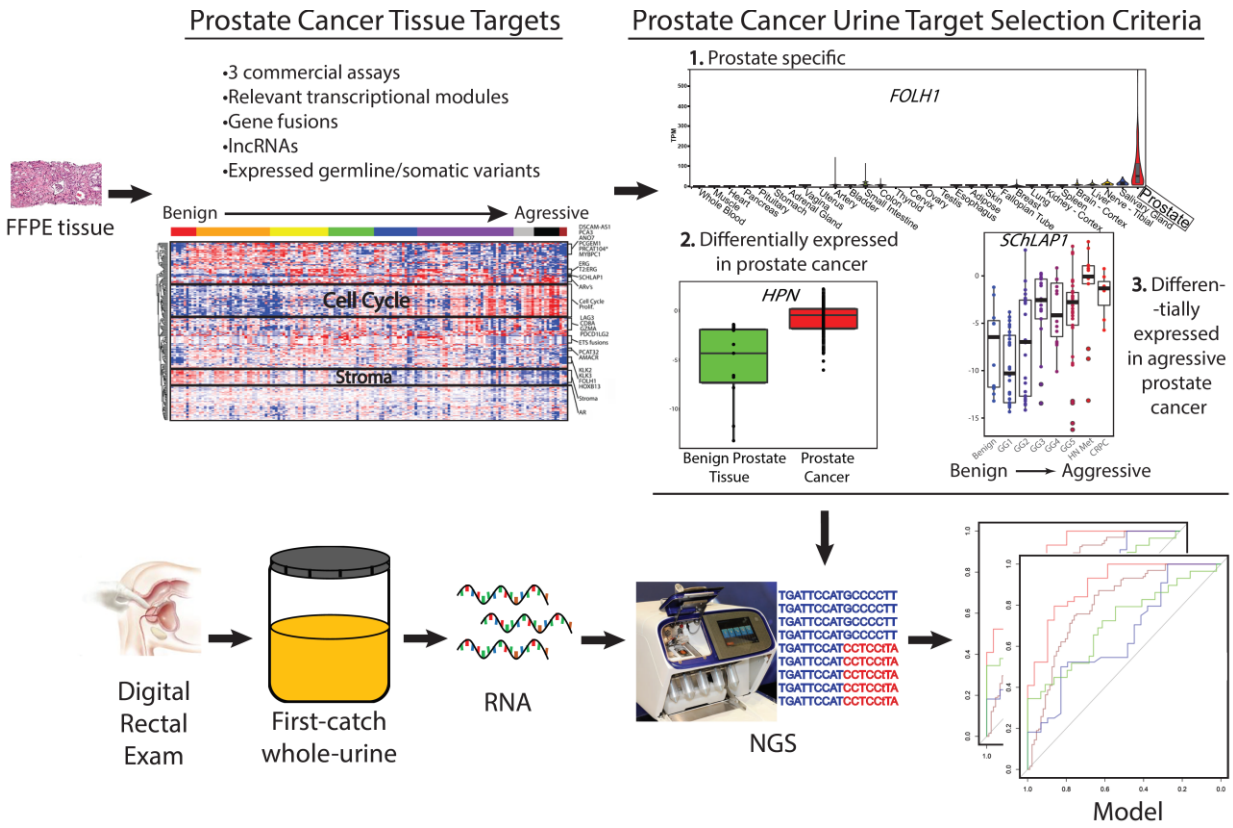


Figure 12. Workflow for development of urine RNA NGS assay (NGS-MiPS) for early detection of aggressive prostate cancer (PCa). Our previously validated FFPE tissue-based PCa prognostic RNA NGS assay (Salami et al. JCI Insight, 2018[158]) served as the starting pool of gene targets and amplicons. Its 306 amplicons include those in the three commercial tissue-based prognostic assays, relevant transcriptional signatures, the vast majority of PCa gene fusions, known and novel PCa-related lncRNAs, expressed somatic/germline variants etc. Heatmap shows tissue expression data with expected patterns of expression across the disease spectrum. We filtered these 306 targets to select 84 transcripts relevant in urine by prioritizing 1) Prostate-specific targets (*FOLH1* expression levels by tissue type in GTEX database are shown), 2) Transcripts differentially expressed in PCa vs. normal prostate tissue (*HPN* tissue expression levels for benign vs. prostate cancer are shown[158]), and 3) Transcripts differentially expressed in aggressive PCa (Gleason Score >6) (*SChLAP1* tissue expression levels for benign vs. prostate cancer are shown[158]). For our urine assay ~30 ml of first-catch urine obtained immediately after a digital rectal examination of the prostate is mixed with RNA-preserving GenProbe urine transport media in a 1:1 ratio. 5 ml of this mix undergoes RNA isolation. NGS with the 84-transcript-panel described above is performed on the Ion Torrent sequencing platform, and target transcript read counts are normalized to sample-specific KLK3 read counts. A machine learning approach was used to train a model for predicting the presence of PCa and aggressive PCa on tissue pathology results.

We reduced this target set down to 84 transcripts (**Table 9**) for urine based assessment by

three major criteria: 1) Prostate specificity (in order to avoid confounding urine RNA contributions from other tissue types, in particular those in the GU system); 2) Differential expression in PCa vs. normal prostate tissue, and 3) Differential expression in aggressive/high grade PCa vs. indolent/low grade PCA (**Figure 12**, violin and box-plots). We thus generated a custom targeted RNA NGS panel for the Ion Torrent platform (NGS-MiPS) using this set of 84 amplicons qualified from our tissue RNA NGS assay.

Gene (Alternate Name)	Spanned Exons	Type (Target)	Gene Fusion	Fused Exons	Gene Fusion	Fused Exons	Gene Fusion	Fused Exons
ACPP	E2-E3	mRNA	*ACSL3-ETV1*	A3-E8	*SLC45A3-ERG*	S1-E4	*TMPRSS2-ETV1*	T1-E4.COSF115
AR-FL5-1-2	E1-E2	mRNA	*C15orf21_HMGN2P46-ETV1*	C1-E3	*SLC45A3-ETV1*	S1-E7	*TMPRSS2-ETV1*	T1-E5
AR-V7-3-CE3	E1-E2	mRNA	*CANT1-ETV4*	C2-E5	*SLC45A3-ETV4*	S1-E3	*TMPRSS2-ETV1*	T2-E5
DLX1	E1-E2	mRNA	*CANT1-ETV4*	C2-E5.COSF1155	*SLC45A3-ETV5*	S1-E8	*TMPRSS2-ETV1*	T2-E8
ERG	E1-E2	mRNA	*DDX5-ETV4*	D3-E4	*SNURF-ETV1*	S2-E7	*TMPRSS2-ETV1*	T3-E8
FOLH1	E5-E6	mRNA	*DDX5-ETV4*	D3-E5	*TMPRSS2-ERG*	T1-E2	*TMPRSS2-ETV4*	T1b-E3
HOXC6	E1-E2	mRNA	*DDX5-ETV4*	D5-E5	*TMPRSS2-ERG*	T1-E3	*TMPRSS2-ETV4*	T1-E1
HPN	E10-E11	mRNA	*ESRP1-RAF1*	E13-R6	*TMPRSS2-ERG*	T1-E4	*TMPRSS2-ETV4*	T1-E2
KLK2	E3-E4	mRNA	*ESRP1-RAF1*	E13-R8	*TMPRSS2-ERG*	T1-E5	*TMPRSS2-ETV5*	T1b-E2
KLK3	E2-E3	mRNA	*EWSR1-ETV4*	E7-E8	*TMPRSS2-ERG*	T1-E6	*TMPRSS2-ETV5*	T1-E2
NKX3-1	E1-E2	mRNA	*FOXP1-ETV1*	F11-E7	*TMPRSS2-ERG*	T1-EIIIc_4	*TMPRSS2-ETV5*	T3-E2
OR51E2	E2-E3	mRNA	*FOXP1-ETV1*	F11-E8	*TMPRSS2-ERG*	T2-E2	*UBTF-ETV4*	U2-E5
PDLIM5	E6-E7	mRNA	*HERPUD1-ERG*	H1-E4	*TMPRSS2-ERG*	T2-E3		
TDRD1	E16-E17	mRNA	*HMGN2P_C15ORF21-ETV1*	H2-E8	*TMPRSS2-ERG*	T2-E4		
HOXB13	E1	mRNA (Germline Variant)	*HNRNPA2B1-ETV1*	H1-E4	*TMPRSS2-ERG*	T2-E5		
SPOP	E5-E6	mRNA (Somatic Mutation)	*KLK2-ETV1*	K1-E9.COSF1116	*TMPRSS2-ERG*	T2-EIIIc_4		
AC009478.1	E3-E4	lncRNA	*KLK2-ETV1*	K1-E9.COSF1117	*TMPRSS2-ERG*	T3-E2		
G053084_T230577 (PRCAT104)	E3-E4	lncRNA	*KLK2-ETV4*	K1-E4	*TMPRSS2-ERG*	T3-E4		
G053084_T230586 (PRCAT104)	E2-E3	lncRNA	*NDRG1-ERG*	N2-E4	*TMPRSS2-ERG*	T4-E4		
PCA3	E2-E3	lncRNA	*NDRG1-ERG*	N3-E4	*TMPRSS2-ERG*	T4-E5		
PCAT4	E1-E2	lncRNA	*RP11_EST14-ETV1*	R1-E7	*TMPRSS2-ERG*	T5-E4		
RP11-314O13.1 (ARLNC1)	E3-E4	lncRNA	*RP11_EST14-ETV1*	R1-E8	*TMPRSS2-ERG*	T5-E5		
RP11-314O13.1 (ARLNC1)	E2-E3	lncRNA	*SLC45A3-BRAF*	S1-B8	*TMPRSS2-ERG.EF194202*			
SCHLAP1	E1-E2	lncRNA	*SLC45A3-BRAF*	S1-B9	*TMPRSS2-ETV1*	T1-E4.COSF14		

Table 9. NGS-MiPS panel targets. Genes or gene fusions targeted in NGS-MiPS are shown, with alternate gene names given in parentheses when available. For genes, exon-exon junctions spanned (or exon in the case of *HOXB13*) by the corresponding NGS-MiPS amplicons are shown. For gene fusions, the exons for the 3' and 5' fusion partners participating in the fusion at the breakpoint are shown, with the exons being named using the initial of the corresponding gene. For genes, the type of RNA [mRNA or lncRNA (long non-coding RNA)] is shown. Expression levels for all targets were assessed with *HOXB13* and *SPOP* amplicons being additionally used to detect germline variant and somatic mutation respectively, indicated in parentheses.

Urine was collected as "first-catch" (first ~30 ml of the urine stream) immediately following DRE (3 strokes per lobe), as this has been shown to be enriched for prostate-derived contents[150]. We isolated RNA from whole urine, which has been shown to be superior to urinary sediment or exosomes for detection of *PCA3* lncRNA and *ERG* mRNA[173]. This established urine collection strategy is the basis for the PROGENSA *PCA3* test (Hologic Inc, Bedford, MA, USA)[150, 174], our previous clinically-approved laboratory-developed MiPS

assay[74] and the SelectMDx assay[154]. Across a total of 233 patient samples used in NGS-MiPS development, we obtained a median of 62.5 ng urinary RNA (interquartile range 34.2 to 109.2 ng) from ~2.5 ml urine with lysis, which is enough to perform NGS-MiPS four times. NGS-MiPS sequencing results were consistent with high quality Ampliseq library preparation and sequencing metrics; specifically, we obtained a median of 100% of reads mapping to target regions (range, 99% - 100%) with 87% of those being end-to-end reads (reads spanning the entire amplicon, range, 82%-90%). Importantly, NGS-MiPS has a current informative sample rate of 98% (**Table 10**).

	Reads	On-Target Reads (%)	End-to-end Reads (%)	High quality bases (%)	Mean Read Length (bp)
Median	3,548,284	100.0%	86.9%	88.6%	125
Mean	3,594,155	99.9%	86.5%	88.7%	124
Range	2,508,053 - 5,613,786	98.8 - 100.0%	81.5 - 89.5%	86.0 - 89.9%	111 - 130

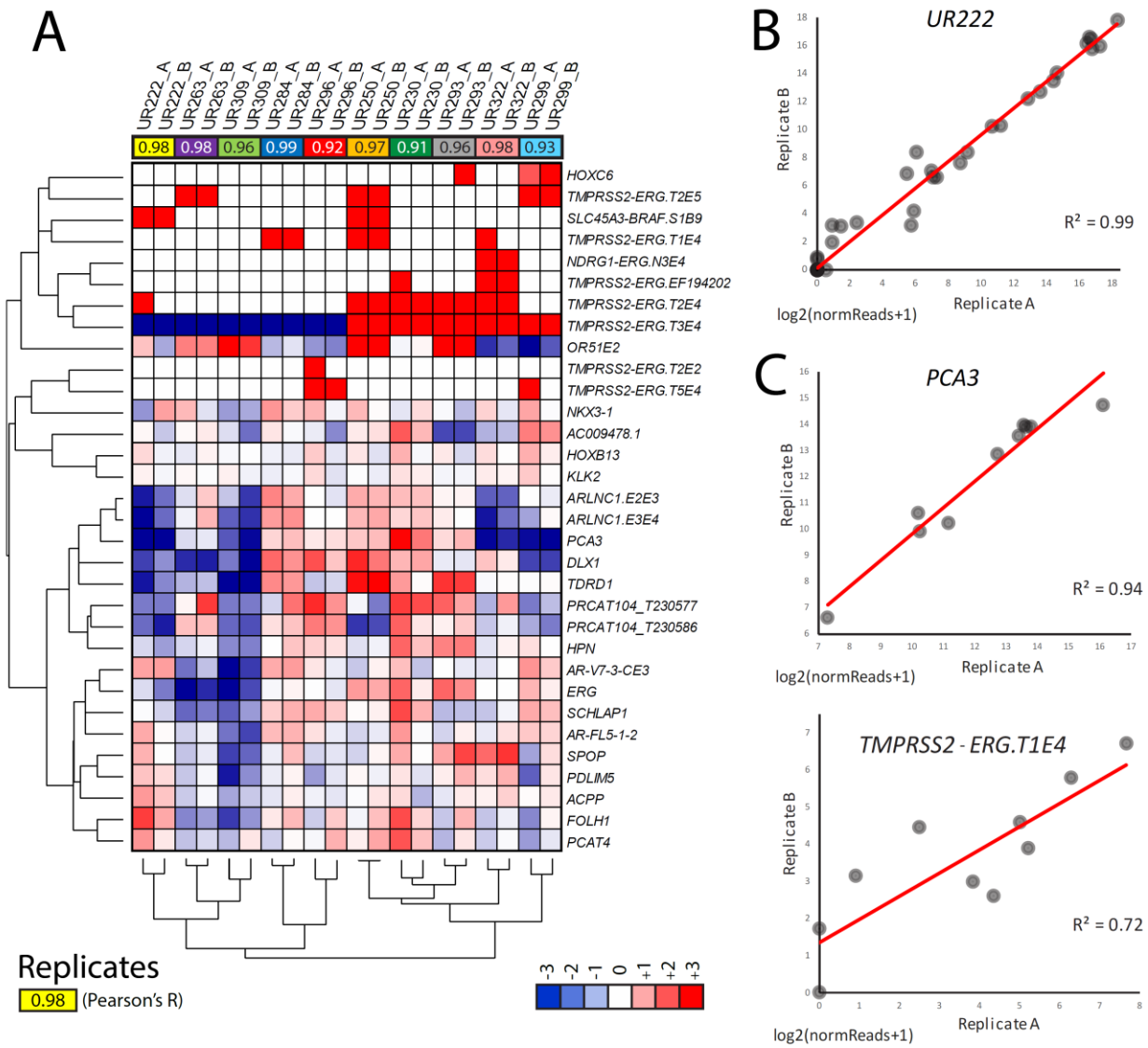
Informative Sample Rate	98.0%

Table 10. NGS-MiPS Sequencing Quality Metrics. Sequencing quality metrics for samples passing quality filters for a representative NGS-MiPS run. Median, mean and range are shown for: total number of raw reads; on-target reads (% of reads mapping to the target list); end-to-end reads (% of on-target reads spanning the entire target amplicon); high-quality bases (% of on target reads of a Q20 or better quality); mean read length (in base pairs). Bottom: informative sample rate (% of samples passing quality filters). NGS-MiPS performs highly in all sequence quality metrics.

3.4.2 NGS-MiPS assay analytical validity testing

To assess NGS-MiPS reproducibility, we constructed two replicate libraries for each of 10 urine RNA PCa samples with library and sequencing performed ~6 months apart on stored RNA samples. Normalized target amplicon expression (urine target reads x 100,000 / urine *KLK3* reads) is shown in the heatmap in **Figure 13A** for targets above-background in at least one sample. As shown by unsupervised hierarchical clustering, paired replicates were highly correlated (median Pearson's R = 0.97, range 0.91 – 0.99); one representative sample is shown in

Figure 13B to demonstrate the dynamic range of gene expression assessable by the multiplexed amplicon based NGS-MiPS assay. Importantly, related amplicons—such as various splic isoforms of the same gene fusion (*TMPRSS2-ERG*) known to be coexpressed—or multiple amplicons targeting the same transcript (e.g. the lncRNAs *ARLNC1* or *PRCAT104*, **Figure 13A**) also showed highly correlated expression across samples, demonstrating internal validity of the individual targets (also shown in **Figure 13C** dot plots for *PCA3* and *TMPRSS2-ERG.T1E4*).



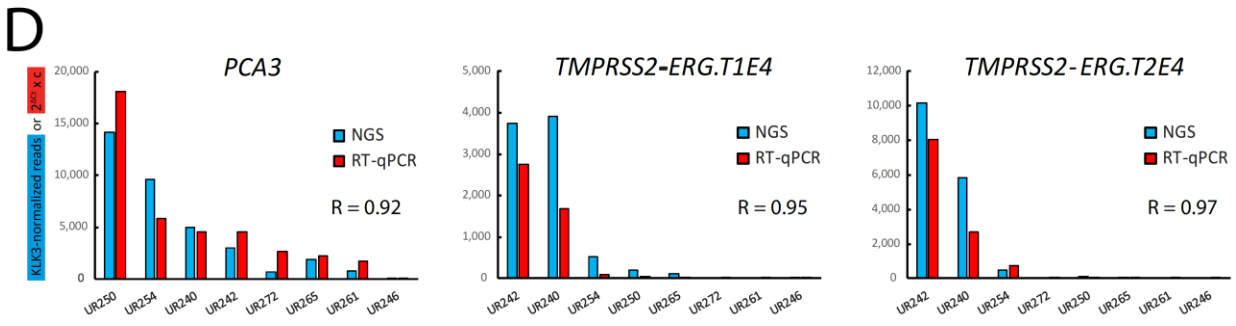
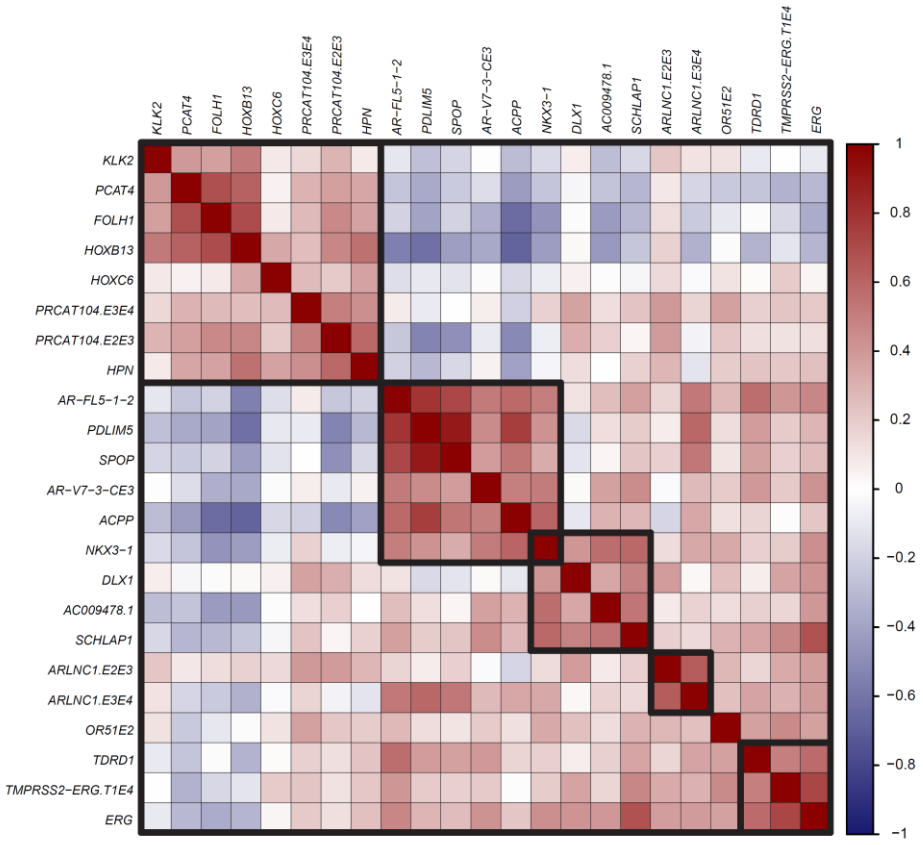


Figure 13. NGS-MiPS shows high technical reproducibility and accuracy. NGS-MiPS showed high performance in reproducibility and accuracy testing. **A.** 20 replicate NGS libraries (UR_A vs. _B in sample ID) for 10 randomly selected PCa patient urine RNA samples were constructed and sequenced on separate days. Paired replicates are denoted with the same color in header above heatmap. Pearson's pair-wise correlation R values for all *KLK3*-normalized targets are shown inside colored boxes and demonstrated high reproducibility (Median R = 0.97, range 0.91 – 0.99). Heatmap shows expression levels (gene-median-centered, log2(normalized reads +1)) where red and blue indicate over- and under-expression, respectively. Gene fusion targets are restricted to those having >32 normalized sequencing reads in at least one sample (read levels below that threshold are zeroed as background noise). Unsupervised hierarchical clustering (uncentered correlation similarity metric and centroid linkage clustering method) assigned paired replicates adjacent to each other and in separate branches for all 10 samples (sample dendrogram). **B.** Expression levels (Log2(normalized reads +1)) over all 84 transcripts for Replicate A vs. B are plotted for one representative sample (UR222) showing high concordance over the entire range of expression. **C.** Expression levels ¬(Log2(normalized reads +1)) over all 10 samples for Replicate A vs. B are plotted for *PCA3* and *TMPRSS2-ERG* isoform T1E4 (targets that are part of the MiPS test) showing high concordance. **D.** Assay accuracy against an orthogonal method was assessed by comparing NGS-MiPS vs. RT-qPCR in 8 samples for *PCA3* and the two main *TMPRSS2-ERG* splice isoforms: T1E4 and T2E4. Bar graphs show *KLK3*-normalized reads for NGS-MiPS and $2^{\Delta Ct}$ x c for RT-qPCR (where the constant c = 30,000, 100,000 and 500,000 for *PCA3*, *TMPRSS2-ERG.T1E4* and *.T2E4*, respectively, for ease of visualization). Samples are sorted left to right by the RT-qPCR value. Pearson's R values are shown demonstrating high concordance between the two orthogonal methods for all three transcripts.

Expression level correlation matrices of pair-wise target Pearson's correlation R coefficients with unsupervised hierarchical clustering yielded expected clusters of related NGS-MiPS targets. This included the two different amplicons each for *ARLNC1* or *PRCAT104* as well as a cluster containing *ERG*, *TMPRSS2-ERG* and *TDRD1*, a known co-expressed *ERG*

target[175] (**Figure 14A**). *TMPRSS2-ERG* splicing isoforms, known to be co-expressed in the same tumor, showed the expectedly high correlation to each-other for the majority of transcripts, although this was not true for all isoform combinations, which supports the potential clinical utility of multiple *TMPRSS2-ERG* isoforms as non-redundant, complementary PCa biomarkers (**Figure 14B**).

A



B

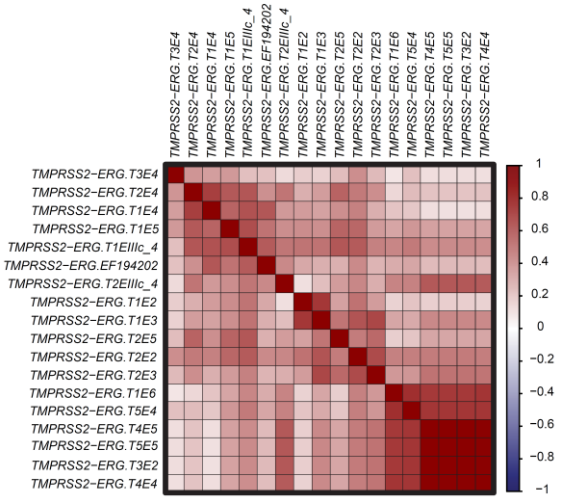
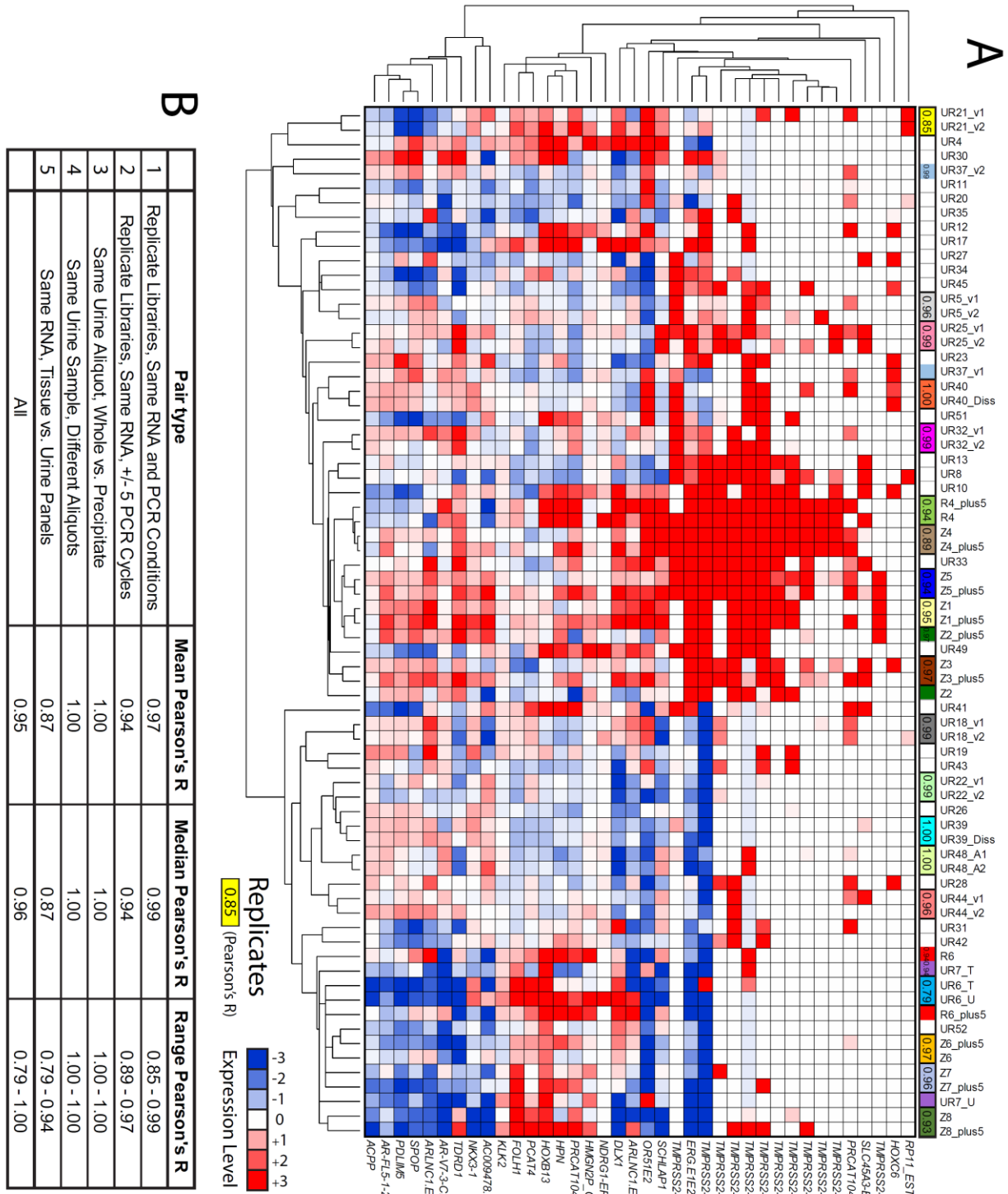


Figure 14. NGS-MiPS target correlation matrices. Unsupervised hierarchical clustering heatmaps of inter-transcript Pearson correlation coefficients for the **A.** non-fusion (except *TMPRSS2-ERG.T1E4*) and *B.* all *TMPRSS2-ERG* fusion NGS-MiPS transcripts in a randomly selected patient cohort. Targets are shown in the same clustered order both across and top-to-bottom. Shades of red and blue represent positive and negative correlation coefficients, respectively and white represents a correlation coefficient of zero. Boxes inside heatmap show several highly correlated clusters of transcripts.

We also assessed robustness of the NGS-MiPS assay to various optimal and suboptimal conditions using 23 total matched pairs of samples with variable pre-analytical and analytical conditions, such as varying the number of PCR cycles, using different urine aliquots, using fully dissolved urine:media mix vs. precipitate, and using NGS-MiPS vs. our PCa tissue NGS assay on urine RNA[158] (**Figure 15A**). Even under suboptimal conditions, NGS-MiPS showed strong reproducibility, with 19/23 (83%) of pairs clustering adjacently by unsupervised hierarchical clustering. The remaining 4 pairs ranged from being in the same main branch to the same higher level sub-branches (see sample dendrogram). We obtained high overall correlation (Pearson's median R = 0.96, range 0.79 – 1.00) with urine vs. tissue assays being the only pairs to show a slightly more inferior concordance (median Pearson's R = 0.87, range 0.79 – 0.94, **Figure 15B**).

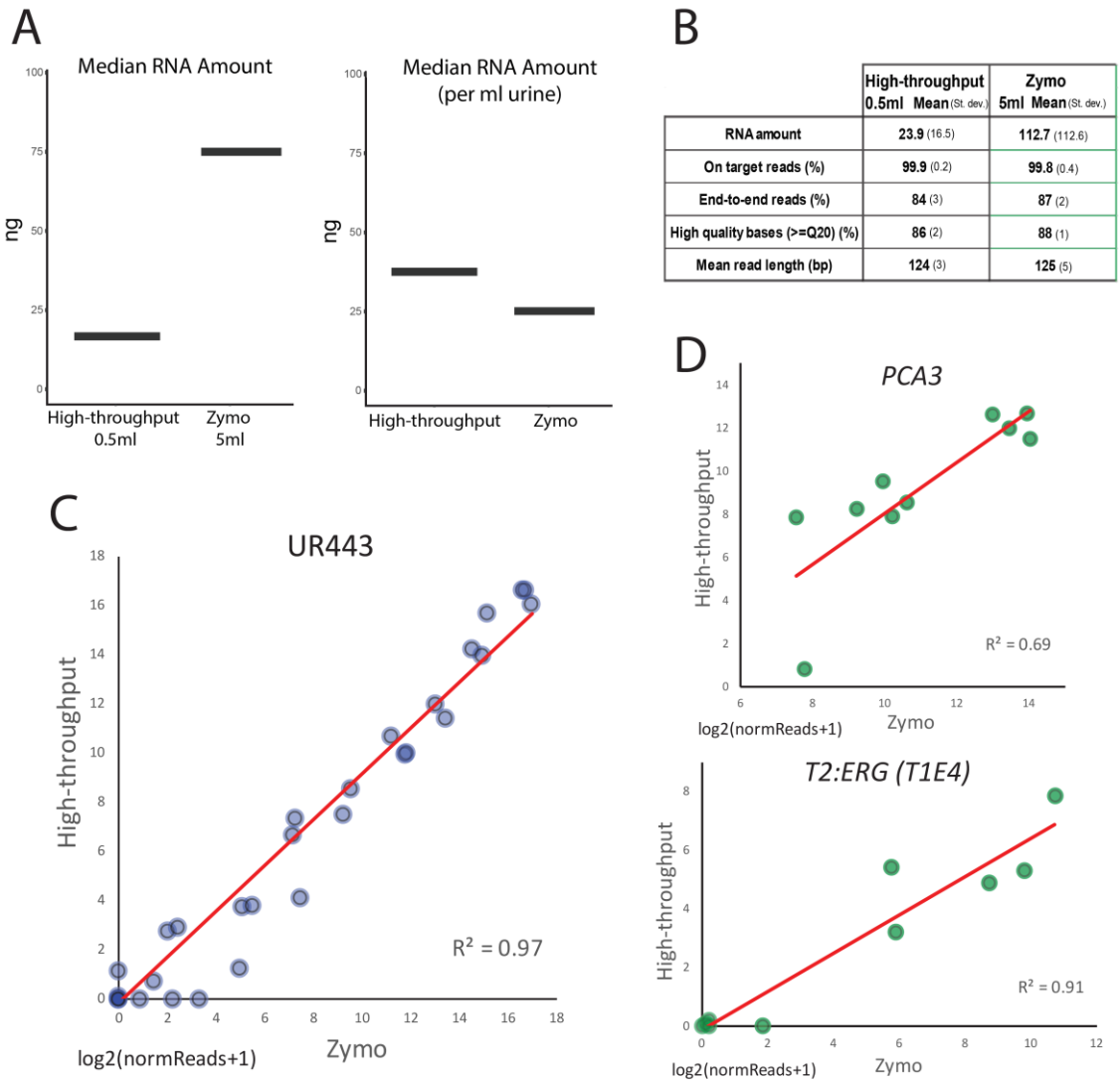


	Pair type	Mean Pearson's R	Median Pearson's R	Range Pearson's R
1	Replicate Libraries, Same RNA and PCR Conditions	0.97	0.99	0.85 - 0.99
2	Replicate Libraries, Same RNA, +/- 5 PCR Cycles	0.94	0.94	0.89 - 0.97
3	Same Urine Aliquot, Whole vs. Precipitate	1.00	1.00	1.00 - 1.00
4	Same Urine Aliquots	1.00	1.00	1.00 - 1.00
5	Same Urine Sample, Different Aliquots	0.87	0.87	0.79 - 0.94
	Same RNA, Tissue vs. Urine Panels	0.95	0.96	0.79 - 1.00
	All			

Figure 15. Robustness of the NGS-MiPS assay. In order to assess robustness of the NGS-MiPS

assay to various optimal and suboptimal conditions, NGS libraries were constructed for 23 total replicate pairs of the following types: *1)* Using the same RNA sample and identical conditions as in main Figure 2 (denoted by _v1 vs. _v2 in sample ID). *2)* As in 1. above, but with one replicate receiving 5 additional PCR cycles during library construction (_plus5 in sample ID). *3)* Using RNA from whole urine vs. urine precipitate (_Diss (dissolved precipitate) in sample ID). *4)* Different aliquots of the same urine sample (_A1 vs. _A2 in sample ID). *5)* Using NGS-MiPS vs. our tissue PCa RNA NGS assay on the same urine RNA with analysis restricted to NGS-MiPS targets (_U vs. _T in sample ID). **A.** Paired replicates are denoted with the same-color boxes. Pearson's pair-wise correlation R values for all *KLK3*-normalized targets are shown inside colored box in header. Heatmap shows expression levels (gene median centered, log2(normalized reads +1)) where red and blue indicate over- and underexpression, respectively. Unsupervised hierarchical clustering (uncentered correlation similarity metric and centroid linkage clustering method) assigned paired replicates exactly adjacent to each other for 19/23 (83%) pairs with the other 4 ranging from being in the same branch to the same sub[10] branch (sample dendrogram). *TMPRSS2-ERG* splice isoforms clustered together (target dendrogram). **B.** Mean, median and range of Pearson's correlation R coefficients are tabulated for the 5 pair types. No drastic decrease in correlation R values was observed for the various condition aberrations (see table), with the exception of replicates done with the tissue vs. urine assays which had a slightly lower median R of 0.87 (range 0.79 – 0.94).

We further assessed our assay's compatibility with a high-throughput automated RNA isolation method using 10-fold lower urine input volume (0.25 ml) (**Figure 16A**). We obtained ~1/5 of the standard RNA yield, yet 10/10 (100%) samples passed NGS-MiPS quality criteria and were not inferior in quality to samples processed with the standard method (**Figure 16B**). Standard and high-throughput RNA expression levels over the entire NGS-MiPS target set were highly concordant (median Pearson's R over 10 patient samples = 0.92, range 0.54 – 0.98) also shown in **Figure 16C** dot plot for one representative sample and in **Figure 16D** dot plots for *PCA3* and *TMPRSS2-ERG.T1E4* over the 10-patient set.



	High-throughput 0.5ml Mean (St. dev.)	Zymo 5ml Mean (St. dev.)
RNA amount	23.9 (16.5)	112.7 (112.6)
On target reads (%)	99.9 (0.2)	99.8 (0.4)
End-to-end reads (%)	84 (3)	87 (2)
High quality bases (>=Q20) (%)	86 (2)	88 (1)
Mean read length (bp)	124 (3)	125 (5)

Figure 16. NGS-MiPS assay compatibility with a high-throughput urine RNA isolation method. Urine aliquots from the same urine collection from 10 PCa patients underwent RNA isolation with the standard method (Zymo Research, Irvine CA) using 5ml urine : buffer mix as well as a high-throughput, automated RNA isolation method using 0.5ml urine : buffer mix. **A.** Median total and adjusted per ml RNA amounts are plotted. **B.** NGS-MiPS with RNA from both methods was performed showing no inferiority in sequencing quality metrics for the high-throughput method as compared to standard RNA: on-target reads (% of reads mapping to the target list); end-to-end reads (% of on-target reads spanning the entire target amplicon); high-quality bases (% of on-target reads of a Q20 or better quality); mean read length [in base pairs (bp)]. **C.** Expression levels (Log2(normalized reads +1)) over all 84 transcripts for High-throughput vs. Zymo RNA NGS-MiPS are plotted for one representative sample (UR443) showing high concordance over the entire range of

expression. **D.** Expression levels (Log2(normalized reads +1)) over all 10 samples for High-throughput vs. Zymo RNA NGS-MiPS are plotted for *PCA3* and *TMPRSS2-ERG* isoform T1E4 (targets which constitute the clinical MiPS test) showing high concordance.

Lastly, we assessed NGS-MiPS robustness to using ordinary whole urine not preceded by DRE (pre-DRE). While we obtained ~1/3 of the patient-matched post-DRE RNA yield, 11/11 (100%) patient pre- and post-DRE samples passed sequencing quality criteria and pre-DRE samples were non-inferior to their matched post-DRE samples (**Figure 17A, B**). Furthermore, we detected robust levels of highly expressed transcripts (*KLK2* and -*3*) in pre-DRE urine, in high concordance with their matched post-DRE samples (Pearson's R = 0.96 and 0.72, respectively, **Figure 17C**). More modestly expressed transcripts such as *PCA3* and especially *TMPRSS2-ERG.T1E4* had generally reduced and more variable pre-DRE levels unlike in their post-DRE counterparts (Pearson's R = 0.38 and 0.62, respectively, **Figure 17C**) leading to the conclusion that robust detection of PCa RNA biomarkers in standard non-DRE urine, at least with respect to whole urine, needs optimization for low expression genes.

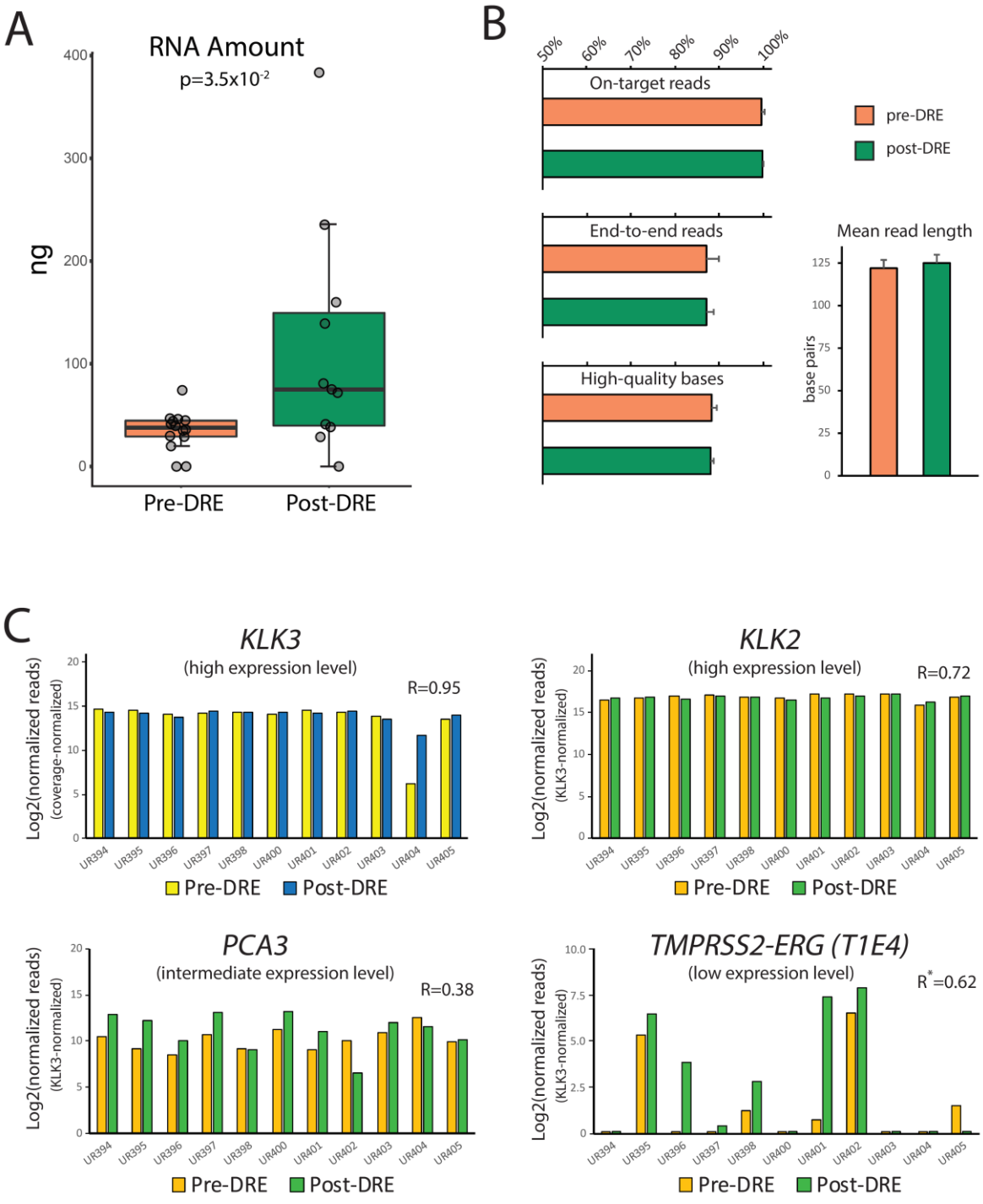


Figure 17. NGS-MiPS assay compatibility with pre-digital rectal exam (pre-DRE) urine. 11 pairs of matched pre- and post-DRE urine samples from 11 randomly selected prostate cancer

patients were subjected to NGS-MiPS. **A.** Boxplot shows RNA amount isolated (in ng) being lower in pre-DRE urine (Mann-Whitney p-value is shown). **B.** Sequencing quality metrics: on-target reads (% of reads mapping to the target list); end-to-end reads (% of on-target reads spanning the entire target amplicon); high-quality bases (% of on-target reads of a Q20 or better quality); mean read length (in base pairs), showed no differences between pre- and post-DRE urine. **C.** Bar graphs show expression levels (in log2 scale) for highly expressed transcripts *KLK3* (coverage-normalized) and *KLK2*, intermediate expression *PCA3*, and low expression level transcript *TMPRSS2-ERG.T1E4*. Pearson's R values for the log2 transformed data are shown. For TMPRSS2-ERG.T1E4, R* value denotes Pearson's R limited to the subset of samples (*n*=7) that are not fusion-negative in both pre- and post-DRE urine. Pre-post-DRE concordance decreased with decreasing expression level.

Finally, to assess our assay's accuracy against an orthogonal method, we also performed RT-qPCR on patient post-DRE urine RNA samples using commercial pre-designed TaqMan primer-probe sets for *PCA3*, the two main splicing isoforms of *TMPRSS2-ERG,* T1E4 and T2E4, and *KLK3* for normalization (**Figure 13D**). NGS-MiPS and RT-qPCR expression levels for these three target transcripts were highly correlated (Pearson's R = 0.92, 0.95 and 0.97, respectively), thus indicating high NGS-MiPS accuracy in detection of RNA biomarker levels. Taken together, these data support NGS-MiPS as a reproducible, robust and accurate assay for quantifying PCa transcriptomic biomarkers from whole urine.

3.4.3 NGS-MiPS accurately recapitulates the clinically validated MiPS assay for *PCA3* and *TMPRSS2-ERG.T1E4* expression

The clinical MiPS test previously developed, validated and clinically offered by our group in a CLIA/CAP accredited, New York state approved laboratory[74], uses transcription mediated amplification (TMA) to measure urinary *PCA3* and *TMPRSS2-ERG.T1E4* expression levels (normalized to urine *KLK3* expression). We compared clinical MiPS scores to derived NGS-MiPS scores calculated similarly, but using NGS-MiPS *PCA3* and *TMPRSS2:ERG.T1E4* expression instead of TMA-quantified *PCA3* and *TMPRSS2:ERG.T1E4* scores in a total of 48

PCa patients. MiPS and derived NGS-MiPS scores for the two transcripts were highly concordant (Spearman's Rho = 0.75 and 0.96 for the 48 and 32 patient samples with available *PCA3* and *TMPRSS2-ERG.T1E4* MiPS scores, respectively, **Figure 18A & 18B**). MiPS also combines these urine *PCA3* and *TMPRSS2:ERG.T1E4* scores with serum PSA into two logistic regression models that predict risk of having PCa or high-grade [Grade Group (GG) > 1 (Gleason Score > 6)] PCa on biopsy[74]. MiPS and high-grade MiPS model biopsy risk predictions calculated using MiPS and derived NGS-MiPS scores above were also highly concordant (Spearman's Rho = 0.86 and 0.82 for the 32 patients with available scores for both transcripts, for all-grade and high-grade PCa, respectively, **Figure 18C & 18D; Table 11**). Taken together these data support NGS-MiPS as a highly reproducible, accurate and robust assay that is concordant with the analytically and clinically validated MiPS test.

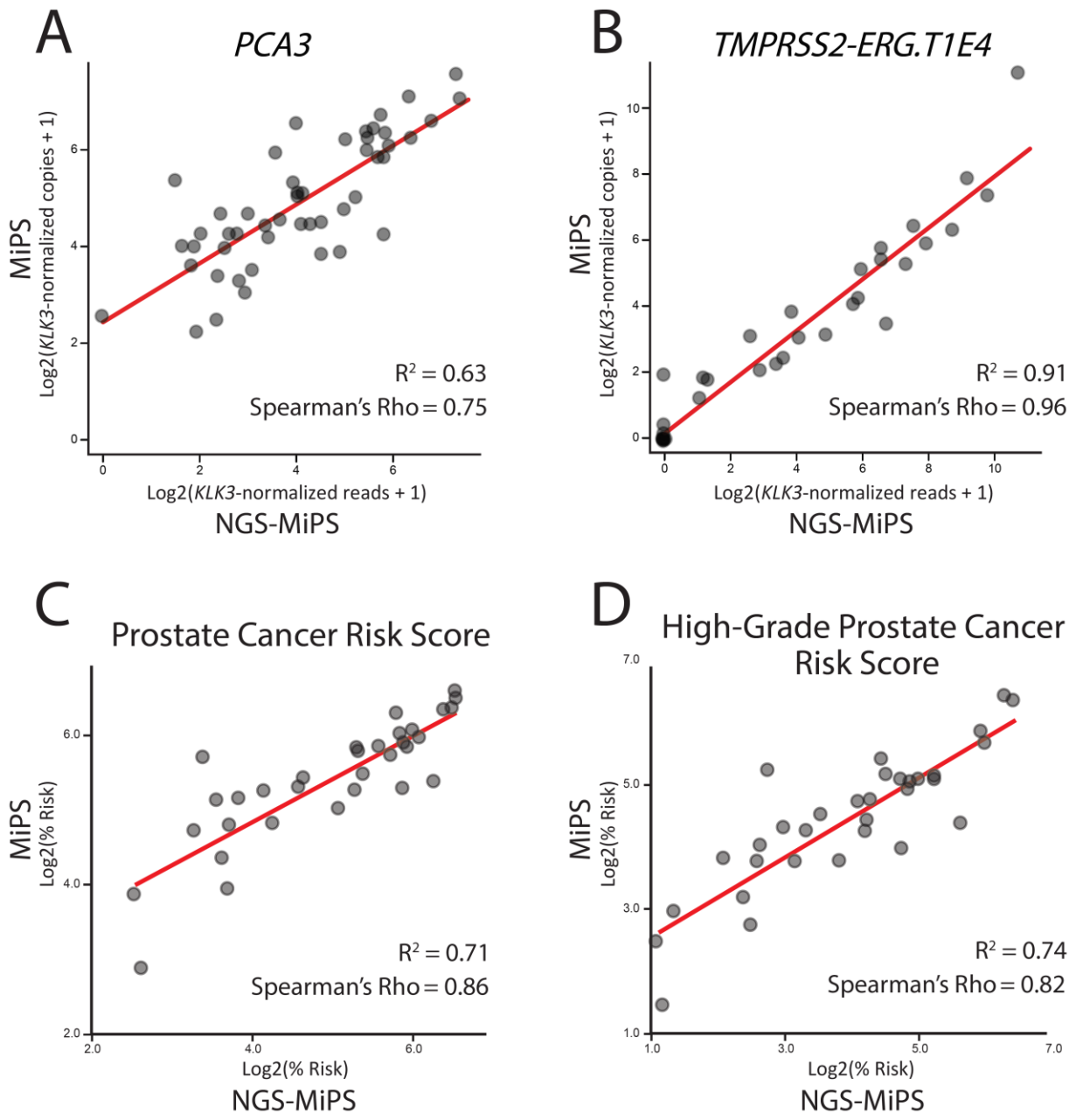


Figure 18. NGS-MiPS shows high accuracy compared to the clinical MiPS laboratory developed test. We performed NGS-MiPS on RNA isolated from stored aliquots of same urine void from PCa patients that had undergone our laboratory developed transcription-mediated amplification (TMA) based MiPS test (measuring *PCA3* and *TMPRSS2-ERG.T1E4*). **A.** For MiPS scores, clinical TMA-based *PCA3* number of copies is normalized to sample-specific *KLK3* number of copies and multiplied by a coefficient of 1,000. Similarly, for NGS-MiPS derived scores, *PCA3* sequencing reads were normalized to sample-specific *KLK3* reads and multiplied by 1,000. Expression level (Log2(normalized number of copies or sequencing reads +1)) are plotted for 48 samples with available clinical MiPS (or PROGENSA) *PCA3* data. *PCA3* scores were highly concordant between the two methods (Spearman's Rho = 0.75, linear fit R^2 = 0.63) **B.** *TMPRSS2-ERG.T1E4* scores for

the two methods were calculated and plotted as for *PCA3* in panel A. above (with the exception that a coefficient of 100,000 is used as opposed to 1,000) for 32 samples with available clinical MiPS *TMPRSS2-ERG.T1E4* data. *TMPRSS2-ERG.T1E4* scores were highly concordant between the two methods (Spearman's Rho = 0.96, linear fit $R^2 = 0.91$). **C. & D.** Concordance plots of urine-based risk probabilities for presence of **C.** prostate cancer, and **D.** high grade prostate cancer (Gleason Score > 6, Grade Group > 1), on biopsy. The clinical MiPS algorithm (Tomlins SA, et al., *European Urology*, 2016[74]) which combines serum PSA with urine *PCA3* and *TMPRSS2-ERG.T1E4* into a model validated to predict biopsy pathology was used to calculate risk probabilities for clinical (TMA) and NGS-derived MiPS scores for the two transcripts. Log2(% risk) is plotted for the 32 samples having clinical MiPS data. Risk predictions were highly concordant between the two methods (Spearman's Rho = 0.86 and linear fit $R^2 = 0.71$ for all-grade prostate cancer; Spearman's Rho = 0.82 and linear fit $R^2 = 0.74$ for high-grade prostate cancer).

		Prostate Cancer						High-Grade Prostate Cancer (Gleason Score > 6)					
		MiPS			NGS-Mips			MiPS			NGS-Mips		
Sample	Serum PSA (ng/ml)	Risk	95.CI.Lower	95.CI.Upper	Risk	95.CI.Lower	95.CI.Upper	Risk	95.CI.Lower	95.CI.Upper	Risk	95.CI.Lower	95.CI.Upper
UR424	2.5	7.3%	4.8%	11.1%	6.1%	3.9%	9.6%	2.7%	1.6%	4.7%	2.3%	1.3%	4.1%
UR411	2.3	14.8%	10.9%	20.0%	5.8%	3.6%	9.2%	5.6%	3.6%	8.5%	2.1%	1.2%	3.9%
UR419	5.4	15.2%	11.2%	20.4%	12.8%	9.2%	17.6%	6.7%	4.4%	10.0%	5.7%	3.7%	8.8%
UR447	1.1	26.2%	18.8%	35.2%	9.6%	5.7%	15.9%	7.8%	4.8%	12.3%	2.6%	1.3%	5.1%
UR442	4.8	20.3%	16.3%	25.0%	12.3%	8.8%	16.9%	9.0%	6.6%	12.3%	5.3%	3.4%	8.2%
UR443	5.5	27.6%	23.0%	32.8%	13.0%	9.5%	17.6%	13.6%	10.4%	17.5%	6.1%	4.0%	9.2%
UR409	5.5	32.0%	25.4%	39.4%	34.0%	26.4%	42.6%	13.6%	9.7%	18.9%	14.3%	9.8%	20.3%
UR435	4.9	28.2%	23.2%	33.8%	19.2%	15.1%	24.1%	13.7%	10.3%	17.9%	8.9%	6.4%	12.3%
UR431	3.4	34.3%	29.5%	39.5%	11.7%	8.0%	16.8%	14.1%	11.0%	18.0%	4.3%	2.6%	7.1%
UR445	3.6	39.0%	32.7%	45.8%	58.9%	50.7%	66.6%	15.6%	11.7%	20.6%	26.9%	20.5%	34.3%
UR446	4.3	37.8%	32.8%	42.9%	17.8%	11.6%	26.3%	16.3%	13.0%	20.4%	6.3%	3.6%	11.0%
UR417	5.6	38.1%	33.9%	42.5%	38.9%	34.5%	43.5%	18.8%	15.6%	22.6%	18.6%	15.2%	22.4%
UR420	5.7	39.4%	35.2%	43.8%	23.7%	18.0%	30.5%	19.2%	15.9%	23.0%	10.1%	6.8%	14.6%
UR448	9.5	35.4%	29.5%	41.8%	14.3%	9.9%	20.2%	19.9%	15.4%	25.3%	7.9%	5.0%	12.3%
UR429	6.1	41.7%	37.2%	46.3%	77.3%	71.5%	82.1%	20.7%	17.2%	24.7%	50.3%	43.6%	56.9%
UR418	6.0	44.4%	39.5%	49.4%	41.6%	34.3%	49.3%	21.8%	18.0%	26.1%	18.9%	14.0%	25.0%
UR427	6.7	42.4%	37.7%	47.2%	25.0%	19.8%	30.9%	22.7%	18.9%	26.9%	11.7%	8.4%	16.0%
UR457	2.5	55.0%	45.9%	63.7%	39.8%	32.7%	47.4%	26.6%	19.6%	35.0%	17.0%	12.5%	22.7%
UR410	3.4	57.4%	51.1%	63.4%	47.8%	39.9%	55.8%	27.1%	22.2%	32.6%	19.6%	14.5%	25.9%
UR430	4.7	53.1%	44.2%	61.8%	53.2%	46.8%	59.5%	30.2%	22.7%	38.8%	28.9%	23.5%	34.9%
UR433	3.8	66.3%	58.8%	73.1%	64.1%	52.5%	74.2%	33.6%	26.8%	41.1%	29.3%	20.1%	40.6%
UR432	4.1	62.0%	55.7%	68.0%	67.9%	62.0%	73.3%	34.1%	28.4%	40.3%	38.1%	32.5%	44.1%
UR414	5.3	60.1%	54.1%	65.9%	58.7%	53.9%	63.3%	34.8%	29.3%	40.8%	32.4%	28.3%	36.8%
UR453	6.0	65.1%	56.9%	72.4%	56.8%	45.1%	67.7%	35.0%	27.5%	43.3%	26.7%	18.0%	37.6%
UR428	5.5	57.5%	48.0%	66.5%	60.6%	50.7%	69.8%	35.0%	26.4%	44.7%	37.9%	28.4%	48.4%
UR439	8.0	56.3%	47.2%	65.0%	39.4%	32.7%	46.6%	36.6%	28.2%	46.0%	23.1%	17.8%	29.4%
UR426	15.5	52.0%	41.3%	62.5%	10.4%	6.1%	17.4%	37.8%	27.8%	49.0%	6.7%	3.6%	12.2%
UR438	4.3	78.5%	72.5%	83.6%	49.8%	42.3%	57.3%	50.5%	43.0%	58.0%	22.0%	16.8%	28.3%
UR421	2.6	81.6%	74.3%	87.3%	89.7%	83.6%	93.7%	51.1%	40.7%	61.4%	64.2%	51.9%	75.0%
UR434	9.8	79.9%	74.0%	84.8%	83.1%	77.2%	87.8%	58.0%	50.4%	65.3%	61.8%	53.4%	69.6%
UR412	34.9	89.4%	81.5%	94.2%	93.1%	87.1%	96.4%	81.7%	69.6%	89.7%	86.6%	76.1%	92.9%
UR423	10.7	95.5%	92.0%	97.5%	92.7%	88.0%	95.7%	85.2%	76.5%	91.1%	78.4%	68.3%	85.9%

Table 11. MiPS model urine risk scores for presence of PCa and high-grade PCa on biopsy from clinical MiPS vs. NGS-MiPS derived data. 32 prostate cancer urine samples underwent clinical TMA-based MiPS testing and NGS-MiPS. TMA and NGS *PCA3* and *TMPRSS2*:*ERG.T1E4* scores were applied to the MiPS algorithm (Tomlins, SA, et al. European Urology,2015[74]) to generate risk probabilites for having any-grade and high-grade (Gleason Score > 6, GG > 1) prostate cancer on biopsy. Calculated risk probabilites and upper and lower 95% confidence intervals are tabulated, sorted by increasing high-grade MiPS risk score.

3.4.4 NGS-MiPS model training and validation

We next assessed a new cohort of urine samples from 126 patients undergoing biopsy or prostatectomy (RRP) of which 109 (86.5%) met stringent NGS-MiPS quality control criteria. These patients were categorized into two groups based on their biopsy (or RRP if available) results as benign or Grade Group (GG) 1 (Gleason Score 6) (*n=65*) vs. GG ≥ 3 (Gleason ≥ 4+3=7) (*n=44*), thus excluding GG2 (Gleason 3+4=7) patients (Extreme Design Cohort). This was done in order to identify the more pronounced transcriptomic differences expected between the extremes of the disease spectrum, while grouping commonly indolent GG1 cancer together with benign biopsy cases. Pathologic measures such as tumor size (for prostatectomy cases), positive biopsy core fraction and maximum core involvement showed marked differences between the two groups as expected, **Figure 19A**. In contrast, 50/65 (77%) of benign/GG1 patients had a serum PSA greater than the commonly used threshold of 4.0 ng/ml consistent with that biomarker's modest sensitivity (**Figure 19A)**.

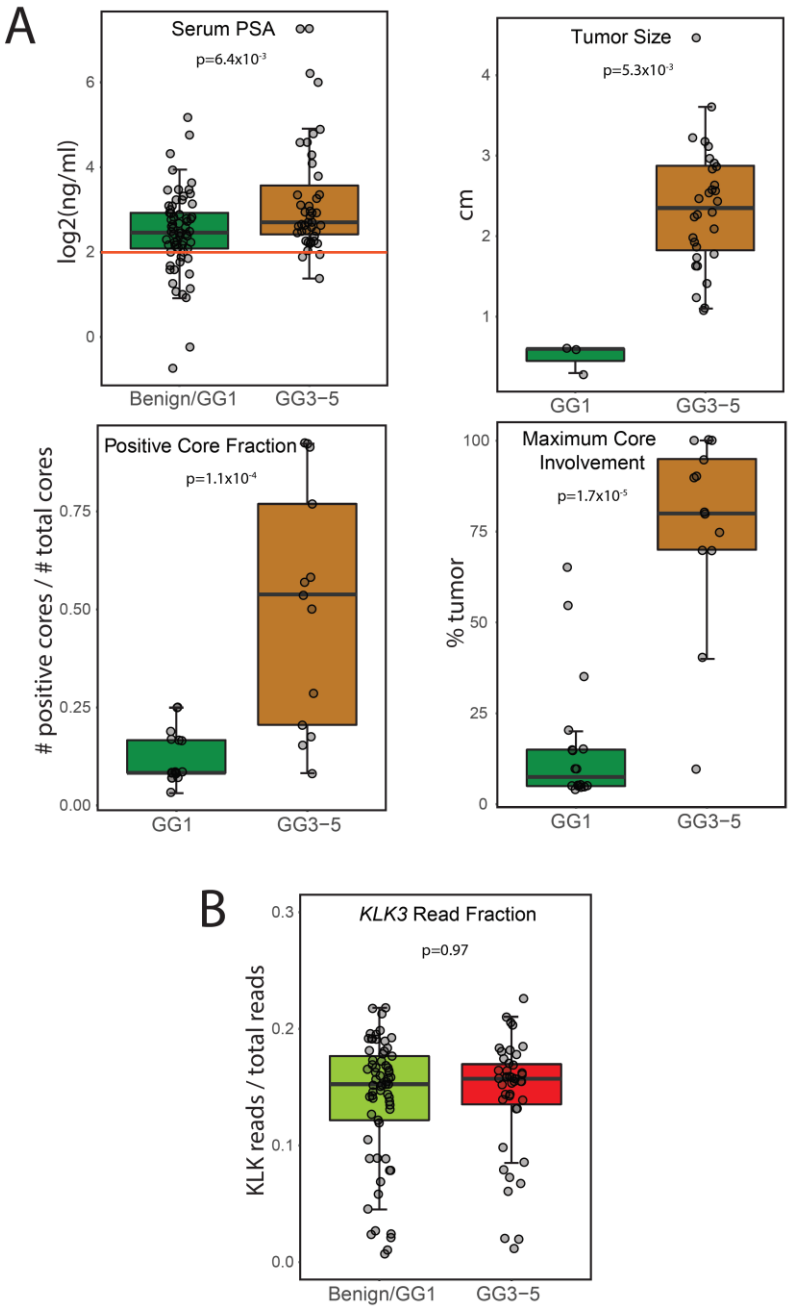
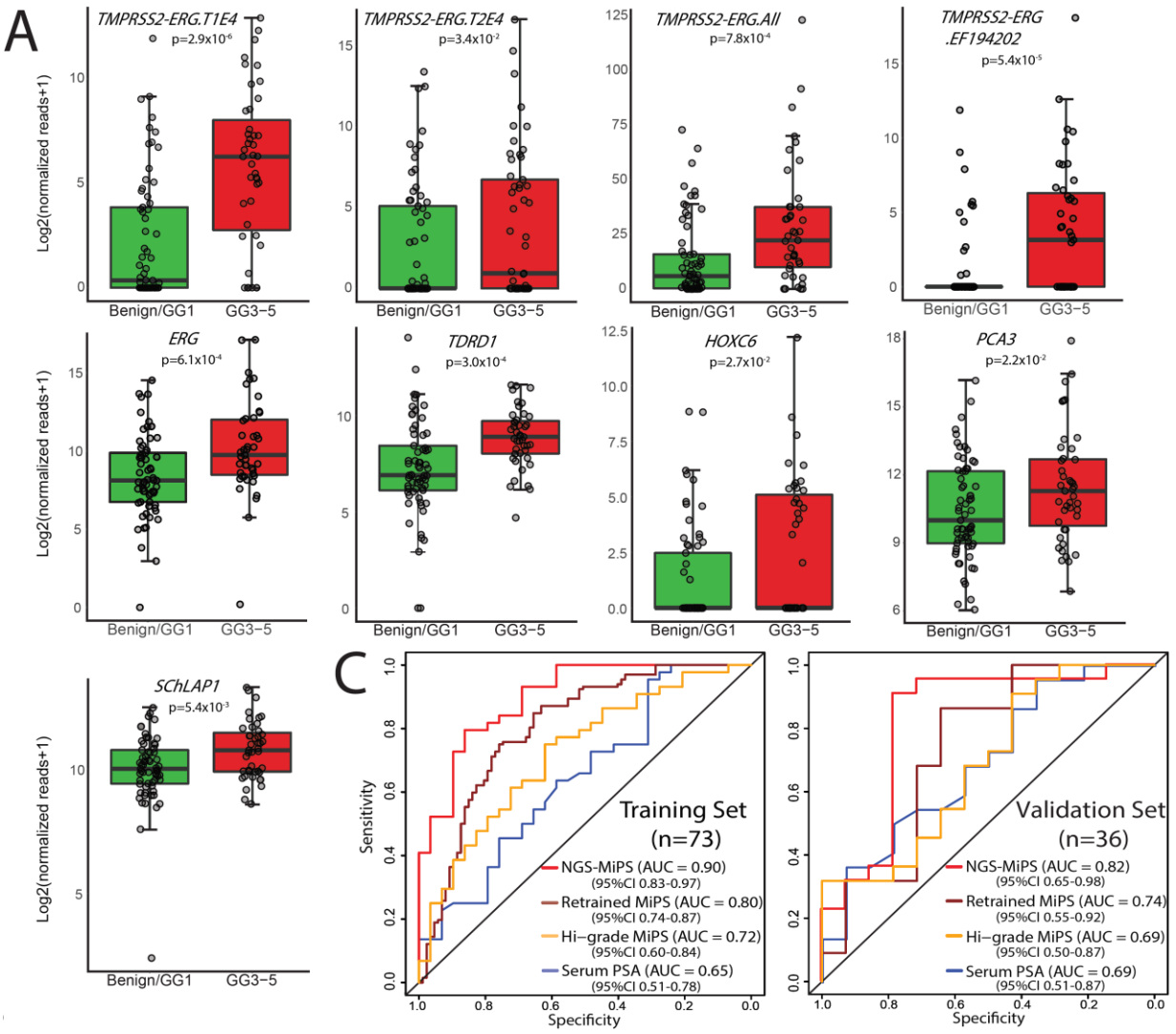


Figure 19. Extreme design cohort clinicopathological characteristics and NGS-MiPS housekeeping gene expression levels. A cohort of 109 patients representing those with benign or Grade Group (GG) 1 vs. those with GG ≥ 3 prostate cancer on biopsy (Extreme Design Cohort) as described in Figure 4, underwent NGS-MiPS. Clinicopathological measures showed the expected differences between the two categories. **A.** Boxplots show serum PSA (in ng/ml); tumor size (in cm, prostatectomy cases only); positive core fraction (fraction of tumor-positive biopsy cores); and maximum core involvement (% tumor in the biopsy core with the highest tumor content). Mann-Whitney test p-values (or Student's T test for normally distributed values) are shown. Serum PSA

boxplot red horizontal line represents the clinically used 4 ng/ml threshold. **B.** *KLK3* expression levels (used as a "housekeeping" gene in NGS-MiPS) normalized by total sequencing coverage show no difference between the two categories. Mann-Whitney test p-value is shown.

NGS-MiPS on Extreme Design Cohort patients yielded nearly identical distributions of urine *KLK3* (PSA) transcript levels between the two groups (**Figure 19B**), supporting its use as a "housekeeping" transcript for NGS-MiPS as in the PROGENSA *PCA3*, MiPS and SelectMDx assays[74, 154]. Importantly, numerous NGS-MiPS targets showed statistically significant differential expression in urine of benign/GG1 vs. GG ≥ 3 patients. Specifically, *PCA3* (p=0.02), *TMPRSS2-ERG.T1E4* (p=0.000003), *ERG* (0.0006), *TDRD1* (0.0003) *and HOXC6* (0.03), transcripts used currently (or previously) in the existing aforementioned clinical urine assays, were more highly expressed in urine of higher biopsy grade cancer patient group than those with benign or low-grade disease as was the aggressive PCa associated lncRNA *SChLAP1*[159] (p=0.005) and others (**Figure 20A**).


A
TMPRSS2-ERG.T1E4
p=2.9x10-6
TMPRSS2-ERG.T2E4
p=3.4x10-2
TMPRSS2-ERG.All
p=7.8x10-4
TMPRSS2-ERG .EF194202
p=5.4x10-5
ERG
p=6.1x10-4
TDRD1
p=3.0x10-4
HOXC6
p=2.7x10-2
PCA3
p=2.2x10-2
SChLAP1
p=5.4x10-3
Log2(normalized reads+1)
Benign/GG1
GG3−5
C
Training Set
(n=73)
NGS-MiPS (AUC = 0.90)
(95%CI 0.83-0.97)
Retrained MiPS (AUC = 0.80)
(95%CI 0.74-0.87)
Hi−grade MiPS (AUC = 0.72)
(95%CI 0.60-0.84)
Serum PSA (AUC = 0.65)
(95%CI 0.51-0.78)
Validation Set
(n=36)
NGS-MiPS (AUC = 0.82)
(95%CI 0.65-0.98)
Retrained MiPS (AUC = 0.74)
(95%CI 0.55-0.92)
Hi−grade MiPS (AUC = 0.69)
(95%CI 0.50-0.87)
Serum PSA (AUC = 0.69)
(95%CI 0.51-0.87)
Sensitivity
Specificity

B

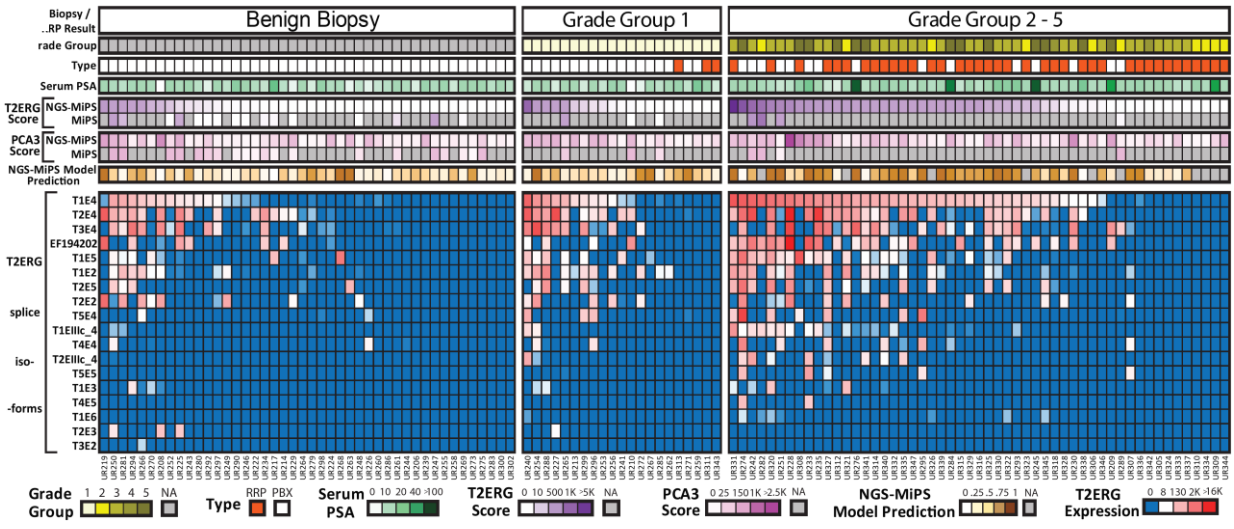

Benign Biopsy
Grade Group 1
Grade Group 2 - 5
Biopsy / ..RP Result
rade Group
Type
Serum PSA
T2ERG Score
NGS-MiPS
MiPS
PCA3 Score
NGS-MiPS Model Prediction
T2ERG
splice
iso-
-forms
T1E4
T2E4
T3E4
EF194202
T1E5
T1E2
T2E5
T2E2
T5E4
T1EIIIc_4
T4E4
T2EIIIc_4
T5E5
T1E3
T4E5
T1E6
T2E3
T3E2
Grade Group
Type
Serum PSA
T2ERG Score
PCA3 Score
NGS-MiPS Model Prediction
T2ERG Expression

Figure 20. NGS-MiPS trained model outperforms serum PSA and derived clinical MiPS models in predicting biopsy results. Urine from a cohort of 109 patients representing those with benign or Grade Group (GG) 1 vs. those with GG ≥ 3 prostate cancer on biopsy (Extreme Design Cohort) was subjected to the NGS-MiPS assay. Sequencing reads for the 84 targets were normalized to sample-specific *KLK3* reads and multiplied by 100,000. **A.** Boxplots for 9 selected targets with significant differential expression between the two groups are shown (median and interquartile range in log2 scale shown). Mann-Whitney test p-values (or Student's T test for normally distributed values) are shown. Boxplots for the rest of the transcripts selected in the trained NGS-MiPS model (see below) are shown in **Figure 21**. **B.** Heatmap shows expression levels for all 18 targeted *TMPRSS2-ERG* splice isoforms (gene-median-centered, log2(normalized reads +1)). Red and blue indicate over- and under-expression, respectively. Headers show tissue pathology results; type of tissue (RRP – radical retropubic prostatectomy, PBX – prostate biopsy); serum PSA (ng/ml); clinical (TMA) and NGS-derived MiPS *TMPRSS2-ERG.T1E4* and *PCA3* scores (calculated as in Figure 3); and the NGS-MiPS model (see below) prediction score from 0 to 1 for having GG ≥ 3 PCa on biopsy. **C.** NGS-MiPS data for all 84 targets from 73/109 Extreme Design Cohort patients (Training Set) randomly selected to have a group-wise ratio proportional to that of the entire cohort, underwent a random forest target reduction method to select the minimal number of the most informative targets (29 transcripts). A regularized logistic regression model was built on the Training Set and showed a higher area under the receiver operating characteristic (ROC) curve (AUC = 0.9) compared to serum PSA or the clinical high-grade (Hi-grade) and a retrained MiPS model using derived NGS-MiPS *PCA3* and *TMPRSS2-ERG.T1E4* scores (left panel). The NGS-MiPS model also outperformed these three models in the held out set of 36 samples (Validation Set), (AUC = 0.82), right panel. AUC 95% CIs are shown.

Further, a number of additional *TMPRSS2-ERG* splicing isoforms not assessed by the clinical MiPS test also showed elevated levels in the high-grade group, a subset of whom with very high specificity (**Figure 20A** and **21A**). As shown in the expression-level heatmap of all 18 targeted NGS-MiPS *TMPRSS2-ERG* splice isoforms, T1E4 was the most commonly expressed isoform, mirroring PCa tissue data, followed by T2E4 and other rarer isoforms (**Figure 20B**). Furthermore, a number of isoforms were co-expressed in any individual urine sample with less common ones having lower expression levels, an observation that is also similar to PCa tissue. Notably, a number of benign biopsy cases contained urine *TMPRSS2-ERG* RNA albeit at lower levels and fewer isoforms, consistent with the notion that random systematic or even MRI-guided biopsy may under-sample or entirely miss smaller cancer foci whereas urine biomarker testing may give a readout of the entire prostate[139, 158]. Interestingly, while the isoforms

present in these benign/low grade cancers were limited to low levels of the more common ones, such as T1E4, higher grade cancers also expressed rarer isoforms such as T1EIIIc_4 or T2EIIIc_4 while these were negligible or absent in benign/low grade cases (**Figure 20B**). Taken together, these data suggest potential utility of multiple *TMPRSS2-ERG* splicing isoforms, including less common, moderately expressed ones, as specific urine biomarkers for non-invasive high-grade PCa detection.

To assess the feasibility of developing an optimized risk predictor from the multiplexed NSG-MiPS assay, we used a machine learning approach to develop a pre-biopsy risk predictor using the Extreme Design Cohort NGS-MiPS urine data set. A third of the data (*n=36* patients) proportional to the cohort-wide category ratio (benign/GG1 vs. GG $\geq$ 3), was randomly sampled and set aside, to be later used as a held-out validation set. A random forest feature reduction process on the remaining 2/3 of the Extreme Design Cohort (training set, *n=73*) decreased the number of targets from 84 down to 29 likely to be the most informative and non-redundant. Next, training of a regularized logistic regression model on this training set assigned non-zero weights to 15 of these 29 targets, to yield a model which included several *TMPRSS2-ERG* splice isoforms, other currently used clinical biomarkers, and additional mRNAs and lncRNAs (expression level distributions and weights for the complete list of model targets are shown in **Figure 21A** boxplots). A simple unweighted unsupervised hierarchical clustering of the Extreme Design Cohort dataset limited to these 15 targets suggested potential utility of these transcripts in discerning the benign/GG1 cases from high-grade ones (as well as was able to cluster related targets such as *TMPRSS2-ERG* isoforms or *AR* splicing variants adjacent to each other; heatmap is shown in **Figure 21B**).

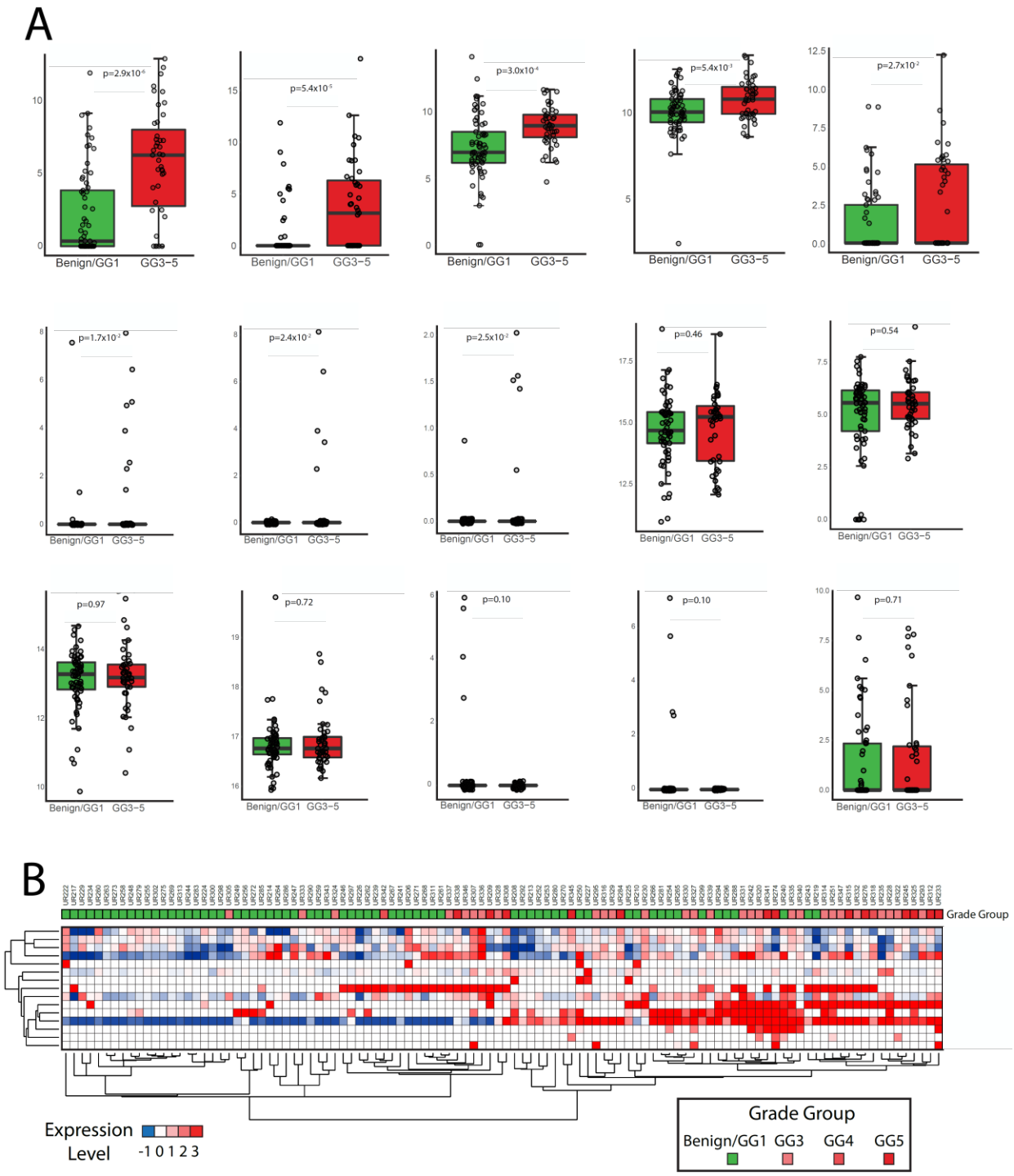


Figure 21. Expression levels for NGS-MiPS preliminary model targets in the extreme design cohort. A. Expression-level boxplots for the 15 transcripts selected in the trained NGS-MiPS model (Figure 20C) are shown. Urine from a cohort of 109 patients representing those with benign or Grade Group (GG) 1 vs. those with GG ≥ 3 prostate cancer on biopsy (Extreme Design Cohort) was subjected to the NGS-MiPS assay. Raw sequencing reads for each target were normalized to sample-specific raw *KLK3* reads and multiplied by 100,000. Boxplots show the interquartile range and

median in log2 scale. Mann-Whitney test p-values (or Student's T test for normally distributed values) are shown. Model weights (W) for each target are shown. **B.** Heatmap shows expression levels for all 15 NGS-MiPS model targets (gene-median-centered, log2(normalized reads +1)). Red and blue indicate over- and underexpression, respectively. Header shows tissue pathology results by grade group colored as in figure legend. Unweighted, unsupervised hierarchical clustering (uncentered correlation similarity metric and centroid linkage clustering method) by gene and by sample was applied.

Receiver operating characteristic (ROC) curves (**Figure 20C**, training set, left panel) were plotted for serum PSA, the NGS-derived high-grade MiPS model[74], a retrained NGS-derived MiPS model (limited to the three MiPS variables but retrained on the Extreme Design Cohort training set), as well as the full 15-transcript NGS-MiPS trained model. Expectedly, serum PSA showed the poorest discriminatory ability (AUC = 0.65 (95%CI 0.51-0.78). The NGS-derived high-grade MiPS and retrained MiPS models had AUCs of 0.72 and 0.80 (95%CIs 0.60-0.84 and 0.74-0.87, respectively), thus confirming MiPS biomarker and model superiority to serum PSA alone. Our NGS-MiPS model had the highest training set AUC of 0.90 (95%CI 0.83-0.97). Importantly, the 36-patient held-out validation set showed a similar AUC distribution to the training set where serum PSA, high-grade MiPS and retrained MiPS had AUCs of 0.69 (95%CI 0.51-0.87), 0.69 (95%CI 0.50-0.87) and 0.74 (95%CI 0.55-0.92), respectively. As in the training set, the 15-transcript NGS-MiPS model was the strongest performing, with a validation set AUC of 0.82 (95%CI 0.65-0.98) supporting the potential value of additional urine transcripts beyond *PCA3* and *TMPRSS2-ERG.T1E4* in PCa early detection using urinary biomarkers (**Figure 20C**, validation set, right panel).

Finally, we used the Extreme Design Cohort to assess the performance of our 15-transcript NGS-MiPS model compared to NGS-MiPS-derived, retrained models using the transcripts in the SelectMDx (*HOXC6 + DLX1*)[154] and ExoDX Prostate IntelliScore (*PCA3 + ERG*)[155] urine tests described earlier. AUCs for the 15-transcript NGS-MiPS models were

higher than the retrained NGS-MiPS derived *HOXC6+DLX1* and *PCA3+ERG* models in both our training and validation sets [Training set: NGS-MiPS AUC = 0.90 (95%CI 0.83-0.97) vs. *HOXC6+DLX1* AUC = 0.69 (95%CI 0.57-0.81) and *PCA3+ERG* AUC = 0.69 (95%CI 0.57-0.81); Validation set: NGS-MiPS AUC = 0.82 (95%CI 0.65-0.98) vs. *HOXC6+DLX1* AUC = 0.50 (95%CI 0.50-0.50) and *PCA3+ERG* AUC = 0.66 (95%CI 0.47-0.85), **Figure 22**]. Although the latter two were only models retrained by us using our moderately-sized cohort, used NGS-derived expression levels as opposed to these assays' actual RT-qPCR method, and, in the case of ExoDx, used whole urine as opposed to that test's intended substrate, urinary exosomes, our data suggest superiority of our biomarker set over currently used ones. Taken together these findings support the potential utility of our non-invasive multi-transcript urine biomarker strategy which promises to improve the early detection of aggressive PCa over currently available serum and urine biomarker-based approaches.

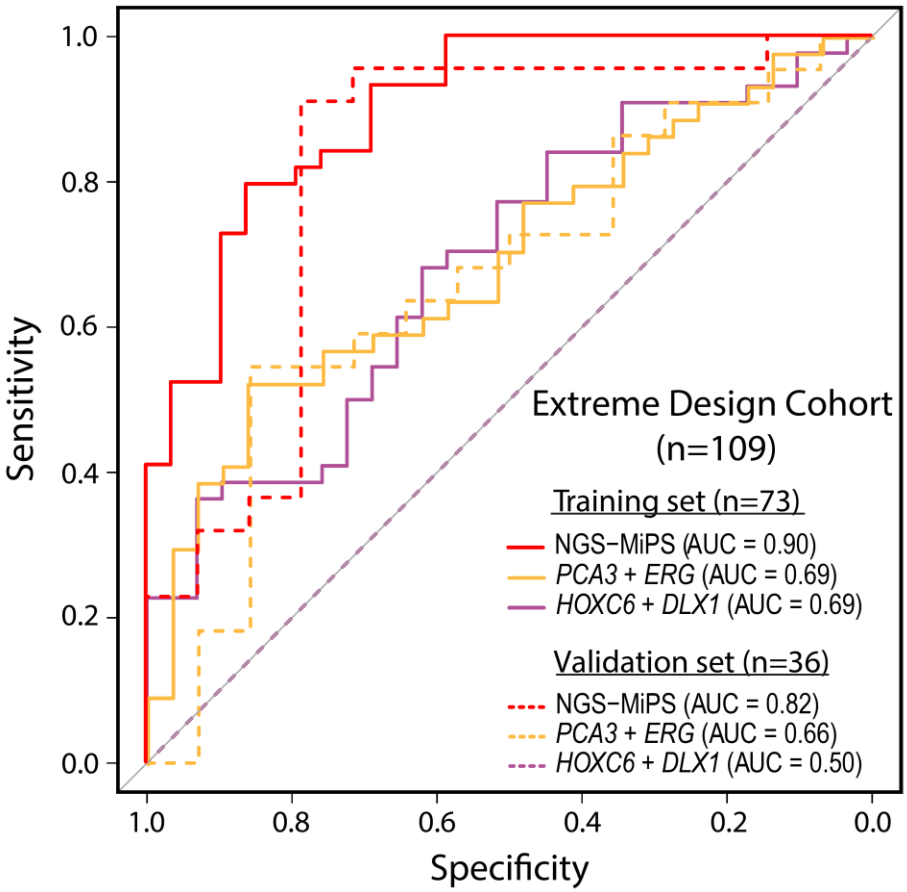


Figure 22. NGS-MiPS vs. derived retrained SelectMDx and ExoDx Prostate Intelliscore models. Models using *HOXC6 + DLX1* (biomarkers comprising the SelectMDx assay) and *PCA3 + ERG* (ExoDx Prostate IntelliScore assay) were trained using the NGS-MiPS Extreme Design Cohort training data set (*n=73*) and were applied to its validation set (*n=36*). Receiver operating characteristic (ROC) curves for predicting pathology results for training (filled lines) and validation (dashed lines) sets were plotted and areas under the curve (AUCs) were calculated for each model.

3.4.5 NGS-MiPS performance in the active surveillance (AS) setting

Given the extreme design cohort results, we next explored our model's performance in a pilot AS cohort, a much more difficult early detection/prognostic scenario. AS is an observation-based management strategy for low-grade PCa that centers on treatment delay and regular prostate biopsy assessment in order to prevent overtreatment of disease likely to be indolent.

Urine from a cohort of 45 AS patients (collected prior to on-AS biopsy) was subjected to the NGS-MiPS assay. Patients were categorized as having benign biopsies (1), or very low (2), low (3) and intermediate (4) risk prostate cancer on biopsy according to National Comprehensive Cancer Network (NCCN) AS guidelines. Patients were then classified into two categories as group 1 + 2 (Benign/Very Low) vs. 3 + 4 (Low/Intermediate). Risk probabilities of being in each of the two biopsy categories above, calculated using urine NGS-MiPS data (or serum PSA) and the corresponding ROCs were plotted for the 4 models compared in Figure 4 earlier (**Figure 23A, B**). Serum PSA exhibited expectedly poor discriminating ability between the two categories in this challenging clinical scenario, with an AUC of 0.53 (95%CI 0.35-0.71). High-grade and retrained MiPS models also performed unsatisfactorily in this patient cohort with AUCs of 0.56 (95%CI 0.37-0.75) and 0.43 (95%CI 0.27-0.65), respectively, as did the all-grade MiPS model (**Figure 24**). However, our 15-transcript NGS-MiPS model, although trained to distinguish the more extremes of PCa aggressiveness spectrum, displayed promising signal with statistically significant differences in risk prediction between the two categories (Benign/Very Low vs. Low/Intermediate) (**Figure 23A**) and an AUC of 0.66 with the entire 95%CI range being at or above 0.50 (95%CI 0.50-0.83, **Figure 23B**). Although our AS cohort contained only a limited number of patients upgraded on biopsy, thus precluding formal assessment of NGS-MiPS performance in this key event in the AS patient's possible clinical course, taken together, these data support the potential and continued investigation of our multi-biomarker NGS-MiPS assay in this patient setting where effective biomarkers that can aid in clinical decision-making are much urgently needed.

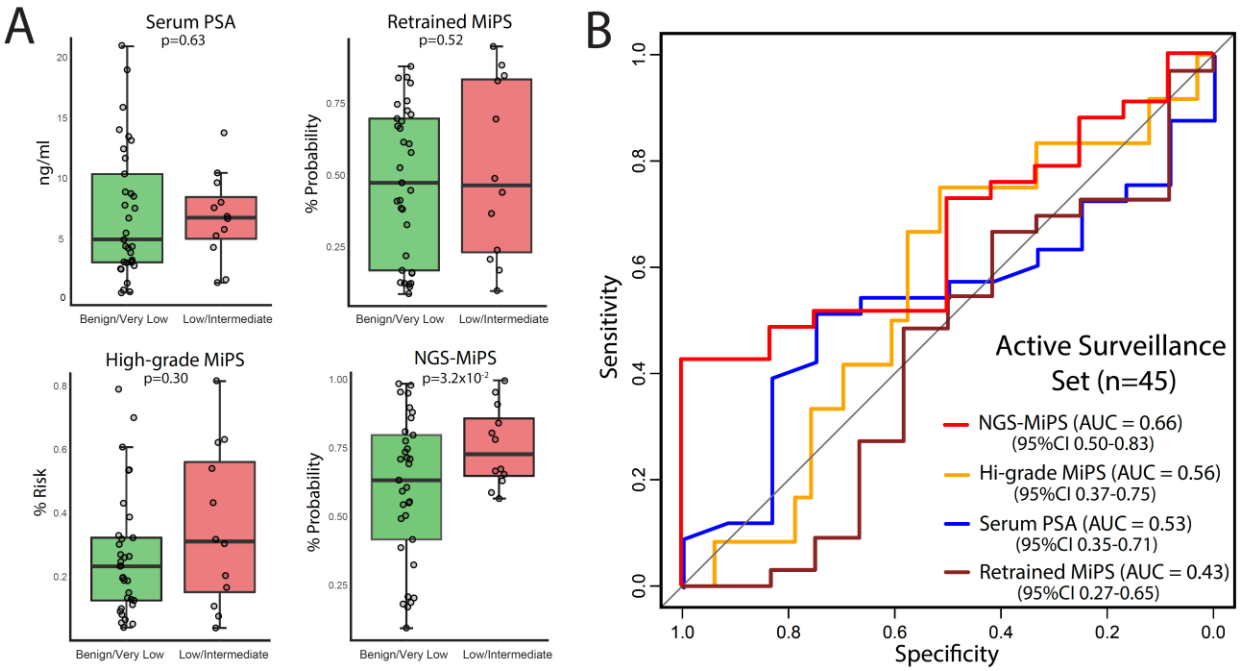


Figure 23. NGS-MiPS trained model detects differences in the active surveillance setting. Urine from a cohort of 45 patients on prostate cancer active surveillance (AS, an observation based management strategy for patients with low grade/indolent prostate cancer on biopsy) was subjected to the NGS-MiPS assay. Patients were categorized as having benign biopsies (1), or very low (2), low (3) and intermediate (4) risk prostate cancer (based on AS biopsy pathology) according to NCCN AS guidelines. Patients were then classified into two groups as category 1 + 2 (Benign/Very Low) vs. 3 + 4 (Low/Intermediate). **A.** Boxplots for serum PSA (ng/ml), and risk/probability scores for having high grade prostate cancer on biopsy are shown for clinical high-grade and retrained MiPS models as well as our trained 15-transcript NGS-MiPS model. All risk/probability scores were calculated using NGS-derived *PCA3* and *TMPRSS2-ERG.T1E4* scores. Boxplots represent the interquartile range and median in log2 scale. Mann-Whitney test p-values (or Student's T test for normally distributed values) are shown with the NGS-MiPS model showing the greatest discriminating power. **B.** ROC curves (with AUCs and 95%Cis indicated) show the trained NGS-MiPS model outperforming the other three models in panel A. in this AS patient cohort.

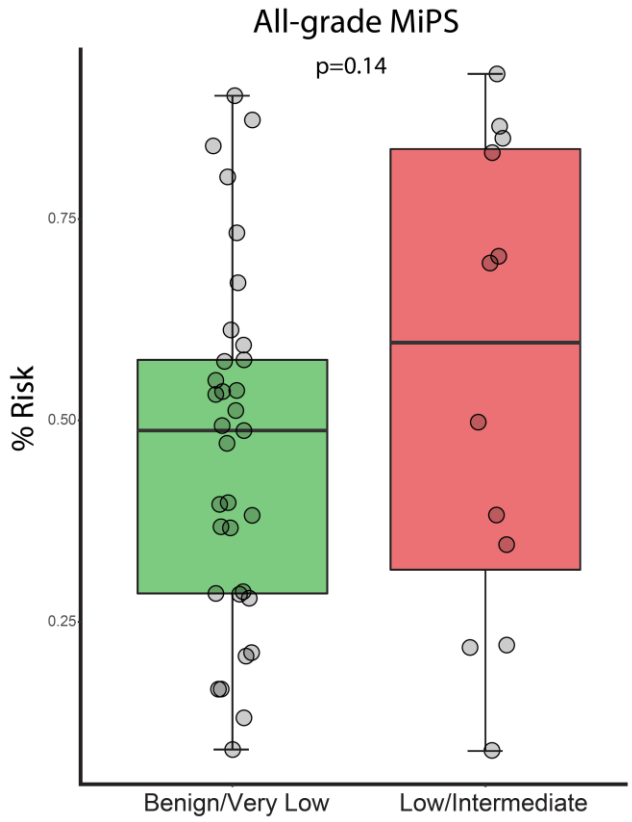


Figure 24. All-grade PCa MiPS model prediction in the active surveillance (AS) cohort. Urine from a cohort of 45 patients on prostate cancer AS (from Figure 23) was subjected to the NGS-MiPS assay. Boxplot of % risk scores for having PCa of any-grade on biopsy for the clinical MiPS model is shown. All risk scores were calculated using NGS-derived *PCA3* and *TMPRSS2-ERG.T1E4* scores. Boxplot represents the interquartile range and median in log2 scale. Student's T test p-value is shown.

3.4.6 NGS-MiPS detects expressed PCa predisposing germline variants and somatic driver mutations

Common germline SNPs have shown limited ability to improve upon serum-based PCa risk predictors[176], and the lack of highly recurrent somatic mutations has limited the performance of somatic DNA-based PCa early detection approaches. Indeed, with the exception of the *TMPRSS2-ERG* gene fusion, none of the currently offered urine (or tissue) assays target

such driver events. The PCa genomic landscape however may provide several potentially useful candidate early detection biomarkers. The *HOXB13* p.G84E risk variant has been associated with increased risk of PCa (1.4% frequency in PCa patients vs. 0.02% in healthy population), especially in familial early onset cases[177, 178]. Using our FFPE tissue-based RNA NGS approach, we previously demonstrated the ability to detect *HOXB13* p.G84E in tumors from patients known to harbor this germline alteration, as well as detection of somatic mutations (such as those in *SPOP, NRAS, BRAF*)[158]. Hence, to determine whether such genomic events could similarly be detected in RNA from urine, we included in NGS-MiPS amplicons targeting *HOXB13* p.G84E and the most common SPOP hotspot somatic mutations. Although germline *HOXB13* p.G84E status was not independently known for any patients in our cohort, NGS-MiPS successfully detected the *HOXB13* G84E variant in the urine RNA of one patient (UR_B, **Figure 25**) at a variant allele frequency (VAF) of 42%, consistent with a potential heterozygous germline SNP (assuming approximately equal expression from the mutant and wild type alleles). We are confirming the variant in the sequencing library and urine RNA of this patient by Sanger sequencing. As no biospecimen other than urine was available from this patient, we used 1.5 ml of urine in Progensa media to isolate DNA (yield, 1.26 ng). We are confirming this mutation in this patient's genomic urine DNA sample by Sanger sequencing.

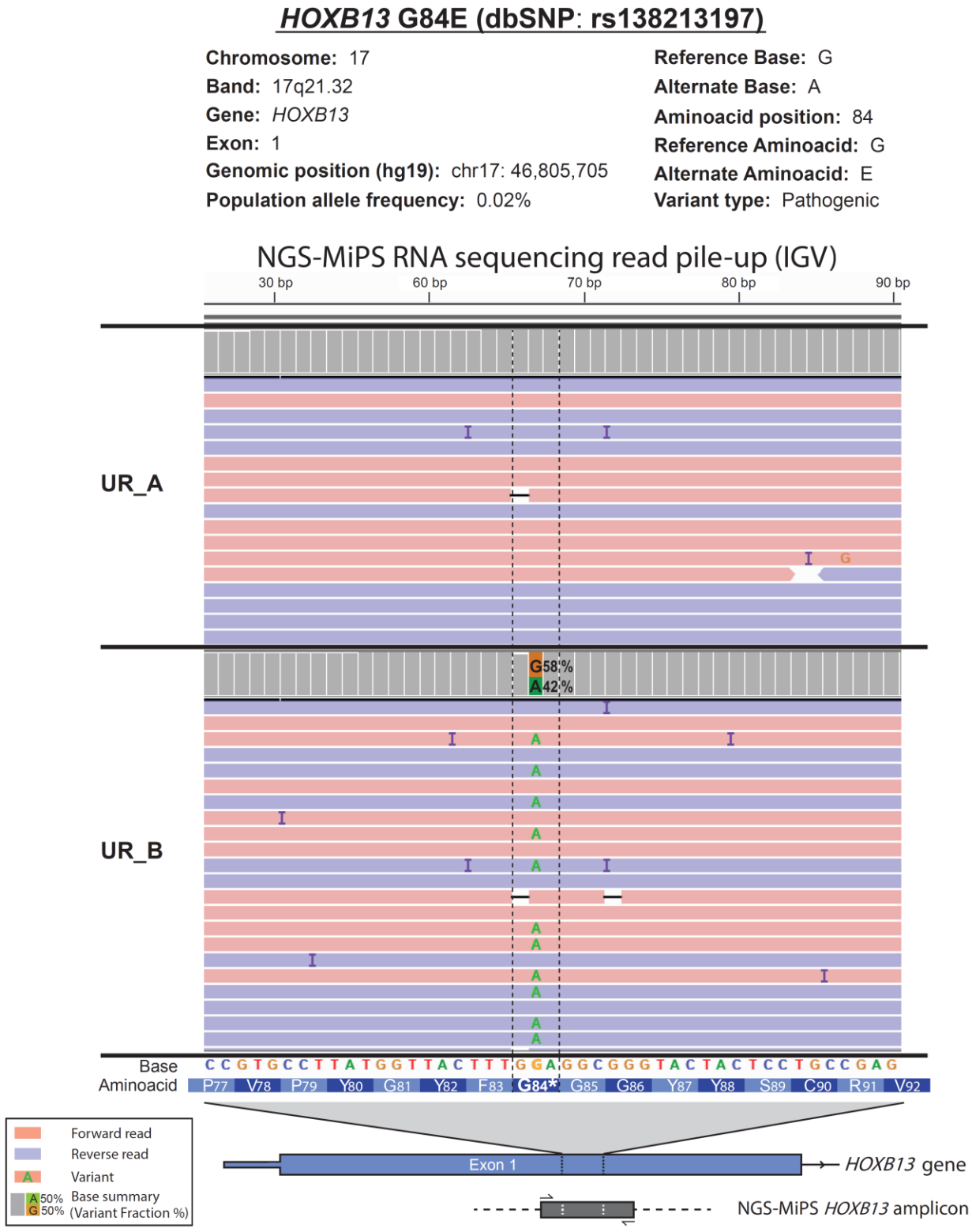
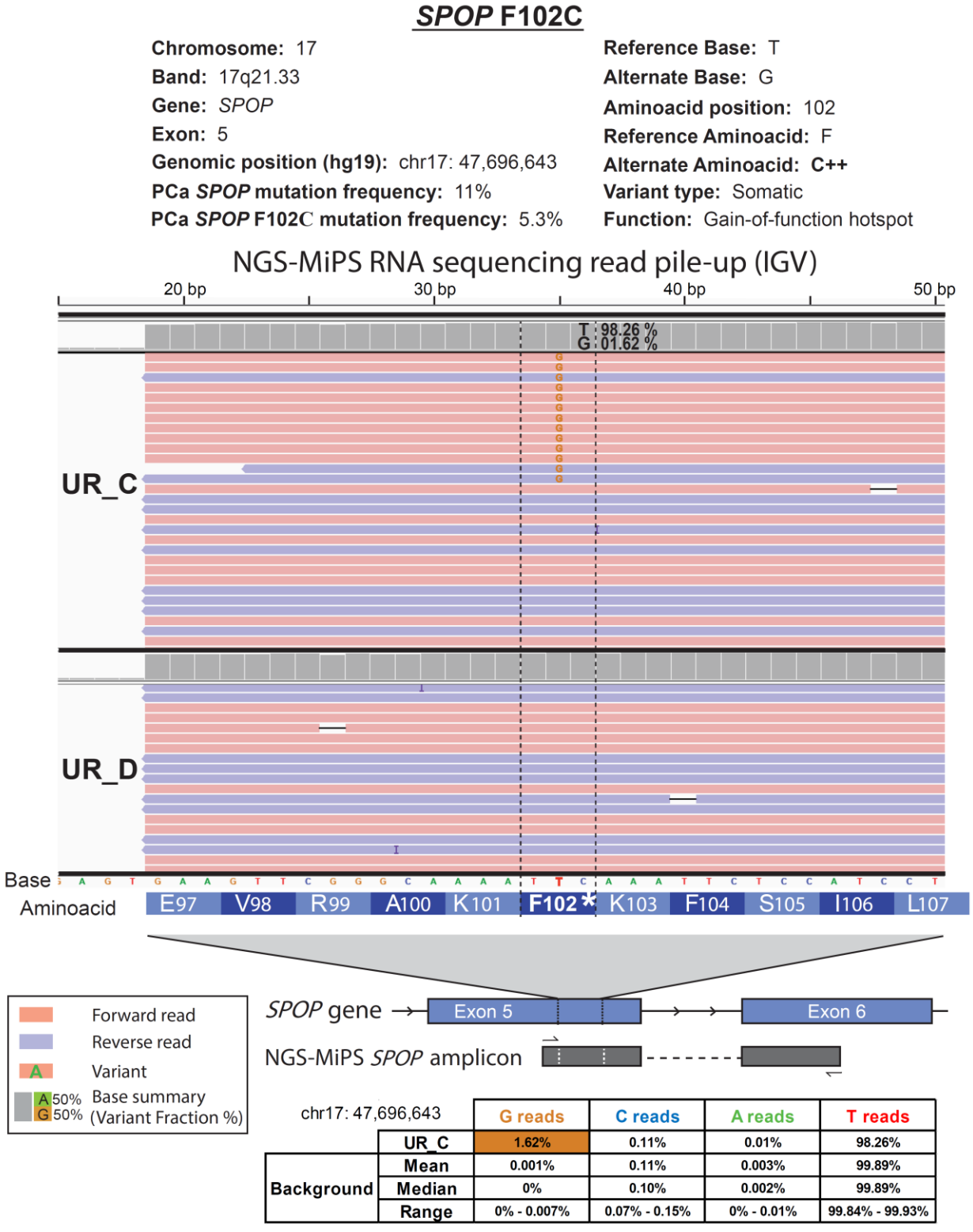


Figure 25. Detection of a *HOXB13* p.G84E prostate cancer predisposing germline variant by NGS-MiPS. Figure shows detection by NGS-MiPS of an expressed germline SNP (population allele

frequency of 0.02%) in a patient urine sample. RNA sequencing read pile-up from Integrated Genome Viewer (IGV) is shown for two samples (UR_A, wild-type; UR_B, variant carrier). Header for each sample (gray bars) shows a summary of that position over the entire read set covering it. Patient UR_B carries the expressed variant at near half of all reads (42%) indicating a heterozygous SNP assuming approximately equal expression from each allele. Both forward (pink) and reverse (blue) strand reads show presence of the SNP. Corresponding *HOXB13* exon 1 aminoacid position (p.G84) as well as NGS-MiPS *HOXB13* target amplicon position are shown (bottom).

As mentioned above, NGS-MiPS also targets expressed somatic driver mutations in the hotspot-rich MATH-domain region of the E3 ubiquitin ligase SPOP, the most highly recurrent mutated gene in PCa (frequency = 11%)[39]. Although *SPOP* mutational status was not independently known for any patients in this cohort, NGS-MiPS was also able to successfully detect two *SPOP* hotspot somatic mutations, p.F102C and p.F125I in two different PCa patients' urine samples (UR_C and UR_D, **Figures 26** and **27**, respectively). These mutations exhibited low VAFs (1.62% and 0.42%, respectively), consistent with the expected low overall fraction of tumor derived *SPOP* RNA in the urine due to high *SPOP* expression of in benign prostate and urothelium[179] (**Figure 28**), However these were high-confidence mutational calls as their VAFs were well above the low background variant rates for those positions (tables in bottom of **Figures 26** and **27**). Negligible *TMPRSS2-ERG* sequencing reads were present in these samples, consistent with mutual exclusivity between *SPOP* mutations and ETS gene fusions[39, 44]. Given these somatic mutations' low urine VAF, we are confirming these mutations in sequencing library, urine RNA and DNA with Sanger sequencing by employing a method that uses wild-type allele PCR suppression strategy using 3'dideoxy-modified primers while allowing extension of the mutant 3'unmodified primer (QuanTAS-PCR [180]).



chr17: 47,696,643		G reads	C reads	A reads	T reads
	UR_C	1.62%	0.11%	0.01%	98.26%
Background	Mean	0.001%	0.11%	0.003%	99.89%
	Median	0%	0.10%	0.002%	99.89%
	Range	0% - 0.007%	0.07% - 0.15%	0% - 0.01%	99.84% - 99.93%

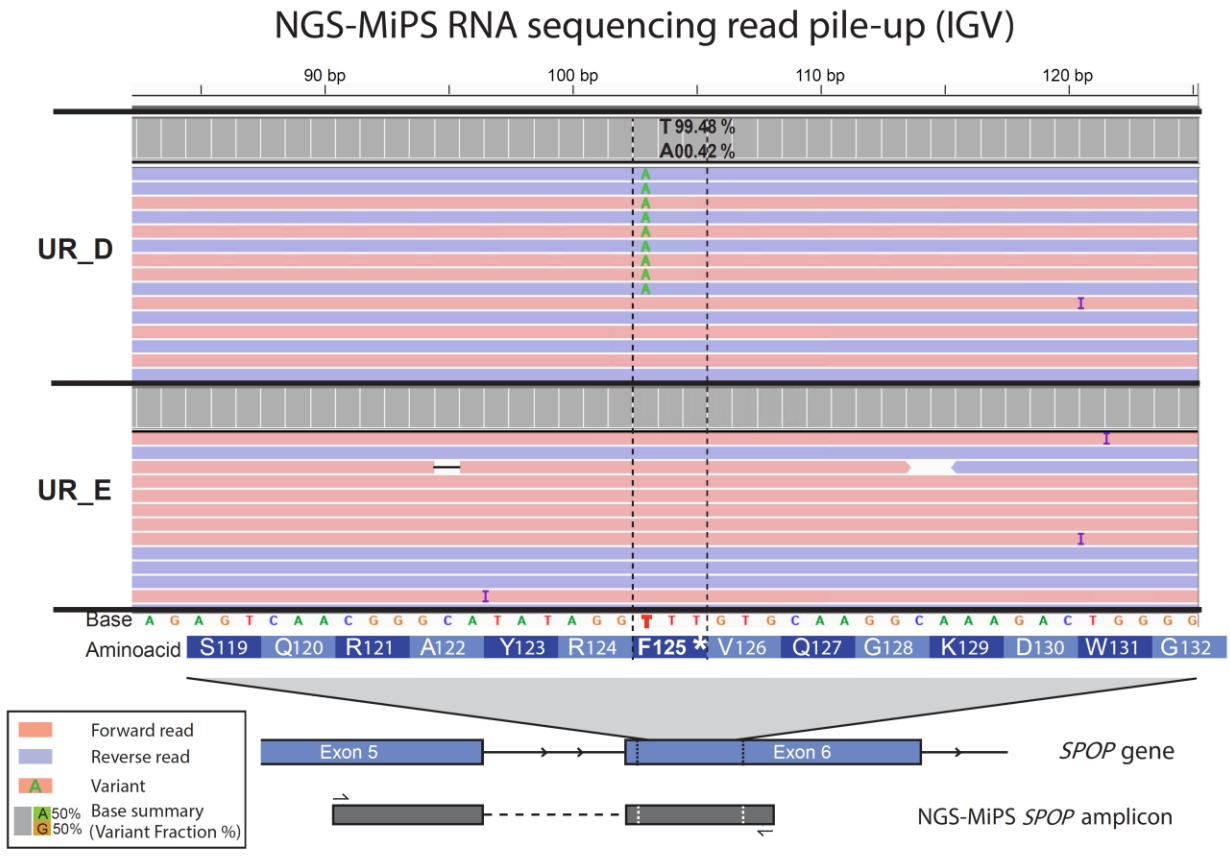
Figure 26. Detection of an *SPOP* F102C prostate cancer hotspot somatic mutation by NGS-MiPS. Figure shows detection by NGS-MiPS of an expressed *SPOP* hotspot somatic mutation in a

patient urine sample. RNA sequencing read pile-ups from Integrated Genome Viewer (IGV) are shown for two samples (UR_C, mutant; UR_D, wild type, for *SPOP* F102C). Header for each sample (gray bars) shows a summary of that position over the entire read set covering it. Mutant and wild type read sections are shown for patient UR_C. Patient UR_C carries the expressed mutation at 1.62% variant allele frequency indicating a somatic mutation. Both forward (pink) and reverse (blue) strand reads show presence of the mutation. Corresponding *SPOP* exon 5 aminoacid position (p.F102) as well as NGS-MiPS *SPOP* target amplicon position are shown below IGV image. Table in bottom, shows VAFs at that base position for variants of each nucleotide for sample UR_C as well as background variant read distribution as mean, median and range over 10 randomly selected urine samples. *SPOP* p.F102C mutation VAF >> background VAF.

SPOP F125I

Chromosome: 17
Band: 17q21.33
Gene: *SPOP*
Exon: 6
Genomic position (hg19): chr17: 47,696,450
PCa *SPOP* mutation frequency: 11%
PCa *SPOP* F125I mutation frequency: 2.6%

Reference Base: T
Alternate Base: A
Aminoacid position: 125
Reference Aminoacid: F
Alternate Aminoacid: I
Variant type: Somatic
Function: Gain-of-function hotspot



chr17: 47,696,450		G reads	C reads	A reads	T reads
	UR_D	0.06%	0.03%	0.42%	99.48%
Background	Mean	0.06%	0.03%	0.003%	99.91%
	Median	0%	0.03%	0.002%	99.91%
	Range	0.01% - 0.11%	0.006% - 0.05%	0% - 0.02%	99.85% - 99.96%

Figure 27. Detection of an *SPOP* p.F125I prostate cancer hotspot somatic mutation by NGS-MiPS. Figure shows detection by NGS-MiPS of an expressed *SPOP* hotspot somatic mutation in a patient urine sample. RNA sequencing read pile-ups from Integrated Genome Viewer (IGV) are shown for two samples (UR_D, mutant; UR_E, wild type, for *SPOP* p.F125I). Header for each sample (gray bars) shows a summary of that position over the entire read set covering it. Mutant and

wild type read sections are shown for patient UR_D. Patient UR_D carries the expressed mutation at 0.42% variant allele frequency indicating a somatic mutation. Both forward (pink) and reverse (blue) strand reads show presence of the mutation. Corresponding *SPOP* exon 5 aminoacid position (p.F125) as well as NGS-MiPS *SPOP* target amplicon position are shown below IGV image. Table in bottom, shows VAFs at that base position for variants of each nucleotide for sample UR_C as well as background variant read distribution as mean, median and range over 10 randomly selected urine samples. *SPOP* p.F125I mutation VAF >> background VAF.

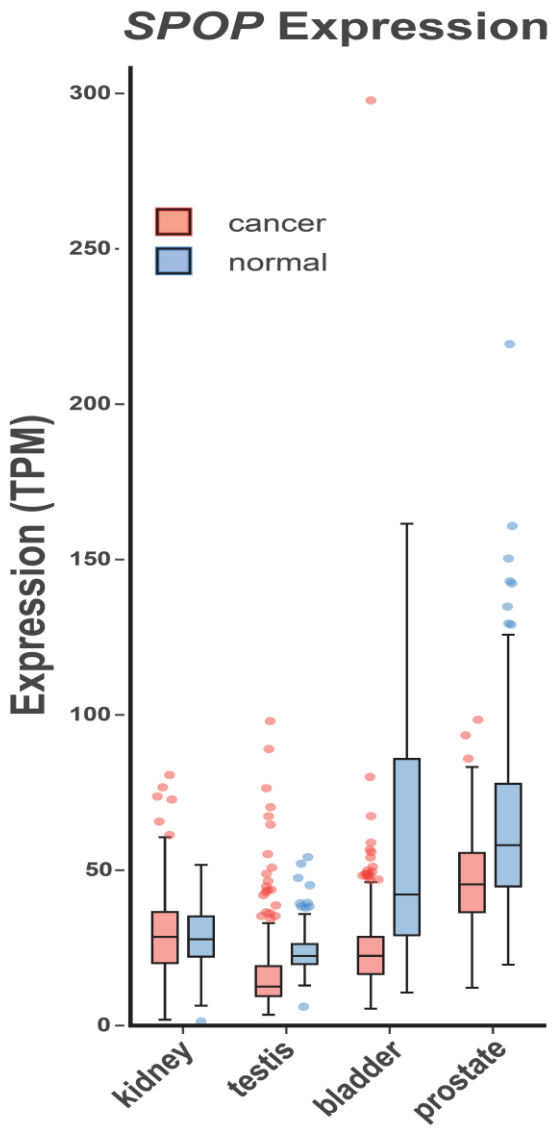


Figure 28. *SPOP* expression in cancer and normal tissues of the urinary tract. Boxplots of *SPOP* expression levels in the cancer and normal tissues presumed to be the major contributors to the urinary tract. Boxes represent interquartile range with median value and are plotted in transcripts per million (TPM). Data were obtained using the MiPanda tool (mipanda.org). Higher expression levels in normal tissues compared to cancer are observed for bladder and prostate.

Taken together, these data support NGS-MiPS' ability to detect expressed PCa predisposing germline variants and somatic mutations in urine, a novel approach in the PCa urinary biomarker field. Larger training cohorts can determine whether these genomic alterations as well as other rarer non-*TMPRSS2-ERG* PCa gene fusions can be combined with the biomarkers selected earlier in our preliminary model into improved models for aggressive PCa early detection with high specificity and sensitivity.

3.5 Discussion

3.5.1 Overall review of findings

We demonstrated the potential utility of our whole-urine RNA NGS assay, NGS-MiPS, in the non-invasive detection of high grade PCa. NGS-MiPS was highly concordant with the TMA-based clinical-grade MiPS assay for the two MiPS transcripts. Furthermore, NGS-MiPS comprises a broad additional target set of 84 total transcripts that upon training in an extreme design patient cohort yielded a 15-transcript model that outperformed current biomarkers/models in predicting tissue pathology results. Additionally, NGS-MiPS offers novel aspects including detection of an expressed germline PCa risk SNP (*HOXB13* p.G84E) and expressed somatic PCa driver mutations in *SPOP*. These data suggest continued pre-clinical and clinical development of our assay to guide initial PCa biopsy decisions as well as potential benefit in other settings such as patients with negative MRI findings or negative biopsies, those on active surveillance, identification of patients/families with potential inherited risk alleles, tumors with targetable or risk-stratifying somatic alterations etc.

Our preliminary 15-transcript urine model was trained to predict tissue pathology results in a 73-patient cohort representing the ends of disease spectrum (Extreme Design Cohort). As discussed earlier, urine following a thorough DRE is expected to contain information from all regions of the prostate as opposed to the more localized approach of tissue core biopsies. This may overcome issues of molecular and histological intratumoral heterogeneity and multi-focality that pathological tissue evaluation may under-sample or entirely miss. Despite using tissue biopsy pathology results as the training gold standard, our 15-transcript model had a higher AUC (0.82) than both serum PSA (0.69) and the high-grade clinical MiPS (0.69) models (using NGS-assessed expression levels) in predicting pathology results in the 36-patient held-out validation set. We also allowed for retraining the MiPS model in our Extreme Design Cohort using its variables in order to eliminate any cohort-specific effects in the comparison. The retrained MiPS model, although having higher AUC (0.74) than MiPS, still underperformed our 15-transcript model. This supports potential improved utility in using biomarkers beyond *PCA3* and *TMPRSS2-ERG*, (as well as fusion splicing isoforms beyond *TMPRSS2-ERG.T1E4*), in detecting the presence of aggressive, high-grade PCa from urine. Furthermore, we used our Extreme Design Cohort to retrain models using NGS-MiPS-measured expression of transcripts that are part of two clinically available urine PCa tests, SelectMDx (*HOXC6+DLX1*) and ExoDx Prostate IntelliScore (*PCA3+ERG*). Although we used retrained models, and in the case of ExoDx, used our whole-urine RNA data as opposed to its intended substrate, urine exosomal RNA, our 15-transcript model's AUC (0.82) was higher than either of these retrained models (0.50 and 0.66, respectively), suggesting that additional biomarkers can improve upon those included in current clinically-used tests.

Given NGS-MiPS' encouraging results in our Extreme Design Cohort we also explored

whether NGS-MiPS could detect any discriminatory signal in active surveillance patients, who usually fall in GG1 (and in some cases, benign or less aggressive end of GG2) region of the spectrum. These patients follow a strategy of regular (usually bi-yearly) prostate biopsies and delayed treatment until biopsy results show evidence that their disease has advanced into a higher grade. There is an urgent need for non-invasive biomarkers that can identify patients at high risk of upgrading in order to avoid delaying the required treatment, while confirming stable indolent disease in other patients in order to alleviate patient anxiety. Importantly, although trained to discern the more extreme ends of the PCa spectrum, our 15-transcript model, had higher AUC (0.66) than serum PSA (0.53) and the existing high-grade (0.56) and retrained (0.43) MiPS models in this patient population. Notably, our model's entire AUC 95%CI range was at least 0.50, supporting promise of our approach in the active surveillance setting.

Our assay design strategy centered on casting a wide net for candidate biomarkers which included validated transcripts currently in clinical use as well as many from tissue clinical assays and non-clinical studies. We expanded our target set to assay most reported splice isoforms of the *TMPRSS2-ERG* gene fusion, the most highly recurrent PCa alteration (as opposed to only assessing for the most commonly expressed isoform, T1E4, as in MiPS), a novel approach in the field. We also assay several isoforms each of many non-*TMPRSS2-ERG* fusions involving other androgen driven genes (e.g. SLC45A3) and/or other Ets transcription factor family members (e.g. ETV1, -4, -5)[67] as well as non-Ets gene fusions involving oncogenic Raf family members. Being less recurrent, these latter biomarkers will require larger training patient cohorts in order to show their potential utility. In addition, we included several recently reported aggressive PCa-associated lncRNAs (e.g. *SChLAP1* and *ARLNC1*) in addition to the already clinically used *PCA3*. Lastly, in another innovative approach, we targeted and successfully

detected at the RNA level, expressed germline PCa-predisposing genomic variants (*HOXB13* p.G84E) and cancer driving hotspot somatic mutations (e.g. *SPOP* p.F102C and p.F125I). These genomic biomarkers may also improve upon the potential utility of our assay in larger training cohorts. Importantly, all NGS-MiPS target primers/amplicons were previously qualified in tissue in a large study of FFPE tissue RNA samples (*n=193*) as part of our tissue-based PCa RNA multiplexed NGS assay[158].

3.5.2 Urine-based prostate cancer assay design considerations

In order to have the most accurate signal and avoid confounding information from non-prostatic GU or other tissue types, we prioritized transcripts expressed in prostate, PCa and/or aggressive PCa for our final assay design. This included the transcripts described above, but excluded potentially useful targets such as cell proliferation markers for the purposes of this assay. Their use in urinary PCa detection warrants further investigation. Given the relatively large number of post filtering targets (*n=84*), we used a targeted RNA NGS approach based on the Ion Torrent Ampliseq platform leveraging our expertise in the field. The only literature example of a many-biomarker PCa urine assay to our knowledge was the recently reported assay by Connell et al.[181], an exosomal RNA test using the NanoString platform. That report's targets modeled to stratify PCa outcome risk, only overlapped with NGS-MiPS in a subset of the current clinically used biomarkers (e.g. *HOXC6, TMPRSS2-ERG.T1E4*, *ERG*, *PCA3*). Furthermore, its performance compared to current assays (or their derived versions) remains to be shown.

Our use of whole-urine collected as "first-catch" immediately following a DRE is

supported in findings by Hendriks et al.[173] as being superior to urinary sediments or exosomes for detection of prostate/PCa-derived transcripts. Furthermore, whole-urine has been the method of choice in the FDA approved PROGENSA *PCA3* assay and the SelectMDx and MiPS laboratory-developed tests (LDT). DRE's effect of enriching for prostate contents in urine was confirmed by our comparison of matched pre- and post-DRE samples where we saw a somewhat diminished ability to accurately detect some less strongly expressed transcripts in normal non-DRE urine. Low-expression transcripts include important clinically validated biomarkers such as *TMPRSS2-ERG*, *DLX1*, *HOXC6* etc. and the difficulty of assessing them in pre-DRE urine may be overcome by increasing sequencing depth and/or future NGS-MiPS assay iterations that exclude unnecessary highly-expressed transcripts in order to allow more sequencing depth for more scantly expressed ones.

We extensively tested and successfully showed reproducibility of our assay using identical replicates as well as robustness to varying optimal and suboptimal experimental conditions. The vast majority of urine samples passed quality control filters for a current informative sample rate of 98%. Importantly, NGS-MiPS accurately measured expression levels as compared to orthogonal methods such as RT-qPCR and the TMA-based predecessor LDT, MiPS, which has been implemented in our CLIA certified, CAP accredited, NY-State-approved clinical laboratory. These data support high overall accuracy, as previously shown using the same amplicons and approach in FFPE tissue samples[158].

3.5.3 Conclusions

Serum PSA is a widely used biomarker for PCa early detection, however, its poor

specificity especially in patients near the clinically used cut off of 4.0ng/mL has led to many unnecessary biopsies and overdetection of low-grade, likely indolent PCa. In addition to other serum based tests incorporating other PSA isoforms or closely related proteins, several urine-based RNA tests have been introduced clinically that are able to improve on serum PSA in detecting aggressive PCa. Importantly, all of these tests assess at most two target genes. Our data support our multi-biomarker urine RNA NGS approach, as exemplified in our preliminary 15-transcript model, as showing the potential to improve upon these current tests supporting the continued development of this assay for guiding initial biopsy decision-making and other settings for highly accurate early detection of aggressive prostate cancer.

CHAPTER IV

Discussion

4.1 Current clinical landscape and need in the sex hormone-driven cancer biomarker field

As mentioned in Chapter I, sex hormone-driven molecular signaling is a central theme in the biology of most breast and prostate cancers. Furthermore, additional molecular alterations are involved in the oncogenesis and progression of these tumors from localized all the way to lethal metastatic disease. These genomic and transcriptomic alterations themselves, and the molecular mechanisms they engender, have served as biomarkers for detection, prognostication and prediction of rational pharmacological targeting of disease in these and other cancers. Tests such as breast cancer OncotypeDx and MammaPrint are tissue expression-level assays measuring RNA for 21 and 70 genes, respectively, and have been used successfully to assess primary breast cancer aggressiveness and prognosticate risk of recurrence as well as predict benefit from chemotherapy[182]. Likewise, OncotypeDx (GPS), GenomeDx (Decipher) and Prolaris (CCP) assays are similar tissue expression-based tests assessing similar numbers of genes and have prognostic value in prostate cancer[183]. In advanced prostate cancer, detection of AR-V7 mRNA, an overactive AR splice variant lacking a ligand binding domain, is predictive of lack of benefit even from next generation androgen deprivation therapies (enzalutamide and abiraterone), but of benefit from chemotherapy[184].

With the advent of the precision oncology era, several oncogenic molecular alterations have entered the clinic as biomarkers predictive of response for various actionable cancer-driving defects in breast cancer. These include the use of trastuzumab (Herceptin) and its variants in tumors overexpressing Her-2; the CDK4/6 inhibitor palbociclib (Ibrance) to target the cell cycle, (although alterations in the involved genes do not seem to confer additional benefit compared to non-altered tumors); the recently-approved inhibitor of mutant PI3 kinase (*PIK3CA* gene) alpelisib (Piqray); and the PARP inhibitors such as olaparib (Lynparza) in patients with germline *BRCA* gene mutations[185-188]. Furthermore, non-invasive liquid biopsy approaches have become common in the clinic including CTC detection in the blood of advanced breast (and prostate) cancer patients, which has been shown to be a prognostic biomarker of progression-free and overall survival[73]. However, CTCs' use beyond cell counts, phenotypic marker staining and some limited genetics, has been not fully explored, yet it remains clinically promising. Likewise, in prostate cancer, non-invasive detection approaches in urine have been coupled with the validation of several transcriptomic biomarkers. These include oncogenic Ets transcription factor *ERG* (alone or fused to androgen-driven gene *TMPRSS2*), the lncRNA *PCA3* and a few others[189]. In advanced disease, PARP inhibitors have also shown success in prostate cancer patients with mutations in *BRCA* and other DNA damage repair (DDR) genes[190]. Although AR amplifications and mutations are genomic alterations conferring resistance to traditional androgen deprivation therapy, they are not used as clinical biomarkers since the majority of patients that have failed ADT will respond to next generation androgen inhibitors enzalutamide (Xtandi) and abiraterone (Zytiga), and are thus treated with them.

The clinical success of the molecular biomarkers mentioned above in sex hormone-driven cancer supports continued development in the field. Thus, as the breast cancer precision

medicine approaches mentioned above, and others that are sure to follow in the future, become important, there will be a need for continued monitoring of the tumor molecular evolution and mechanisms of treatment resistance in real-time. However, tissue biopsies at multiple metastatic sites and multiple time points are likely to continue to remain impractical due to a host of obvious reasons. Another drawback to tissue biopsy sequencing is the limited tumor region sampled by the biopsy needle, which can provide an incomplete picture of tumor heterogeneity[191].

Advances that attempt to address these shortcomings have been made in the detection of circulatory biomarkers shed by the tumor, one of which being tumor-derived circulating cell-free (cf) DNA and the cancer-driving alterations it harbors. This strategy has shown considerable recent success in the clinic. Its use, however, is somewhat limited by the need for high sensitivity in detecting these minute amounts of circulating tumor DNA diluted in large fractions of cf-DNA originating from normal healthy cells elsewhere in the body. The small amount of DNA can limit the comprehensiveness of profiling to the assessment of genomic alterations only. (This may not even include deleterious mutations in large tumor suppressors since they require thorough sequencing of their entire coding regions). And obviously, detection of RNA or phenotypic changes (as in cell staining techniques) are not possible in cf-DNA profiling. As an example, PD-L1 expression remains the main and currently approved predictive biomarker for checkpoint blockade-based immunotherapy (in several malignancies, including triple-negative breast cancer[192]). However, its assessment requires direct detection of PD-L1 protein on the cell surface which is possible only by tissue immunohistochemistry (IHC, the current clinically approved checkpoint inhibitor companion diagnostic) or staining in single cells. Presumably, quantitation of PD-L1 mRNA in tissue or cells would be an equivalent approach to this. And in

general, use of any other antibody based therapies, including antibody-drug conjugates (ADCs) will require assessment of expression of these target proteins on cancer cells, yet any of these strategies are obviously not possible in cf-DNA. In addition, like bulk tissue biopsies, tumor cf-DNA represents the sum total of tumor genomic events thus precluding the full elucidation of relationships between subclonal tumor cell populations and the effect that this heterogeneity may have on clinical disease progression at the single cell resolution.

Looking for precision medicine biomarkers in individual CTCs appears to fill this clinical need by representing a "best of both worlds" scenario. Specifically, CTC profiling allows non-invasive tracking of the tumor molecular biomarker landscape at regular time points (not possible in tissue biopsies) This is coupled with robust sequencing of tumor genome (tumor content over normal cells = 100%, not usually the case in tissue and drastically less so in cf-DNA) and the ability to analyze individual cells and thus reveal potential subclonal tumor cell populations (severely impaired in tissue biopsies or cf-DNA) that may harbor varying breast cancer molecular resistance mechanisms or targetable vulnerabilities. Additionally, as mentioned above, phenotypic analysis is amenable in CTCs, as in tissue IHC, by staining with fluorescently-labeled antibodies, as is transcriptomic analysis by single cell RNA sequencing. This can be of interest in breast cancer when trying to assess PD-L1 levels for immuno-oncology in the triple-negative subtype etc. Our proof of concept approach, described in Chapter II, of applying comprehensive genomic profiling in metastatic breast cancer patient individual (and pooled) CTCs, although currently limited to DNA, is an attempt to begin to fill just such a need in the clinic.

There is a great need for biomarkers with clinical utility in prostate cancer as well, however, at the opposite end of disease spectrum. Thus, instead of metastatic disease, we focused

our prostate cancer work in the early detection setting, as described in Chapter III. As mentioned, given the variable clinical course of prostate cancer and complications/cost related to biopsies, there is currently a great need for biomarkers that have high specificity in addition to sensitivity, and that can be detected non-invasively to allow for low-complication/cost detection of aggressive prostate cancer early and tracking its progression temporally. As mentioned, serum PSA and the few current urine biomarkers in clinical use have limitations especially in their specificity. Furthermore, the combination of a large number of biomarkers offered by the now fully elucidated prostate cancer molecular landscape, modeled via artificial intelligence approaches such as machine learning into algorithms that can improve early detection, is only beginning to be explored pre-clinically[181]. Moreover, there is a great need in the active surveillance prostate cancer setting for biomarkers that can predict cancers likely to be upgraded and thus become eligible for treatment by surgery or radiation. Our approach, in this setting, consisted in combining a large number of such candidate RNA biomarkers into a urine based NGS assay. It showed very encouraging results in the ability to predict presence of aggressive prostate cancer as opposed to low grade (likely indolent) one or complete absence of cancer using tissue pathology results as known truth standard. Furthermore, we detected a discriminating signal of our assay in the active surveillance setting, a challenging patient population which, as explained, is in great need of good biomarkers.

4.2 Metastatic breast cancer CTC comprehensive genomic profiling is feasible and complements and improves that of tissue.

As mentioned in Chapter II earlier, the degree to which the tumor genomic landscape

obtained from CTC sequencing represents that of bulk tissue biopsies/resected tumor material, has not been thoroughly examined. This question is of considerable translational importance as it addresses whether CTC profiling can serve as a proxy for tissue biopsies which are not always available/feasible especially when monitoring patient treatment response or even at the initial metastatic biopsy setting. To answer this question, we performed targeted DNA sequencing in multiple individual CTCs and CTC pools composed of several cells from each of 12 patients which had concurrent whole-exome sequencing of their metastatic tissue biopsies. We found a concordance of 85% (defined as fraction of CTC samples having at least one –but usually multiple– alterations found in the matched tissue sample). There were both kinds of private genomic events, those found only in tissue and those limited to CTCs only. This supports both approaches as complementary. Thus one can imagine a clinical scenario where advanced breast cancer patients undergo broad, comprehensive genomic profiling on their available biopsied/resected tissue. This can then be followed by continuous monitoring of their disease progression and molecular evolution in response to treatment by CTC profiling. The latter can take the shape of tracking a targeted set of that individual patient's tumor genomic alteration repertoire found in tissue (potentially complemented by cf-DNA, as needed) or more comprehensive assessment of CTC genomic landscape as in our study. This latter approach can also be complemented with cf-DNA profiling, especially in patients for whom CTCs can not be collected in adequate numbers.

Other important contributions of our work include findings of actionable genomic alterations in CTCs including in some of the genes mentioned earlier such as *PIK3CA* hotspot mutations now targetable by an FDA-approved drug, alpelisib. Another curious finding was the presence of a minority of circulating epithelial cells that do not appear to harbor any genomic

alterations, thus presumably having non-tumor origin. These may come from normal epithelial tissues anywhere in the body or may be enriched at the blood collection site due to needle damage of normal epithelial tissues of the skin and deeper glandular structures. Additional important findings from this project include the detection of multiple molecular resistance mechanisms to endocrine therapy within one patient's CTC repertoire as exemplified by multiple *ESR1* hotspot mutations. These included a functional mutation not found in tissue and previously unreported in literature. Another finding was the detection of emergent genomic alterations upon disease progression on therapy in patient CTCs over time which highlights the ability of temporal tracking of disease evolution non-invasively.

An interesting finding in one patient was detection of CTCs containing *ESR1* mutant cells that also harbored an *ESR1* one-copy loss (presumably a deletion of the wild type copy resulting in loss of heterozygosity, LOH). This LOH event is expected to generate all-mutant ER dimers hypothesized to be stronger transcriptional activators compared to mutant/wild type or wild type/wild type dimers (**Figure 29**). Such LOH events in mutated oncogenes have been shown to recur in cancers of various kinds[193]. Both of these genomic *ESR1* alterations (mutation and copy-loss) constituting the LOH event were present in bulk tissue sequencing although the one-copy loss was less apparent due to dilution of tumor signal from intratumoral normal cells. Due to these factors, LOH in tissue sequencing could only be presumed. This event was quite clear in individual CTCs on the other hand since the tumor cell fraction is essentially 100% and mono- and bi-allelic mutations will appear at discrete mutant read fractions of approximately 0.5 and 1.0, respectively. Intriguingly, another, more minor CTC subpopulation was present in that patient, which harbored the *ESR1* mutation (at heterogeneous read fraction) but not the wild type copy loss. This may represent a cell in that patient's cancer genomic evolution which may

descend from a time prior to the LOH event happened and subsequently became the dominant clone under the selective pressure imposed by endocrine therapy. This subpopulation was undetectable in bulk tissue sequencing for reasons explained so far and would be much less so by cf-DNA profiling as well.

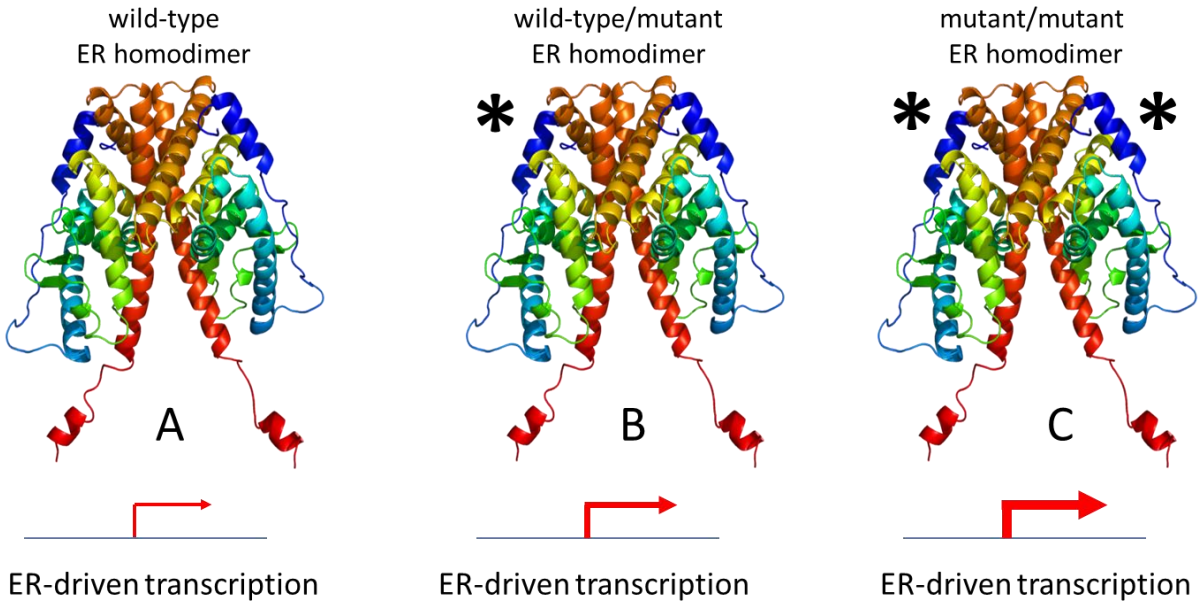


Figure 29. Potential effect of *ESR1* activating mutations in the context of LOH on ER-driven transcription. ER dimers containing all mutant ER (C) may be stronger transcriptional activators than those composed of wild type/mutant (B) or all wild type (A) dimers. Tumors without *ESR1* mutations will have only type A dimers, those with heterozygous mutation will have a distribution of all three types (likely predominated by type B dimers) and those with LOH mutations will only have type C ER dimers presumably showing elevated resistance to endocrine therapy. (ERα homodimer ribbon crystal structure modified from Tanenbaum et al.,[194])

Limitations of our approach center around the fact that the number of CTCs collected is variable among patients, with a subset being CTC-negative despite having abundant metastatic burden on imaging. Furthermore, the number of individual CTCs that can be feasibly sequenced for a given patient, given that they are even available, is finite. This may lead to the inability to detect highly subclonal alterations present only in a small percentage of cells. This can perhaps

be overcome by high depth tissue sequencing or ultra-sensitive ddPCR on cf-DNA. However, a highly subclonal alteration is not likely to be the main driver of disease burden, at least at that point in time. In the event that this subpopulation becomes a driver of therapy resistance and a dominant subclone, its representation in the CTC repertoire will likely increase and its detection will become possible.

Thus, the various aspects of our CTC study listed above underscore the potential of CTC genomic analysis to non-invasively unravel cancer's complex molecular makeup in a personalized fashion for individual patient precision medicine approaches in breast and likely other cancers as well.

4.3 Our prostate cancer urine detection assay has the potential to outperform current clinically approved assays

As mentioned in Chapter III, the results of our prostate cancer urine RNA next generation sequencing assay (NGS-MiPS) support its continued development toward a test with potential clinical utility. Thus in this highly translational project with potential for clinical implementation we demonstrated high analytical validity using an NGS approach which was consistent with clinically validated methods of expression-level measurement such as transcription mediated amplification and RT-qPCR. We then used machine learning to select a subset of our 84 total candidate biomarkers previously validated in prostate cancer tissue, which was trained into a final 15-transcript model that outperformed currently used models/biomarkers in predicting presence of prostate cancer on tissue pathology. Furthermore, it showed potential signal in the active surveillance setting which, as aforementioned, is in need of clinically useful biomarkers.

Lastly, we demonstrated an entirely novel aspect of our assay, specifically in the detection of several prostate cancer somatic hotspot mutations and a germline risk SNP.

A few aspects contribute to this assay's encouraging results and its potential for clinical development. Firstly, our assay is non-invasive (compared to prostate biopsies and even circulatory biomarkers such as CTCs which require a blood draw). The required digital rectal exam (DRE) prior to urine collection for NGS-MiPS is usually performed clinically regardless of whether urine is to be collected or not. Thus our assay does not generally add any invasive procedures on the patient. Furthermore, we showed encouraging results of our assay's performance in pre-DRE urine, thus opening up the potential to disrupt prostate cancer detection with a direct-to-patient approach much like direct-to-consumer genomic ancestry tests. Additionally, the small amount of RNA in urine is not likely to be a limitation since we showed very good NGS-MiPS performance in a high-throughput approach that uses several fold less RNA. Another advantage lies in our use of NGS, which allows for detection of expressed somatic mutations/germline SNPs more readily compared to the current non-NGS approaches. Furthermore, NGS allows for assessment of a large number of transcripts simultaneously which enabled us to include multiple splice isoforms of most observed prostate cancer gene fusions, m- and lncRNAs, transcriptional signatures etc.. Thus, making use of a broad set of candidate biomarkers, coupled with the power of machine learning, is likely to have contributed to our model outperforming clinically used urine models/biomarkers which are currently limited to a few biomarkers at a time.

Limitations of this work include the relatively low tumor content (fraction of tumor-derived as opposed to normal tissue material) in urine. While this is obviously not a drawback in the case of our germline SNP detection, which will be expressed by all of that patient's cells, it

can be a disadvantage for the rest of the biomarkers which are expression-level or somatic mutation based. We can deduce the approximate tumor content in our case of somatic *SPOP* mutation (Chapter III, **Figure 26**) detected at 1.6% variant allele frequency. Assuming a heterozygous mutation as well as considering the fact that normal tissues of the urinary tract express *SPOP* at a few fold higher levels than cancer cells (**Figure 28**), we can estimate a tumor content of 5-10% in that patient. This is perhaps an acceptable trade-off considering the non-invasiveness of urine testing compared to tissue biopsies which themselves will still not have a perfect tumor content even with any tumor content enrichment methods. (We are exploring the possibility of isolating and performing DNA/RNA profiling of single tumor cells purified form urine, a step that increases tumor content to 100%, as explained above for CTCs). Massaging the prostate by performing a DRE prior to urine collection contributes to enriched prostate contents in urine which may mitigate some of the issues with urine tumor content. Furthermore, the majority of our biomarkers are prostate cancer (e.g. Ets fusions) or even aggressive prostate cancer specific (e.g. *SChLAP1*) and thus are likely unaffected by any diluting signal from normal cells. Another limitation is our training/validation patient cohort size and thus our 15-transcript model can be considered preliminary and will likely require retraining in larger cohorts in order to reveal its full potential. Despite any limitations, our overall approach and findings support continued development of our assay as a tool with potential clinical utility in aiding first biopsy decision-making and other prostate cancer clinical settings.

4.4 Future directions of our breast cancer CTC and prostate cancer urine biomarker work

Further work remains to be done, primarily towards clinical implementation of our breast and prostate cancer work. Having shown proof of concept data supporting analytical validity and potential utility of genomic profiling of breast CTCs, we are directing our attention to applications of our method to correlative exploratory studies in patients participating in various clinical trials assessing novel therapies. E.g. our CTC genomic profiling in patients on a Phase I study of a new oral SERD[23, 24] shows signs of diminution of *ESR1* mutant cells in circulation post-treatment (data not shown). Analysis of additional pre- and post-treatment CTCs from these patients will allow accurate determination of the drug's effect on *ESR1* mutant CTCs with statistical certainty. Furthermore, given intra-patient heterogeneity of *ESR1* resistance mutations, a proposed future clinical trial could test whether this is clinically important by determining whether presence of heterogeneity of *ESR1* molecular resistance mechanisms predicts response to endocrine therapy (or lack thereof) compared to non-heterogeneous cases. A trial outcome to measure for this might be duration of endocrine therapy until time of disease progression.

In another exciting recent application of our model, we have detected a *TP53* p.Y205D hotspot mutation in CTCs purified from cerebrospinal fluid (CSF) of a breast cancer patient with brain metastasis, a mutation which was also present in the primary tumor FFPE tissue biopsy sequenced clinically on the FoundationOne platform (data not shown). This pilot study shows proof-of-concept for presence and successful isolation and genomic profiling of brain metastasis cells in CSF. In contrast to blood CTCs, which may come from multiple metastases (if present), CSF CTCs are likely derived from the brain metastasis only, thus potentially allowing tracking tumor evolution of brain vs. other metastases separately. Further work will be needed to fully explore this aspect by more comprehensively comparing tissue biopsy genomics vs. CSF vs.

blood CTCs and determining the degree of their complementarity, discordance and exploring potential clinical application for patients with brain metastases.

Other liquid biopsy strategies with future potential for the purification of CTCs, in addition to blood and CSF, include urine for cancers of the urinary tract (kidney, bladder, prostate, penis), saliva for salivary gland and head and neck cancers, fecal matter for colorectal cancer etc. Another exciting potential application to explore is detection of tumor mutation burden in individual CTCs, a biomarker shown to predict response to checkpoint inhibitor immunotherapy[195]. This application can be of particular importance in triple-negative breast cancer (ER/PR-, Her2-) as well as other cancer types such as non-small cell lung cancer and melanoma where immunotherapy has become mainstay[196]. This can be coupled with real-time monitoring of CTCs while on immunotherapy for genomic alterations known to confer checkpoint blockade resistance such as *JAK2* mutations, loss of β2-macroglobulin (*B2M*) etc.[197].

As scarcity of CTCs in relatively small blood samples, especially in a subset of patients, remains a limitation of CTC clinical implementation, apheresis-based strategies are emerging in order to overcome this problem[198-200]. This centers on the use of external nanofluidic devices connected to a surface vein which allows blood flow through the device for CTC capture by various affinity methods before blood is redirected back to circulation. This increases the amount of blood that can be sampled to 50 – 100 ml or more (depending on duration of collection) compared to static blood draws of several ml each[199]. This has resulted in an increase in the number of cells captured per patient and can thus greatly extend the number of CTC+ patients for which circulating precision therapy approaches described in our method will be possible. Future work in this nascent and exciting area includes confirming that our genomic profiling is

compatible with CTCs purified by this approach, device improvement, portability outside of the clinic to allow for longer collection times in order to screen greater volumes of blood etc.

Potential limitations to clinical implementation of our CTC work include the cost of genomically profiling multiple individual CTCs per patient. This is important in light of the fact that detailed assessment of tumor heterogeneity, despite providing a fascinating window for peering into cancer evolutionary development, has yet to have a real impact in the clinic. In fact, current data supports more robust targeted therapy responses when acting on clonal (truncal) mutations as opposed to subclonal (branch) ones. Some data also suggests that in some cases of excessive tumor heterogeneity, especially after resistance to targeted therapy, a more generalized approach such as cf-DNA may be better than detailed assessment of individual subclones[201], which can be described as a situation of "missing the forest for the trees". In such cases, CTC analysis can be easily converted to a bulk sequencing approach similar to tissue or cf-DNA by pooling multiple individual CTCs from that patient into one sample. Although subclonal alterations present only in one or a minority of cells will be somewhat diluted by other fellow pool cells that don't harbor them, the high tumor content in a CTC pool (100%) is an advantage compared to tissue or cf-DNA profiling which can suffer from low tumor content. We performed just such an analysis on a number of pooled patient CTC samples as a proof of concept study (**Figure 4**) showing similar distribution of alterations in pools as in bulk tissue sequencing. The issue of which approach is optimal (individual or pooled CTCs) could be addressed by a clinical trial whereby *ESR1* resistance mutations in a patient are tracked by both methods in order to compare how well each approach provides a thorough assessment of such heterogeneity.

Given the existence of a few different approaches for tumor molecular profiling (tissue

biopsy, cf-DNA, CTCs), each with its advantages and disadvantages explained above, clinical trials will be required that can tease out the most optimal strategy, or combination of strategies, that provides the best precision oncology outcomes at the lowest cost and invasiveness on the patient's body. Low-pass (depth) cf-DNA sequencing has been proposed as an initial low-cost screening method to first determine tumor content in a patient's circulation as a molecular triage of sorts in order to determine follow up strategies (higher depth cf-DNA, CTCs etc.[114, 202]) Future clinical trials following this approach vs. current precision oncology strategy could determine its clinical utility or lack thereof.

Future work for our urine prostate cancer detection assay should be centered on retraining and re-validating a model in larger patient cohorts composed of at least 200-300 patients each. Although our preliminary 15-transcript model showed an advantage compared to current clinically-used models/biomarkers, the fact that it was trained in a relatively small cohort (*n=73* patients) and validated in two small cohorts (the held out portion of the extreme design cohort, *n=36* patients and the active surveillance cohort, *n=45* patients) suggests that our model may not capture the full potential of our biomarker set. In fact, the difference in ROC AUCs between our training set (0.90) and validation set (0.82) suggests some degree of model overfitting to the specific samples in our training cohort. Furthermore, larger training cohorts can allow our model to take advantage of less recurrent biomarkers/alterations (such as rare gene fusions) that are targeted by our assay but are likely to be misrepresented in small cohorts.

Another aspect of our assay that requires further development is streamlined detection of germline SNP and somatic mutations. Our NGS-MiPS mutation detection method currently rests on a rather manual approach. Due to relatively low expected variant read fractions for somatic mutations, bioinformatics strategies should center around an analysis pipeline for automated

mutant read detection compared to the pre-determined background variant rate noise for those positions. This should be coupled with statistical analysis in order to determine whether the purported mutation variant fraction is truly above background in a statistically significant manner.

As previously discussed, our preliminary data in matched pre- post-DRE urine samples suggested good performance in pre-DRE urine also, especially for targets with high or intermediate expression levels. Potential pre-DRE NGS-MiPS optimization strategies should center around redesigning the assay to decrease the number of targets by removing unnecessary or redundant highly-expressed transcripts. This in order to allow more sequencing reads to go to more useful transcripts that are more scantly expressed. A similar effect can be accomplished by higher sequencing depth with the current assay, although this would result in cost increase. Other potential assay improvements to allow compatibility with pre-DRE urine could include pre-amplification of the low levels of RNA, reverse transcription by gene-specific (as opposed to our current random) priming etc. In addition, looking for cancer biomarkers in various urine compartments such as pellet or exosomes as opposed to whole urine should be explored for pre- (and perhaps even for post-) DRE urine. (ExoDx Prostate Intelliscore, a clinical prostate cancer detection test, uses pre-DRE urine exosomal RNA[155]). Lastly, isolation and testing of our assay on RNA from individual intact urinary cells (after whole transcriptome amplification) is an exciting area that warrants further investigation.

After improvements to our NGS-MiPS assay/model are addressed, its clinical implementation can be explored. This should be done by giving careful consideration to the proposed clinical time point(s) of intended use in the disease course of the prostate cancer patient, the current standard of care with respect to available biomarkers at that time point and

how introduction of our assay intends to change clinical decision-making. This includes whether our assay will be used as an Opt-In factor (supporting performing an intervention where the standard of care is not to), Opt-Out (supporting foregoing an intervention where the standard of care is to perform it) or Opt-Alt (supporting the decision to choose one standard of care intervention over others)[203]. As previously mentioned, one of the clinical junctures of our assay's intended use is to help decision-making at the time of first biopsy for patients presenting with a serum PSA > the clinically used threshold of 4.0 ng/ml. Current standard of care (at least until recently), recommends that a prostate biopsy be performed, a decision that requires better biomarkers as explained earlier. In this scenario, NGS-MiPS would be an Opt-Out test (requiring high sensitivity) which would detect patients likely to have no cancer or very indolent one and would support foregoing the biopsy at that time. By that reasoning, it would also further reinforce the decision to proceed with a biopsy in those patients with PSA over the threshold where urine supports presence of aggressive disease. As multi-parametric magnetic resonance imaging (mpMRI) for suspected prostate cancer is becoming increasingly adopted before biopsy for men presenting with PSA over the threshold, our assay could likewise be an Opt-Out factor in the decision whether to preform MRI or not.

Further down the prostate cancer clinical road, if biopsy results are positive, but only for low grade (likely indolent) cancer, leading to the patient being placed on active surveillance, current standard of care is to repeat biopsy yearly until evidence of upgrading is found. Our assay, could serve here as an Opt-In factor (requiring high specificity) to recommend more frequent repeat biopsies or treatment if urine shows evidence of aggressive disease. Our assay's potential clinical utility in this scenario is supported by NGS-MiPS results in our active surveillance cohort (**Figure 23**). Likewise, for high PSA patients whose biopsy and mpMRI

show no evidence of cancer, the standard of care in such cases being to refrain form any interventions (other than periodical repeats of PSA testing), an NGS-MiPS result showing aggressive cancer signal could serve as an Opt-In factor supporting repeat biopsy (potentially with more random cores than the usual 12) until there is evidence of that aggressive cancer on pathology. This could be an important contribution of our assay since as explained earlier, massaging of the prostate and isolation of the global urine RNA allows NGS-MiPS to assay the entire prostate. This is in contrast to the random core biopsy approach that may miss small/MRI-invisible cancer foci which could be a reason for the discrepant results between the biopsy and NGS-MiPS. Other potential clinical paths one could entertain for NGS-MiPS are cases of low PSA but presence of other factors that add suspicion of cancer. These include age, race, family history, abnormal DRE findings etc. In this case, where the standard of care is to conduct and MRI, our assay could be a factor supporting opting out of that path if urine shows no signs of (aggressive) cancer. In contrast, a direct-to-consumer approach of a potential pre-DRE version of our assay would position it as an assay that suggests opting into making a doctor's appointment if there is some evidence of aggressive prostate cancer in their pre-DRE urine.

Thus, thorough knowledge of the prostate cancer clinical work-up flow chart and current standard of care at each step allows for a value-based approach for implementation of our assay's use in clinical decision-making in order to provide maximum clinical utility.

4.5 Potential for inter-applicability of our precision medicine and biomarker approaches between hormone dependent cancers.

Lessons learned through our breast cancer work have the potential to be applied to

prostate cancer and *vice versa*. Thus, our platform for CTC genomic analysis can be implemented in advanced prostate cancer where, similarly to breast, CTC enumeration has been shown to be a prognostic marker. In addition to blood CTC collection in metastatic disease, urine could also serve as a point of access for collecting prostate cancer cells. In fact, given the prostate's anatomical relationship with the distal urinary tract, prostate cancer tumor cells in urine may be collected for the localized primary tumor thus moving up the detection timeframe to earlier disease stages. Our whole-urine detection platform does not differentiate between cell and cell-free RNA and employs a bulk sequencing approach. Isolation of individual prostate cancer cells in urine, coupled with our already-established CTC analysis platform, can open new opportunities for precision medicine approaches in this malignancy. By extension, isolation of individual (or bulk) cancer cells from breast secretions in women with breast cancer could enable detection, enumeration and phenotypic/genomic profiling on the localized primary tumor or localized recurrence. Likewise, our prostate cancer transcriptomic biomarker NGS approach could be applied to breast cancer by isolating RNA in bulk (cell and cell-free) from breast secretions as mentioned above. Additionally a similar breast cancer transcriptomic NGS assay can be applied to a collection of CellSearch-enriched or individually DEPArray-purified and then combined pure CTCs. Expressed recurrent breast cancer mutations could be targeted and detected in RNA in a similar fashion to our prostate cancer *HOXB13* and *SPOP* mutations, thus enabling precision medicine in addition to cancer detection.

APPENDICES

Appendix I

Precision medicine / liquid biopsy approaches to eye lymphomas

A.1.1 Ocular and orbital adnexal lymphomas

Ocular and orbital adnexal lymphomas (OOALs) are non-Hodgkin's lymphoid malignancies of the eye and surrounding extra-orbital tissues. When originating in the eye, they are the most common primary eye tumor in adults, accounting for up to 15% of eye tumors, and their incidence is on the rise[204, 205]. The vast majority of cases are of B cell origin and are subdivided into several subtypes (marginal zone (MZL), follicular (FL), diffuse large B cell lymphoma (DLBCL) etc.) mirroring systemic lymphomas[206, 207]. Although locally well-controlled with radiation therapy (accompanied by some eye toxicity), OOAL distant metastatic disease, especially for the DLBCL subtype, can have poor prognosis and a 5-year survival rate of 30-70% depending on node involvement[208]. Genomic studies of systemic lymphomas have defined their molecular landscape[209, 210], however, it is unclear how well lymphomas of the eye and surroundings reflect the landscape of their systemic counterparts. Furthermore, precision oncology approaches that target tumor driver alterations and pathways are desperately needed to treat OOALs. To that end, we subjected FFPE lymphoma tissue DNA from a set of 38 OOAL patients to targeted next generation sequencing (NGS) using a panel covering ~130 genes altered in cancers including lymphoid malignancies. These data are presented in appendix A.2 below. We detected similarities as well as differences in the OOAL genomic landscapes compared to those reported for systemic lymphomas and observed potential actionable

alterations. Defining for the first time the landscape of genomic alterations driving eye lymphomas this work can serve as a foundation for precision medicine approaches to these malignancies.

A.1.2 Intraocular (vitreoretinal) lymphomas

Intraocular or vitreoretinal lymphomas (VRL) are defined as the presence of lymphoma cells in the vitreous/retinal space of the eye often presenting as floaters (floating particles in the visual field) that can impair clear vision[211]. They are commonly of B cell origin and in up to 90% of the time are derived from a lymphoma primary to the brain (primary CNS lymphoma, PCNSL)[212]. In addition to radiotherapy, treatment strategies include surgical vitrectomy which consists in aspirating the fluid of the vitreous body containing tumor cells and replacing it with cell-free fluid in order to improve vision. The collected fluid (whether the original undiluted vitreous or the diluted subsequent washes) thus provides a liquid biopsy for the original PCNSL or primary VRL and cytology and flow cytometry is commonly performed clinically to stain for lymphoma markers in any cells present in the collected fluid. This method however can have limited sensitivity and specificity and may be confounded for uveitis, the inflammation of the uveal layer of the eye (choroid, iris and ciliary body). Molecular diagnostic methods that take advantage of the knowledge of the PCNSL genomic landscape are missing and the landscape of primary VRL is not well defined. In Appendix A.2 we present genomic profiling of vitreous fluid-derived DNA obtained from four VRL patients with confirmed PCNSL-derived VRL or suspected VRL without PCNSL or vitrous cytology/flow cytometry confirmation. We used a targeted DNA next generation sequencing (NGS) approach using a panel covering ~130 genes altered in cancers including lymphoid malignancies. We successfully

detected know lymphoma tumor driver alterations in all four cases such as *MYD88* mutations (major and minor hotspots) and highly recurrent *CDKN2A* deletions. In one case of unconfirmed malignancy, our *MYD88* mutation detection in vitreous fluid occurred 2 years ahead of clinical PCNSL diagnosis. This work is a proof-of-concept study for clinical implementation of molecular liquid biopsy approaches for highly specific and sensitive clinical diagnosis in vitreoretinal lymphomas.

Appendix II

Comprehensive genomic profiling of orbital and ocular adnexal lymphomas identifies frequent alterations in *MYD88* and chromatin modifiers: new routes to targeted therapies[3]

A.2.1 Abstract

Orbital and ocular adnexal lymphomas (OOALs) are increasing in incidence, and many forms have high relapse rates, unfavorable prognoses, and complications from localized treatment. Hence, there is an unmet need to identify potentially targetable alterations to improve therapeutic approaches. The objective of this study was to discover potential precision medicine opportunities in patients with OOALs. We performed retrospective targeted next generation sequencing (NGS) profiling of 38 OOAL formalin-fixed paraffin-embedded (FFPE) specimens obtained from a single academic ophthalmology clinic. Potentially actionable mutations and copy number alterations (CNAs) were prioritized based on gain- and loss-of function (GoF, LoF) analyses and catalogued approved and investigational therapies. We performed NGS on DNA from 38 FFPE OOALs (9 follicular lymphomas [FL], 21 marginal zone lymphomas [MZL] and 8 high grade diffuse large B cell lymphomas [DLBCL]) using a panel targeting near-term clinically relevant genes. Of 36 informative samples, 56% harbored a prioritized alteration (median of 1, range 0-5 per sample). *MYD88* was the most frequently altered gene in our cohort,

[3] This study has previously appeared in Cani, A.K., et al. Modern Pathology, 2016. (168. Cani, A.K., et al., *Comprehensive genomic profiling of orbital and ocular adnexal lymphomas identifies frequent alterations in MYD88 and chromatin modifiers: new routes to targeted therapies.* Mod Pathol, 2016. **29**(7): p. 685-97.)

with hot-spot GoF mutations identified in 28% of OOALs (17% and 71% of low [FL and MZL] and high grade OOALs, respectively). Prioritized alterations in epigenetic modulators were common and included GoF *EZH2* mutations and LoF *ARID1A* mutations (both in 8% of OOALs). Single prioritized alterations were also identified in the histone methyltransferases *KMT2B* and *KMT3B*. Lastly, LoF mutations and CNAs in the tumor suppressors *TP53*, *CDKN2A*, *PTEN*, *ATM* and *NF1*, as well as GoF mutations in the oncogenes *HRAS* and *NRAS* were also observed. In conclusion, potentially actionable alterations are frequent in OOALs, including 25% of sample with recurrent GoF *MYD88* or *EZH2* mutations that are inclusion/stratification criteria for ongoing clinical trials in lymphomas or other tumors. Prioritized alterations in histone/chromatin modifying genes were also common in FL and DLBCLs, including LoF *ARID1A* alterations reported to enhance sensitivity to *EZH2* inhibition. Taken together, we demonstrate that NGS can be used to profile routine OOALs for identification of somatic driving alterations and nomination of potential therapeutic strategies.

A.2.2 Introduction

Non-Hodgkin lymphomas of the orbital and ocular adnexa (OOALs) represent 10-15% of all tumors present in the orbit, eyelids, conjunctiva, and lacrimal apparatus (i.e. the "adnexa") and are the most common primary orbital cancer in adults[204, 205]. OOALs are B-cell proliferations that include low-grade tumors such as extra-nodal marginal zone (MZL, also known as mucosa-associated lymphoid tissue lymphoma [MALT]) and follicular lymphoma (FL), as well as high grade tumors such as mantle cell lymphoma (MCL) and diffuse large B-cell lymphoma (DLBCL). MZL, FL, MCL and DLBCL account for ~50%, 10-20%, 5-10% and 5-10% of OOALs, respectively[206, 207]. While the mainstay of treatment for primary OOALs, external beam radiation therapy (EBRT), results in high rates of local control, distant relapse rates are high (up to 40%) and local toxicities can lead to vision loss or blindness[213]. Likewise, despite combinations of EBRT chemotherapy regimens, antibiotic therapy, immunotherapy, and radioimmunotherapy, 5-year survival rates for OOALs range from 70-80% without lymph node involvement, to 30-40% with lymph node involvement[208]. Alarmingly, OOALs have also shown a rapidly increasing incidence (~6% yearly increase in white Americans from an analysis of Surveillance, Epidemioloyg, and End Results [SEER] data)[214]. Taken together, these data demonstrate an unmet need to develop novel treatment strategies for OOALs.

The development of novel therapies for OOALs has been hampered by the limited nature of studies exploring genomic drivers of the disease. Although cytogenetic, copy number and transcript, and immunohistochemistry (IHC) based profiling studies have been reported[215-219], these represent limited cohorts mainly assessing MZL, the most common OOAL subtype that is associated with good prognosis. Importantly, although next generation sequencing (NGS)

studies have defined genetic alterations and potential therapeutic targets in non-orbital lymphomas, such approaches have not been applied to NHLs. Hence, in order to define potentially actionable somatic mutations and copy number alterations (CNAs) in OOALs, we performed NGS on 38 formalin-fixed paraffin embedded (FFPE) OOALs, including MZL, FL, and DLBCL histologic subtypes.

A.2.3 Materials and Methods

A2.3.1 Case Selection

With University of Michigan IRB approval, we identified a cohort of 38 archived, routine clinical FFPE orbital lymphoma specimens from the University of Michigan Department of Pathology Tissue Archive for next generation sequencing. Clinicopathological information for each case was obtained from the clinical archive. Hematoxylin and eosin (H&E) stained slides and immunostains (when available) were reviewed for all cases by board-certified Anatomic Pathologists (A.S.M. and S.A.T.) to ensure sufficient tumor content.

A2.3.2 Targeted Next Generation Sequencing (NGS)

Targeted NGS was performed essentially as described[101-103]. For each specimen, 3-7 x 10um FFPE sections were cut from a single representative block per case, using macrodissection with a scalpel as needed to enrich for tumor content. DNA was isolated using the Qiagen Allprep FFPE DNA/RNA kit (Qiagen, Valencia, CA) and quantified as described[101-103]. Targeted, multiplexed PCR based next generation sequencing (NGS) was performed by Ion Torrent NGS using the DNA component of a beta version of the Oncomine Comprehensive Assay (OCP), a custom panel comprised of 3435 amplicons targeting 126 genes. Genes included in this panel were selected based on pan-solid tumor NGS and copy number

profiling data analysis that prioritized somatic, recurrently altered oncogenes, tumors suppressors, genes present in high level copy gains/losses and known/investigational therapeutic targets[102]. Library preparation with barcode incorporation, template preparation and sequencing using the Ion Torrent Proton sequencer were performed according to the manufacturer's instructions. Data analysis was performed using Torrent Suite 4.0.2, with alignment by TMAP using default parameters, and variant calling using the Torrent Variant Caller plugin (version 4.0-r76860) using default low-stringency somatic variant settings. Variant annotation filtering and prioritization was performed essentially as described using validated in house pipelines[101-103, 105]. Briefly, called variants were filtered to remove synonymous or non-coding variants, those with flow corrected read depths (FDP) less than 20, flow corrected variant allele containing reads (FAO) less than 6, variant allele frequencies (FAO/FDP) less than 0.10, extreme skewing of forward/reverse flow corrected reads calling the variant (FSAF/FSAR <0.2 or >5), or indels within homopolymer runs >4. Called variants were filtered using a panel-specific, in house blacklist. Variants with allele frequencies >0.5% in ESP6500 or 1000 genomes or those reported in ESP6500 or 1000 genomes with observed variant allele frequencies between 0.40 and 0.60 or > 0.9 were considered germ line variants unless occurring at a known hot-spot variant. Variants located at the last mapped base (or outside) of amplicon target regions, variants with the majority of supporting reads harboring additional mismatches or indels (likely sequencing error), those in repeat–rich regions (likely mapping artifacts), and variants occurring exclusively in one amplicon if overlapping amplicons cover the variant were excluded. High confidence somatic variants passing the above criteria were then visually confirmed in IGV. We have previously confirmed that these filtering criteria identify variants that pass PCR validation with >95% accuracy[220]. Copy number analysis from total amplicon read counts provided by

the Coverage Analysis Plug-in was performed essentially as described using a validated approach[101-103, 106]. Genes with a $\log_2$ copy number estimate of <-1 or >0.6 were considered to have high level loss or gain, respectively.

To prioritize potential driving alterations, we utilized Oncomine software tools (powertools.oncomine.com) to annotate called variants, which use pan-cancer NGS data to identify genes as oncogenes or tumor suppressors, based on overrepresentation of hot-spot or deleterious mutations, respectively. Variants in oncogenes are then considered gain-of-function (GoF) if at a hot-spot and variants in tumor suppressors are considered loss of function (LoF) if deleterious or at a hot-spot[101, 102, 105]. Likewise, high level copy number alterations were prioritized if they were concordant with the minimal common region (MCR) analysis used to design the OCP (e.g. high level copy number gain in a gene prioritized as amplified/deleted by MCR analysis.

A2.3.3 Immunohistochemistry (IHC)

IHC was performed on the DAKO Autostainer (DAKO, Carpinteria, CA) using DAKO Envision+ and diaminobenzadine (DAB) as the chromogen. Sections of de-paraffinized OOALs were labeled with ARID1A (mouse monoclonal, clone PSG3, 1:250, Santa Cruz Biotechnology, Dallas TX, #sc-32761), EZH2 (mouse monoclonal, clone 11, 1:100, BD Biosciences, San Jose, CA, #612666), or KMT3B/NSD1 (rabbit polyclonal, 1:250, EMD Millipore, Billerica, MA, #ABE1009) for 60 minutes at ambient temperature. Microwave epitope retrieval in 10 mM Tris/HCl, pH 9 containing 1 mM EDTA was used prior to staining. Appropriate negatives (no primary antibody) were stained in parallel with each set of tumors studied. IHC slides were examined by light microscopy by two pathologists (A.S.M. and S.A.T.). Only staining of the nucleus was marked as positive expression. The staining was scored semiquantitatively and

recorded based on percent nuclei staining (0 = negative, 1 = 1 – 25% immunoreactive cells, 2 = 26% – 50% immunoreactive cells, 3 = 51% - 75% immunoreactive cells, 4 = 76% - 100% immunoreactive cells) and intensity of staining (0 = negative, 1 = weak, 2 = moderate, 3 = strong). Corresponding sections were also H&E stained as previously described[221-223].

A2.3.4 Statistics

Comparisons of the number of mutations and copy number alterations per sample by lymphoma subtype were performed using the Kruskal-Wallis test with post-hoc pairwise comparison of subgroups using R 3.1.0.

A.2.4 Results

We performed targeted NGS on a cohort of 38 FFPE OOALs comprised of 8 DLBCLs, 9 FLs and 21 MZLs; representative photomicrographs are shown in **Figure 30A & B** and clinical characteristics of all informative patients (see below) are presented in **Figure 30C.** We isolated an average of 5.5 ug DNA per case (range 0.1-24.1 ug) and all samples had >70% estimated tumor content by H&E evaluation after macrodissection (as needed). NGS was performed using the DNA component of a beta version of the Oncomine Comprehensive Assay (OCP v1), a custom panel comprised of 3,435 amplicons targeting 126 genes and Ion Torrent based sequencing on the Proton machine. Targeted genes were selected based on pan-solid tumor NGS and copy number profiling data analysis to prioritize somatic, recurrently altered oncogenes, tumors suppressors and genes present in high level CNAs, filtered by available or investigational therapeutic targets.[102]

Of the 38 samples assayed, one MZL (IE-26) and one DLBCL (IE-31) sample yielded low quality sequencing data due to poor genomic DNA quality and were excluded from all

subsequent analyses. Across the thirty-six informative samples, NGS generated an average of 1,871,816 mapped reads yielding 545x targeted base coverage (**Table 12**). We identified a total of 41 high-confidence, prioritized somatic alterations (median *n*=1, range *n*=0-5 per sample) comprised of non-synonymous point mutations (*n*=33), short insertions/deletions (indels; *n*=2) and copy number variations (CNVs; *n*=6). All prioritized somatic mutations and high level CNAs for each case are shown in an integrative heat map (**Figure 30C**) and given in **Table 13**. Prioritized alterations were present in 19 of 36 samples (53%). DLBCLs harbored the majority of total prioritized alterations (6 of 7 [86%] with prioritized alterations; median prioritized alterations *n*=3, range *n*=0-5). Prioritized alterations were identified in 6 of 9 (67%) FL (median prioritized alterations *n*=1, range *n*=0-3) and 7 of 20 (35%) MZL (median prioritized alterations *n*=0, range *n*=0-2). The number of prioritized alterations was significantly different between the lymphoma subtypes (Kruskal-Wallis test, *p*=0.0014), as shown in **Figure 30D.**

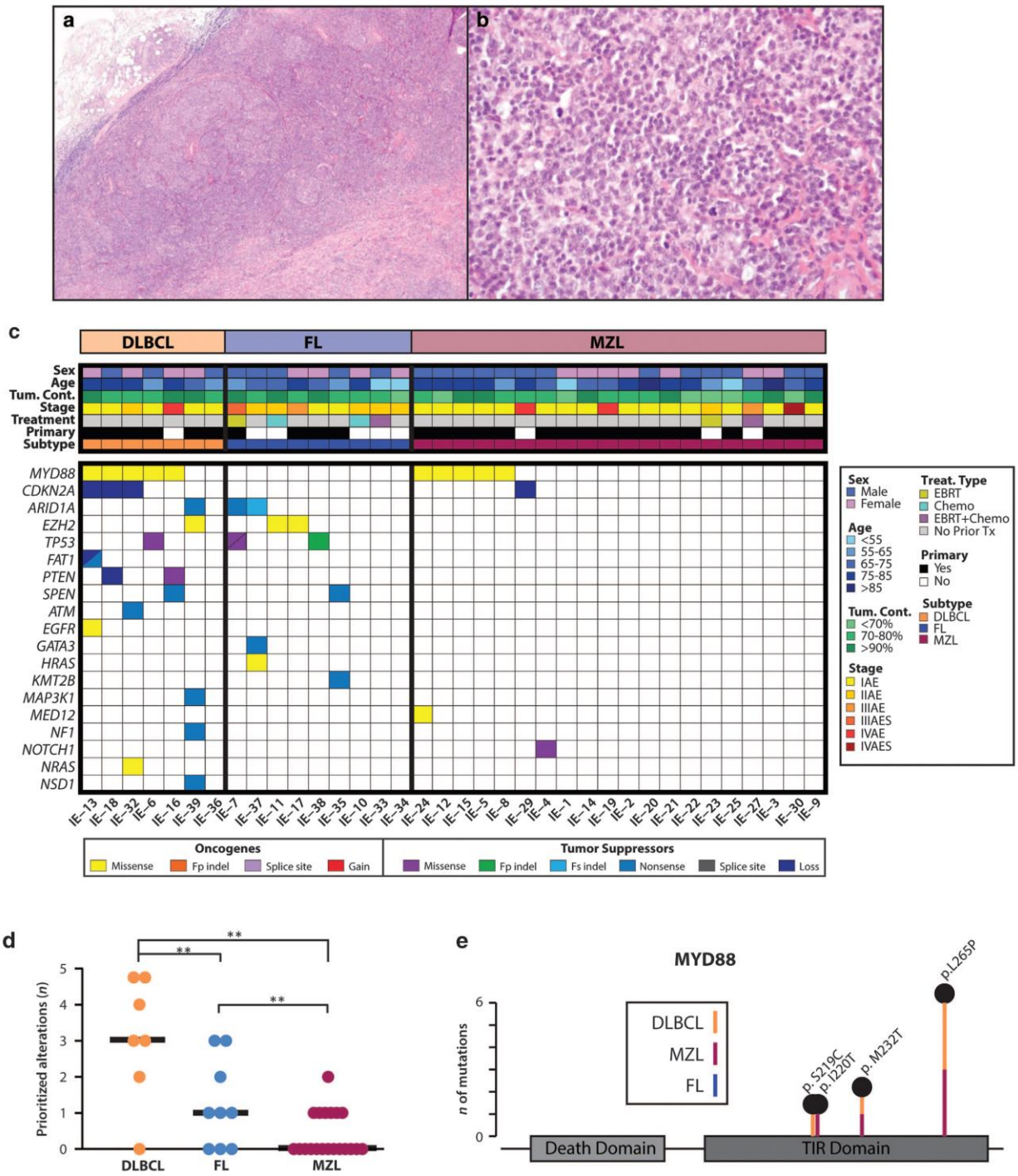


Figure 30. Targeted next-generation sequencing (NGS) of routine formalin-fixed, paraffin-embedded orbital and ocular adnexal lymphomas identifies recurrent informative/potentially actionable alterations. (**A** and **B**) Representative histology by hematoxylin and eosin (H&E) staining of a formalin-fixed, paraffin-embedded lymphoma sample (IE-38, a follicular lymphoma (FL)). Original magnification, × 4 (**A**) and × 40 (**B**). **C** Integrative heatmap of prioritized, driver/potentially actionable mutations and copy number alterations across the 36 informative orbital

and ocular adnexal lymphomas. Clinicopathologic information including patient sex, age, estimated tumor content, and subtypes (diffuse large B-cell lymphoma (DLBCL), FL, and marginal zone lymphoma (MZL)) are indicated in the header according to the legend on the right. All prioritized, high-confidence, gain-of-function or loss-of-function somatic mutations in oncogenes and tumor suppressors, respectively, as well as high-level copy number alterations are shown for each case. Specific alteration types are indicated according to the legend (Fp and Fs indel, frame-preserving and frame-shift indels, respectively; Gain, copy number gain; Loss, copy number loss). Slashed boxes indicate two alterations. **D** Comparison of the number (n) of prioritized alterations per sample (mutations and copy number alterations) by subtype. The median number of alterations per type is shown by thick black bars. **P<0.01 from posthoc subgroup comparison of Kruskal–Wallis test. **E** Stick plot of recurrent domain mutations. Schematic representation of the location and subtype of *MYD88* gain-of-function mutations identified in 10 of 36 orbital and ocular adnexal lymphomas. The number of mutations (n) at each residue is indicated by the stick length according to the scale (left).

Sample	Mapped Reads	On Target	Mean Depth	Uniformity
IE-01	475,068	97.8%	139	90.5%
IE-02	1,198,553	98.3%	355	85.6%
IE-03	1,307,356	98.2%	389	89.5%
IE-04	751,659	98.5%	224	84.9%
IE-05	1,609,024	98.2%	479	88.8%
IE-06	1,071,293	97.6%	313	83.9%
IE-07	1,213,220	98.1%	358	88.4%
IE-08	2,469,002	98.3%	728	83.4%
IE-09	1,791,846	97.2%	520	90.4%
IE-10	831,116	97.8%	242	88.7%
IE-11	2,579,189	96.8%	741	88.1%
IE-12	1,793,249	97.9%	523	89.5%
IE-13	2,625,398	97.2%	748	87.3%
IE-14	2,227,155	97.8%	646	87.5%
IE-15	2,946,426	97.9%	860	87.6%
IE-16	1,398,441	98.2%	415	85.7%
IE-17	1,955,364	98.3%	581	85.9%
IE-18	2,242,925	98.3%	666	87.8%
IE-19	832,754	98.5%	250	90.5%
IE-20	1,333,225	98.6%	403	91.5%
IE-21	834,096	98.5%	250	88.4%
IE-22	3,296,664	98.4%	988	89.4%
IE-23	3,456,374	95.1%	962	62.9%
IE-24	3,586,389	97.8%	1,035	86.1%
IE-25	1,585,151	95.5%	444	89.8%
IE-27	2,904,798	97.2%	838	88.3%
IE-29	1,149,546	97.9%	335	86.9%
IE-30	2,820,224	98.3%	828	84.9%
IE-32	3,732,365	97.7%	1,086	87.3%
IE-33	1,292,583	97.8%	378	88.5%
IE-34	1,421,675	97.2%	413	91.6%
IE-35	1,225,099	96.2%	343	83.3%
IE-36	1,631,165	97.3%	470	89.3%
IE-37	2,432,098	97.1%	693	83.1%
IE-38	2,914,581	98.1%	860	89.4%
IE-39	450,316	92.6%	114	82.5%
Average	**1,871,816**	**97.6%**	**545**	**86.9%**

Table 12. Sequencing statistics for informative orbital and ocular adnexal lymphomas. Orbital and ocular adnexal lymphomas were subjected to next generation sequencing of multiplexed templates using Ion PI chips on the Proton sequencer Multiplexed PCR. For all informative samples, the number of mapped reads, % reads on target, average read coverage depth over targeted bases and uniformity of mapped reads are shown. Average values for the cohort are given.

Sample	Subtype	Alteration Type	Location	Gene	Ref.	Var.	AA Change	Var. freq.	CNA	Log_2 CNRatio	Onc/TSG	Epigenetic
IE-37	FL	Mutation	chr1:27087894	*ARID1A*	GGCCACCCA	G	p.727_730del	0.44	N/A	N/A	TSG	Yes
IE-39	DLBCL	Mutation	chr1:27100375	*ARID1A*	C	T	p.Q1363X	0.11	N/A	N/A	TSG	Yes
IE-07	FL	Mutation	chr1:27057736	*ARID1A*	C	T	p.Q482X	0.26	N/A	N/A	TSG	Yes
IE-32	DLBCL	Mutation	chr11:108183151	*ATM*	G	T	p.E1978X	0.49	N/A	N/A	TSG	No
IE-13	DLBCL	Mutation	chr7:55249028	*EGFR*	C	T	p.R776C	0.11	N/A	N/A	Oncogene	No
IE-17	FL	Mutation	chr7:148508728	*EZH2*	A	T	p.Y646N	0.23	N/A	N/A	Oncogene	Yes
IE-11	FL	Mutation	chr7:148508728	*EZH2*	A	T	p.Y646N	0.16	N/A	N/A	Oncogene	Yes
IE-39	DLBCL	Mutation	chr7:148508727	*EZH2*	T	G	p.Y646S	0.37	N/A	N/A	Oncogene	Yes
IE-13	DLBCL	Mutation	chr4:187557927	*FAT1*	G	A	p.R1262X	0.11	N/A	N/A	TSG	No
IE-37	FL	Mutation	chr10:8100362	*GATA3*	G	A	p.W112X	0.11	N/A	N/A	TSG	No
IE-37	FL	Mutation	chr11:533878	*HRAS*	C	T	p.G60S	0.11	N/A	N/A	Oncogene	No
IE-35	FL	Mutation	chr19:36228097	*KMT2B*	C	T	p.Q2495X	0.29	N/A	N/A	TSG	Yes
IE-39	DLBCL	Mutation	chr5:56171042	*MAP3K1*	C	T	p.Q624X	0.14	N/A	N/A	TSG	No
IE-24	MZL	Mutation	chrX:70339260	*MED12*	A	C	p.N46T	0.99	N/A	N/A	Oncogene	No
IE-24	MZL	Mutation	chr3:38182035	*MYD88*	T	C	p.I220T	0.38	N/A	N/A	Oncogene	No
IE-13	DLBCL	Mutation	chr3:38182641	*MYD88*	T	C	p.L265P	0.90	N/A	N/A	Oncogene	No
IE-18	DLBCL	Mutation	chr3:38182641	*MYD88*	T	C	p.L265P	0.48	N/A	N/A	Oncogene	No
IE-32	DLBCL	Mutation	chr3:38182641	*MYD88*	T	C	p.L265P	0.39	N/A	N/A	Oncogene	No
IE-12	MZL	Mutation	chr3:38182641	*MYD88*	T	C	p.L265P	0.52	N/A	N/A	Oncogene	No
IE-15	MZL	Mutation	chr3:38182641	*MYD88*	T	C	p.L265P	0.36	N/A	N/A	Oncogene	No
IE-05	MZL	Mutation	chr3:38182641	*MYD88*	T	C	p.L265P	0.47	N/A	N/A	Oncogene	No
IE-16	DLBCL	Mutation	chr3:38182259	*MYD88*	T	C	p.M232T	0.27	N/A	N/A	Oncogene	No
IE-08	MZL	Mutation	chr3:38182259	*MYD88*	T	C	p.M232T	0.44	N/A	N/A	Oncogene	No
IE-06	DLBCL	Mutation	chr3:38182032	*MYD88*	C	G	p.S219C	0.31	N/A	N/A	Oncogene	No
IE-39	DLBCL	Mutation	chr17:29664869	*NF1*	G	A	p.W2225X	0.10	N/A	N/A	TSG	No
IE-04	MZL	Mutation	chr9:139397775	*NOTCH1*	C	T	p.V1676I	0.61	N/A	N/A	TSG	No
IE-32	DLBCL	Mutation	chr1:115256530	*NRAS*	G	T	p.Q61K	0.20	N/A	N/A	Oncogene	No
IE-39	DLBCL	Mutation	chr5:176675278	*NSD1*	C	T	p.Q1532X	0.11	N/A	N/A	TSG	Yes
IE-16	DLBCL	Mutation	chr10:89624303	*PTEN*	C	T	p.T26I	0.40	N/A	N/A	TSG	No
IE-35	FL	Mutation	chr1:16259360	*SPEN*	C	T	p.Q2209X	0.16	N/A	N/A	TSG	No
IE-16	DLBCL	Mutation	chr1:16260237	*SPEN*	G	A	p.W2501X	0.36	N/A	N/A	TSG	No
IE-38	FL	Mutation	chr17:7577515	*TP53*	GTGA	G	p.255_256del	0.24	N/A	N/A	TSG	No
IE-07	FL	Mutation	chr17:7577550	*TP53*	C	T	p.G244D	0.21	N/A	N/A	TSG	No
IE-06	DLBCL	Mutation	chr17:7578406	*TP53*	C	T	p.R175H	0.36	N/A	N/A	TSG	No
IE-07	FL	Mutation	chr17:7577538	*TP53*	C	T	p.R248Q	0.23	N/A	N/A	TSG	No
IE-13	DLBCL	CNA	chr9	*CDKN2A*	N/A	N/A	N/A	N/A	two copy loss	-3.73	TSG	No
IE-18	DLBCL	CNA	chr9	*CDKN2A*	N/A	N/A	N/A	N/A	two copy loss	-4.26	TSG	No
IE-32	DLBCL	CNA	chr9	*CDKN2A*	N/A	N/A	N/A	N/A	two copy loss	-2.70	TSG	No
IE-29	MZL	CNA	chr9	*CDKN2A*	N/A	N/A	N/A	N/A	two copy loss	-2.37	TSG	No
IE-13	DLBCL	CNA	chr4	*FAT1*	N/A	N/A	N/A	N/A	two copy loss	-1.75	TSG	No
IE-18	DLBCL	CNA	chr10	*PTEN*	N/A	N/A	N/A	N/A	two copy loss	-1.34	TSG	No

Table 13. Prioritized mutations and high level copy number alterations (CNAs) across informative sequenced orbital and ocular adnexal lymphomas (OOALs). All high confidence, prioritized, somatic non-synonymous mutations and and high-level CNAs (see Methods) identified across the 36 informative OOALs are shown. For each variant the sample, subtype (follicular lymphoma [FL], diffuse large B-cell lymphoma [DLBCL] and marginal zone lymphoma [MZL], alteration type, chromosomal location and gene are indicated. For mutations, the (Ref.) and variant (Var.) alleles, amino acid (AA) change and flow corrected variant allele frequency (Var. freq.) are shown. For CNAs, gain or loss and log2 copy number ratios are shown. Gene type (Oncogene [Onc] vs. Tumor suppressor [TSG]) and known status as an epigenetic regualtor are shown.

Across the 36 OOALs, 10 (28%) harbored prioritized GoF mutations in the Toll/interleukin-1 receptor (TIR) domain of *myeloid differentiation factor 88* (*MYD88*), making it the most frequently altered gene (by prioritized alterations) in our cohort. MYD88 is an adaptor protein that binds to the intracellular domains of Toll-like receptors (TLRs) as well as interleukin 1 receptor on B cells and macrophages, which stimulates the nuclear factor kappa-light-chain-enhancer of activated B cells (NF-κB) signaling pathway, and is involved in innate immunity.[224] In our cohort, *MYD88* hot-spot GoF mutations were present in 5/7 (71%)

DLBCLs, including three p.L265P, one p.M232T and one p.S219C mutations. The former represents the most frequently altered *MYD88* hot-spot across human cancers, while the latter two mutations occurred at minor hot-spots. The remaining *MYD88* (TIR domain) mutations occurred in 5 of 20 (25%) MZL samples. Of note, *MYD88* mutations comprise the majority of total prioritized alterations (5 of 8, 63%) in our MZL samples, with three p.L265P mutations and two other TIR domain hot-spot mutations (one each of p.M232T and p.I220T). All GoF *MYD88* mutations across the cohort are shown in **Figure 30E**. Taken together, our data demonstrate that *MYD88* TIR domain mutations are common in the DLBCL and MZL subtypes of OOALs.

Recurrent somatic alterations in histone/chromatin remodeling proteins are frequent in extra-OOALs. Specifically, 22% of DLBCLs and 7-22% of FLs outside of the orbit and ocular adnexa carry GoF missense mutations (at the p.Y646 hot-spot) in the methyltransferase SET domain of the histone H3 lysine 27 trimethylase *enhancer of zeste 2* (*EZH2*)[225-227]. Herein, we identified 3 of 36 OOALs (8%; 2/9 FLs and 1/7 DLBCLs) with *EZH2* p.Y646 mutations (**Figure 30C**). We also identified LoF mutations in tumor suppressor *AT-rich interactive domain-containing protein 1A* (*ARID1A),* an epigenetic regulator that is an ATPase dependent SWI/SNF nucleosome remodeling complex subunit. We identified a prioritized *ARID1A* nonsense mutation (p.Q1363X) in 1 of 7 (14%) DLBCLs; two of 9 (22%) FLs harbored LoF *ARID1A* mutations (p.Q482X and p.727_730del). Interestingly, the *ARID1A* mutated DLBCL sample (IE-39; p.Q1363X) also contained one of the aforementioned activating *EZH2* Y646 mutations (p.Y646S) (**Figure 30C**). Lastly, we identified prioritized mutations in other chromatin modifying genes, including LoF mutations in *KMT2B* (also known as *MLL2* or *MLL4;* p.Q2495X) and *KMT3B* (also known as *NSD1;* p.Q1532X) in one case, each, of FL and DLBCL, respectively (**Figure 30C**).

All mutations in the above chromatin modeling genes in our current were heterozygous based on estimated tumor content and variant allele frequencies. Hence, to determine the impact of mutations on protein expression, we assessed ARID1A, EZH2 and KMT3B expression in several OOAL samples by IHC. Although limited by the number of mutant samples, IHC levels of ARID1A in wild type and mutant FL samples showed no qualitative differences, as ARID1A protein was very low or absent in both samples (**Figure 31A, B**). These results suggest that mechanisms besides mutation/CNA may lead to decreased *ARID1A* expression in OOALs. In contrast, the DLBCL harboring the LOF *KMT3B (NSD1)* p.Q1532X non-sense mutation showed decreased KMT3B expression compared to a DLBCL sample without prioritized *KMT3B* alteration (**Figure 31C, D**). As predicted for GoF mutations, *EZH2* expression was retained in two samples that each harbored prioritized p.Y646N or p.Y646S mutations (**Figure 31E, F**).

In addition to MYD88 and chromatin modifying genes, we also identified samples with prioritized alterations in well-known tumor suppressors and oncogenes (**Fig 30C** and **Table 13)**. Three DLBCLs harbored single LoF mutations in *phosphatase and tensin homolog* (*PTEN*), *ataxia telangiectasia mutation* (*ATM*) or *neurofibromin 1* (*NF1*). Of interest, sample IE-07, a FL, harbored biallelic *tumor protein 53* (*TP53*) LoF hot-spot mutations (R248 and G244). Another FL contained a *TP53* nonframeshift deletion (p.255-256del) while a DLBCL harbored a *TP53* R175 hot-spot missense mutation. LoF nonsense mutations were also observed in the tumor suppressor *split end family protein* (*SPEN*) in one case each of DLBCL and FL. Prioritized alterations in oncogenes included missense mutations in *neuroblastoma RAS Viral oncogene homolog* (*NRAS,* Q61K) and *Harvey rat sarcoma viral oncogene homolog (HRAS,* G60S) in a single DLBCL and FL sample, respectively.

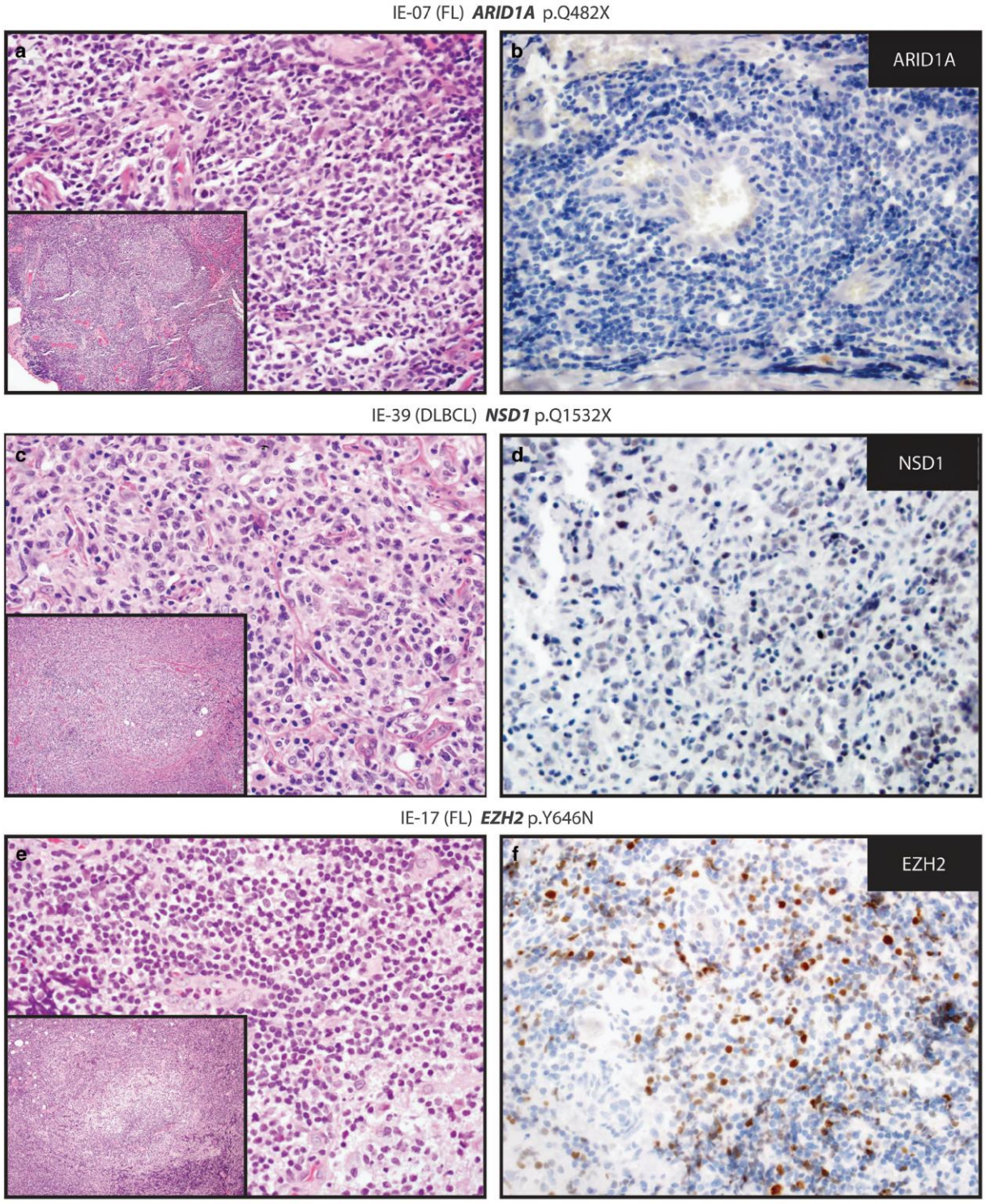


Figure 31. Immunohistochemistry for chromatin-modifying proteins in mutant *ARID1A*, *NSD1*, and *EZH2* lymphoma samples. Hematoxylin and eosin (H&E, **A**, **C** and **E**) stains (large panels, original magnifications: × 400, insets × 100) and corresponding immunohistochemistry (**B**, **D** and **F**, original magnifications: × 400) for indicated chromatin-modifying proteins was performed in

selected
orbital and ocular adnexal lymphomas based on mutational status. (**A** and **B**) IE-07, a usual appearing follicular lymphoma (a) with a loss of-function *ARID1A* mutation (p.Q842X), demonstrates absent *ARID1A* expression (**A**). (**C** and **D**). IE-39, a diffuse large B-cell and marginal zone lymphoma with a truncating stopgain mutation in *NSD1* (p.Q1532X), shows very weak expression of NSD1 (**D**) in tumor cells (wildtype *NSD1* diffuse large B-cell and marginal zone lymphoma samples showed robust expression of NSD1 by immunohistochemistry, see **Figure 32**). IE-17, a follicular lymphoma (**E**) with a predicted gain-of-function variant in *EZH2* (p.Y646N), shows robust *EZH2* (**F**) expression (wild-type *EZH2* marginal zone lymphoma samples did not show *EZH2* expression by immunohistochemistry, see **Figure 32**).

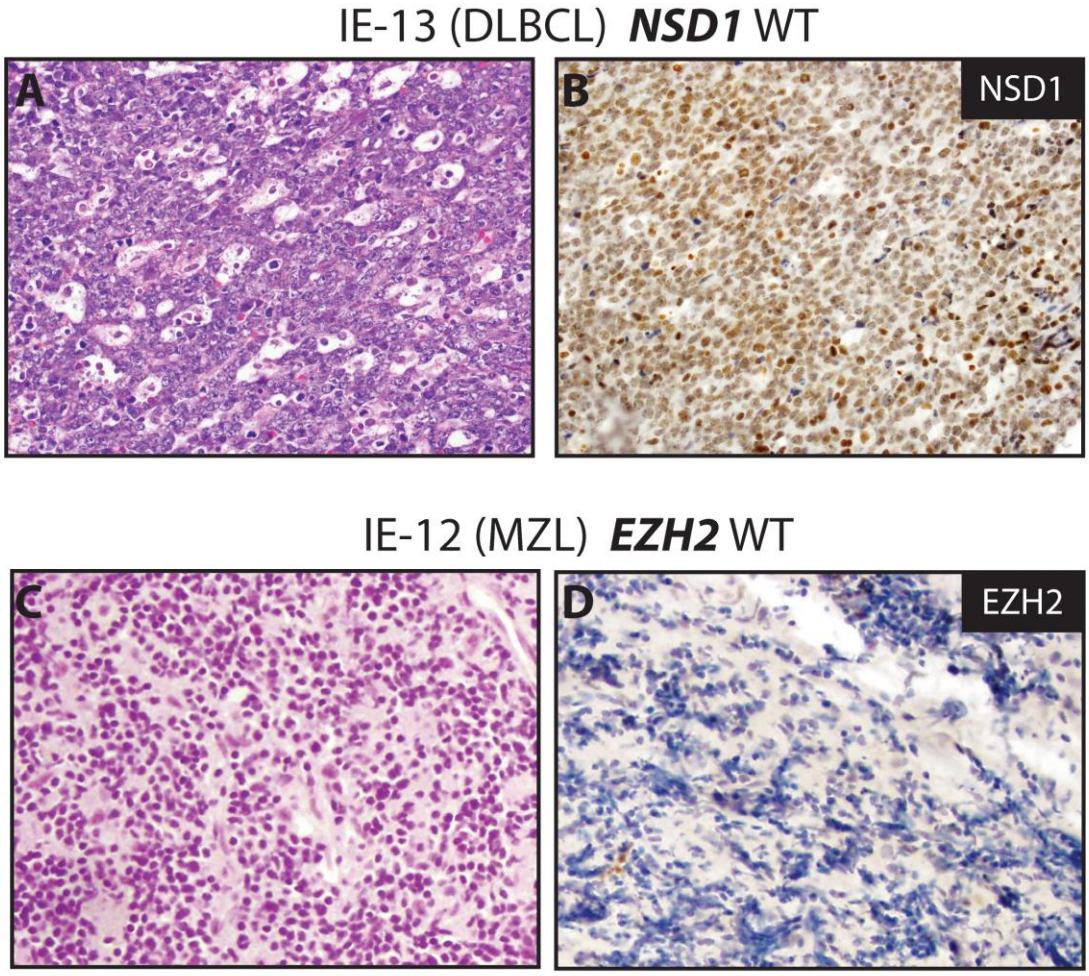
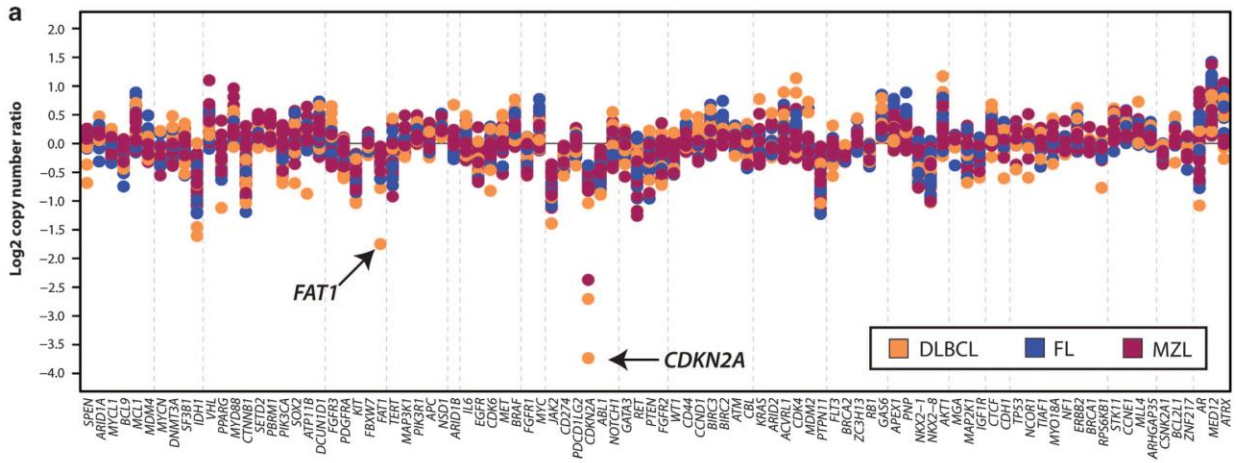


Figure 32. lmmunohistochemistry for chromatin modifying proteins in orbital and ocular adnexal lymphomas with wild-type *NSD1* and *EZH2*. Hematoxylin and eosin (H&E, **A**, **C**) stains (400x magnification) and corresponding immunohistochemistry (**B**, **D**, 400x magnification) for indicated chromatin modifying proteins in representative lymphomas without mutations in the indicated genes. **A** & **B**. IE-13, a diffuse large B-cell lymphoma (DLBCL, **A**) with wild-type *NSD1* shows robust expression of *NSD1* in nearly every cell (**B**, compare with Figure 31D). **C** & **D**. IE-12,

a marginal zone lymphoma (**C**) with wild type *EZH2*, shows absent expression of *EZH2* (**D**, compare with Figure 31F).

In addition to single nucleotide alterations and short indels, we used NGS to assess copy number alterations (CNAs) in genes targeted by our panel using a validated approach[106]. Overall, we identified relatively few focal, high level amplifications or deletions in our OOAL cohort (**Fig 30C**). We identified high level, prioritized deletions in *CDKN2A* (*p16INK4A*) in three DLBCLs and one MZL, as well as a high level deletion in *PTEN* in a DLBCL. Finally, although not prioritized, broad low level gains/losses were observed, including one-copy gains in chromosomes 12 and 3 in a DLBCL and an MZL, respectively, as well as one-copy losses in 17p in two DLBCLs. Cohort wide copy number plots are shown in **Figure 33A**, with individual amplicon level plots for DLBCLs with high level deletions in *CDKN2A* and *PTEN* (IE-18) and CDKN2A (IE-32), respectively, shown in **Figure 33B**.



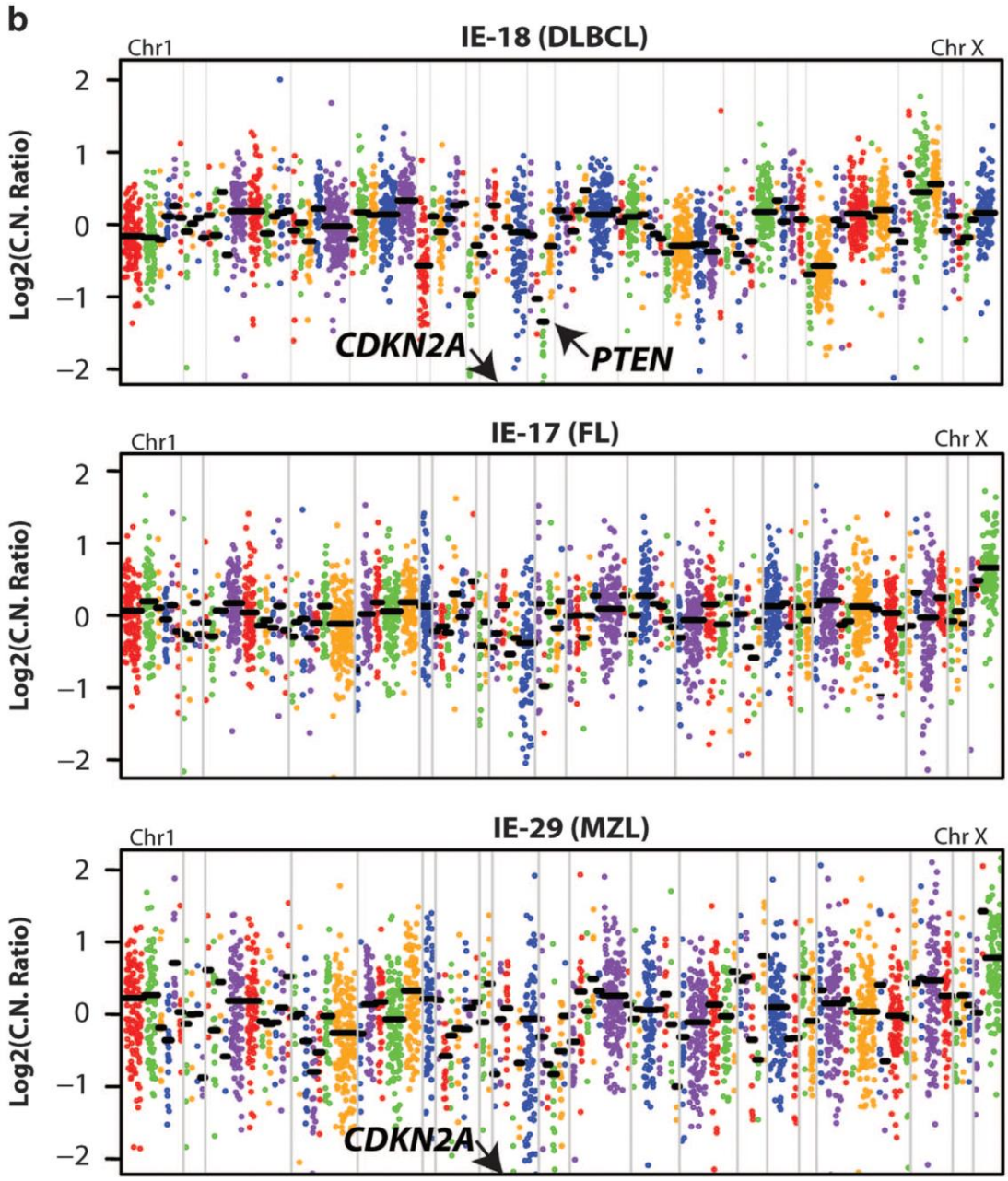


Figure 33. Copy number analysis of orbital and ocular adnexal lymphomas from next-generation sequencing (NGS) data. For each sequenced orbital and ocular adnexal lymphoma, GC-content corrected, normalized read counts per amplicon were divided by those from a composite normal sample, yielding a tumor-to-normal copy number ratio for each amplicon. Gene-level copy number estimates were determined by taking the weighted mean of the per-probe copy number ratios. (**A**) Summary of gene-level copy number ratios (log2) for all profiled orbital and ocular adnexal lymphomas stratified by subtype according to the legend. Selected genes of interest with high-level copy number alteration are indicated. Copy number plot for IE-23 was not informative and was removed from Figure 33A. (**B**). Copy number profiles for three individual orbital and ocular adnexal lymphomas, IE-18 (diffuse large B-cell lymphoma, DLBCL), IE-17 (follicular lymphoma, FL), and

IE-29 (marginal zone lymphoma, MZL). Log2 tumor-to-normal copy number ratios per amplicon are plotted (with each individual amplicon represented by a single dot and each individual gene indicated by different colors), with gene-level copy number estimates (black bars) determined by taking the weighted mean of the per-probe copy number ratios. Selected high-level copy number alterations are indicated. Log2 copy number ratios for *CDKN2A* in both IE-18 and IE-29 are off the scale (IE-18 = − 4.26; IE-29 = − 2.37). IE-17 shows no copy number changes.

A.2.5 Discussion

We performed targeted NGS of 38 FFPE OOALs (MZL, FL, and DLBCL types) to identify somatic mutations and copy number alterations associated with tumorigenesis and identify novel potential therapeutic strategies. Importantly, although numerous studies have profiled extra-ocular NHLs and B-cell neoplasms using NGS, OOALs have not been profiled by comprehensive approaches, and it is unclear if extra-ocular NHLs and OOALs share similar alterations and potential therapeutic targets. Of note, the NGS panel used herein was designed to target pan-cancer altered genes filtered to those with near term potential actionability and a modified, solid tumor-specific version[102] is being used in the MATCH trial, a multi-site, basket trial sponsored by the National Cancer Institute that aims to match patients to investigational therapeutics based on their prioritized mutation profile rather than site of tumor origin.

The most frequently mutated gene in our cohort was *MYD88*, with 28% of OOALs harboring GoF hot-spot mutations. Similar *MYD88* mutations are frequent in non-orbital B cell neoplasms, including Waldenstrom's Macroglobulinemia (79-100%), DLBCL (6-39%), and MZL (4-15%)[224, 228-233]. Interestingly, *MYD88* mutations occur in ~70% of primary DLBCL of the central nervous system (also known as primary central nervous system lymphoma, or PCNSL),[234] and primary vitreoretinal B-cell lymphoma, a subtype of PCNSL[235, 236]. However, *MYD88* mutations have been reported infrequently (0-7%) in

extranodal MZLs[231, 233, 237-240] the most common form (76-100%) of OOAL[231, 233, 237, 238, 241]. Hence, the observed rates of *MYD88* mutations in our OOAL MZLs (25%) and DLBCLs (71%) are notably higher than those in non-orbital NHLs. Importantly, in addition to numerous clinical trials targeting downstream MYD88-dependent factors in the NF-κB signaling pathway (such as phosphorylated Bruton tyrosine kinase [pBTK] and interleukin-1 receptor-associated kinases [IRAK-1 and -4]) and those stratifying response by *MYD88* p.L265P mutation status, a currently recruiting clinical trials in DLBCL assessing a TLR inhibitor requires *MYD88* p.L265P mutation as an entry criterion (https://clinicaltrials.gov/ct2/results?term=myd88; accessed 8/23/15). In summary, our results support evaluation of therapeutic strategies targeting MYD88 and downstream mediators activated by GoF mutations in OOALs.

Alterations in chromatin modifying genes such as those that encode histone lysine methyltransferases (e.g. *KMT2B, KMT3B* and *EZH2*), and ATP-dependent chromatin remodelers (e.g *ARID1A*) are amongst the most common mutations in human cancers, including many forms of B-cell lymphomas, especially DLBCLs and FLs.[210, 225-227, 242-251] Indeed, in our cohort, prioritized mutations in chromatin modifying genes were found only in DLBCLs and FLs (6 of 16 [38%] with prioritized alterations). Specifically, 2 of 9 (22%) FLs and 1 of 7 (14%) DLBCLs harbored prioritized *EZH2* or *ARID1A* mutations, rates that are consistent with extra ocular/adnexal FLs and DLBCLs.[225-227, 252]. Given its association with poor prognosis in several cancers, there has been intense interest in the development of EZH2 inhibitors[253-255]. At present, there are three trials (https://clinicaltrials.gov/ct2/results?term=ezh2; accessed 8/23/15) currently recruiting individuals with relapsed or refractory B-cell lymphomas (including DLBCLs and FLs) for evaluation of oral EZH2 inhibitors in phase 1 and 2 clinical trials. In these trials, the presence of GoF *EZH2* mutations will be tested, in order to determine in which arm an

affected individual will enroll (wild type or mutant *EZH2* status). Likewise, a recent report demonstrated that LoF *ARID1A* mutations in ovarian clear cell carcinomas (OCCCs) render these tumors sensitive to EZH2 inhibitors *in vitro* and *in vivo*[256], suggesting additional ways to target alterations in histone modifiers. Of interest, a DLBCL sample (IE-39) contained both a GoF *EZH2* mutation (p.Y646S) and a LoF *ARID1A* mutation (p.Q1363X), suggesting potentially enhanced sensitivity to EZH2 inhibition. In summary, our results support frequent potentially targetable alterations in histone/chromatin modifiers in OOALs, consistent with results in non-orbital/ocular NHLs.

In recent years, significant interest has emerged in the potential clinical value of NGS of tumoral DNA in order to discover or guide strategies that link personalized therapies to specific genomic alterations present in the cancer. For the first time, we apply comprehensive NGS genomic profiling to routine FFPE OOALs, a cancer subtype for which comprehensive NGS based profiling has not been reported. Of note, the NGS panel used herein was designed to target pan-cancer altered genes filtered to those with near term potential actionability and a modified, solid tumor-specific version[102] is being used in the MATCH trial, a multi-site, basket trial sponsored by the National Cancer Institute that aims to match patients to investigational therapeutics based on their prioritized mutation profile rather than site of tumor origin.

Limitations of our study include a single site cohort, relatively few samples for each subtype, and the use of a targeted panel based approach without assessment of other key alterations (e.g. chromosomal rearrangements, methylation and transcriptional profiling). Hence, larger, prospective, multi-institutional studies will be required to confirm our findings and clinical trials are needed in OOALs to test potential treatment strategies proposed herein.

Here, through NGS targeting informative/potentially actionable genomic alterations, we

identified recurrent mutations and CNAs across 36 OOALs, including MZL, FL and DLBCL subtypes. We identified *MYD88* GoF mutations at higher rates in orbital/ocular adnexal MZL and DLBCLs (25% and 71%, respectively) than those reported in non-orbital/ocular sites. Likewise, histone/chromatin modifiers showed frequent alterations in our cohort (at similar rates to non-orbital/ocular adnexal sites), including potentially targetable GoF mutations in the histone methyltransferase *EZH2* (**Figure 34**). As novel therapeutic approaches are urgently needed for OOALs, our results demonstrate the utility of an NGS-based approach to nominate precision therapeutic approaches for OOALs and other ocular, orbital and adnexal tumors.

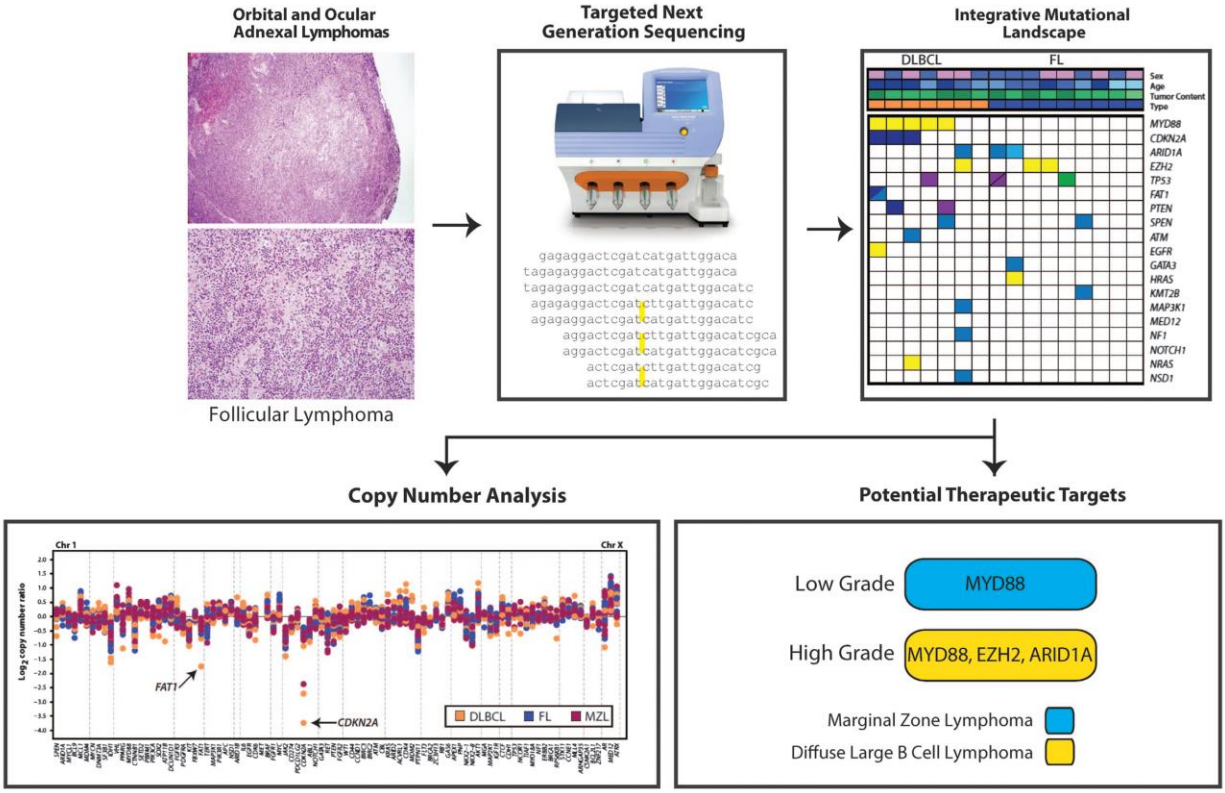


Figure 34 Workflow of determining driver and potentially actionable genomic alterations. Genomic DNA from formalin-fixed, paraffin-embedded orbital and ocular lymphoma tissue enriched by macrodissection is subjected to targeted next-generation sequencing (NGS) using a cancer gene panel. Bioinformatics analysis yields candidate point mutations, small indels and copy number alterations that potentially drive tumor growth and development in orbital and ocular

adnexal lymphomas. Potential therapeutic targets are prioritized and reported.

A.2.6 Acknowledgements

This work was supported by the National Eye Institute (K12EY022299), National Cancer Institute University of Michigan Comprehensive Cancer Center Support Grant (P30CA046592), The Leonard G. Miller Ophthalmic Research Fund at the Kellogg Eye Center, Barbara Dunn Research Fund, Beatrice & Reymont Paul Foundation and March Hoops to Beat Blindness to RCR. RCR is the Leslie H. and Abigail S. Wexner Emerging Scholar and SAT is the A. Alfred Taubman Emerging Scholar, both of the A. Alfred Taubman Medical Research Institute, which supported, in part, this study.

Appendix III

Next generation sequencing of vitreoretinal lymphomas from small-volume intraocular liquid biopsies: new routes to targeted therapies[4]

A.3.1 Abstract

Vitreoretinal lymphoma (VRL), the most common lymphoma of the eye, is a rare form of primary CNS lymphoma (PCNSL). Most frequently a high-grade diffuse large B cell lymphoma, VRL can cause vision loss and its prognosis remains dismal: the overall survival time is 3 years after diagnosis. Radiotherapy and chemotherapy are used but remain frequently ineffective, and no standardized treatment regimen exists. Furthermore, no biologically targeted treatments, based on the genetic profile of the tumor, are available, as VRL has hitherto not comprehensively been profiled. To address these unmet needs, we hypothesized that a next generation sequencing (NGS)-based, National Cancer Institute (NCI) MATCH Trial-modified panel would be able to identify actionable genomic alterations from small-volume, intraocular liquid biopsies. In this retrospective study, we collected diluted vitreous biopsies from 4 patients with confirmed or high suspicion for VRL. Following cytological confirmation of lymphoma (all were diffuse large B cell lymphomas), we subjected genomic DNA from the liquid biopsies to NGS, using a panel containing 126 genes (3,435 amplicons across several hotspots per gene), which was modified from that of the NCI MATCH Trial, a precision oncology clinical trial that matches patients with

[4] This study has previously appeared in Cani, A.K., et al. Oncotarget, 2017. (169. Cani, A.K., et al., *Next generation sequencing of vitreoretinal lymphomas from small-volume intraocular liquid biopsies: new routes to targeted therapies.* Oncotarget, 2017. **8**(5): p. 7989-7998.)

cancers to investigational therapeutics based on their prioritized mutation profile rather than site of tumor origin. Using a validated bioinformatics pipeline, we assessed for the presence of actionable mutations and copy number alterations.

In all four small-volume, intraocular liquid biopsies, we obtained sufficient genomic DNA for analysis, even in diluted samples in which the undiluted vitreous was used for cytology and flow cytometry. Using NGS, we found targetable heterozygous gain-of-function mutations in the *MYD88* oncogene, and confirmed in our cohort the presence the p.L265 mutations, previously described using PCR-based assays. For the first time in VRL, we also identified the *MYD88* p.S243N mutation. We also identified two-copy copy number losses in the tumor suppressor *CDK2NA* in all four cases, and one copy loss of the tumor suppressor *PTEN* in one sample. In one case, in which vitreous biopsies were originally read as cytologically negative, but which was confirmed as lymphoma when a lesion appeared in the brain two years later, our NGS-based approach detected tumoral DNA in the banked, original liquid biopsy.

In conclusion, we performed the first systematic exploration of the actionable cancer genome in VRL. Our NGS-based approach identified exploitable genomic alterations such as gain-of-function MYD88 oncogene mutations and loss of the tumor suppressor CDK2NA, and thus illuminates new routes to biologically targeted therapies for VRL, a cancer with a dismal prognosis. This precision medicine strategy could be used to nominate novel, targeted therapies in lymphomas and other blinding and deadly ocular, orbital, and ocular adnexal diseases for which few treatments exist.

A.3.2 Introduction

Vitreoretinal lymphoma (VRL), the most common lymphoma of the eye, is linked closely to CNS lymphoma (PCNSL), and affects ~380 individuals in the U.S per year[257]. Fifty to 90% of those with primary VRL have concurrent (or will develop) PCNSL, while 15-25% of patients with primary PCNSL develop VRL[257]. Often mistaken for uveitis, delays in VRL diagnosis are common. Diagnosis relies on small-volume intraocular (vitreous) fluid biopsy, yet the limited number of malignant cells recovered complicates cytological analysis, especially in distinguishing from inflammatory lymphoid infiltrates. Flow cytometry, immunohistochemistry, and cytokine analyses are suggestive, but not diagnostic. The high viscosity of vitreous fluid often interferes with sampling and can confound cytological analyses and flow cytometry, which can lead to false-negatives, though this can be dependent on experience of consultant pathologists[211, 258]. Most frequently a high-grade diffuse large B cell lymphoma (DLBCL), VRL can cause vision loss, with progression-free survival of ~1 yr and overall survival less than 3 years after diagnosis. Radiotherapy and chemotherapy are used but are frequently ineffective: no standardized treatments exist. Since the genomic landscape of VRLs have never been comprehensively profiled, no molecularly targeted treatments exist. To address these unmet needs, we hypothesized that a next generation sequencing (NGS)-based panel, a modified version of that used in the National Cancer Institute-Molecular Analysis for Therapy Choice (NCI-MATCH) Trial (NCT02465060),[102, 168] would identify actionable genomic alterations from small-volume, intraocular liquid biopsies and nominate precision medicine-based treatment strategies. The modified NCI-MATCH panel we used herein is designed to detect actionable alterations in both advanced solid tumors and lymphomas (http://www.cancer.gov/about-cancer/treatment/clinical-trials/nci-supported/nci-match)[113]. We analyzed four cytology-confirmed VRL cases, which

represents about 1% of the VRL cases that occur in the U.S. annually (~380/yr)[257].

A.3.3 Materials and Methods

A.3.3.1 Case selection

The study was carried out at the highest ethical standards and with the approval of the University of Michigan Institutional Review Board. We identified a cohort of three archived, flash-frozen vitreous specimens linked to biopsy-proven VRL or PCNSL from the intraocular liquid biobank (Cases 101-103) and one fresh specimen highly suspicious for VRL (Case 104) at the Kellogg Eye Center, Department of Ophthalmology & Visual Sciences at the University of Michigan for next generation sequencing (NGS). Clinicopathological information for each case was obtained from the clinical archive.

A.3.3.2 Targeted next generation sequencing (NGS)

Targeted next generation sequencing was performed essentially as previously described, with few modifications relating to sample preparation[101-103, 168]. Each intraocular liquid specimen was thawed and centrifuged to pellet tumor cells. Genomic DNA was isolated using the Qiagen All-Prep formalin-fixed paraffin-embedded DNA/RNA kit (Qiagen, Valencia, CA) and quantified as previously described. Targeted, multiplexed NGS was performed on the Ion Torrent platform using the DNA component of a beta version of the Oncomine Comprehensive Assay (OCP), a custom panel comprised of 3,435 amplicons targeting 126 genes. Genes included in this panel were selected based on pan-solid tumor next generation sequencing and copy number profiling data analysis that prioritized somatic, recurrently altered oncogenes, tumors suppressors, genes present in high level copy gains/losses and known/investigational therapeutic targets[102]. Library preparation with barcode incorporation, template preparation and sequencing using the

Ion Torrent Proton sequencer were performed according to the manufacturer's instructions. Data analysis was performed using Torrent Suite 4.0.2, with alignment by TMAP using default parameters, and variant calling using the Torrent Variant Caller plugin (version 4.0r76860) using default low-stringency somatic variant settings. Variant annotation filtering and prioritization was performed essentially as described using validated in house pipelines[101-103, 259, 260]. Briefly, called variants were filtered to remove synonymous or non-coding variants, those with flow corrected read depths (FDP) less than 20, flow corrected variant allele containing reads (FAO) less than 6, variant allele frequencies (FAO/FDP) less than 0.10, extreme skewing of forward/reverse flow corrected reads calling the variant (FSAF/FSAR <0.2 or >5), or indels within homopolymer runs >4. Called variants were filtered using a panel-specific, in house blacklist. Variants with allele frequencies >0.5% in Exome Sequencing Project 6500 (ESP6500) or the 1000 Genomes project, and those reported in ESP6500 or 1000 genomes with observed variant fractions between 0.40 and 0.60 or > 0.9 were considered germline variants unless occurring at a known hot-spot variant. Variants located at the last mapped base (or outside) of amplicon target regions, variants with the majority of supporting reads harboring additional mismatches or indels (likely sequencing errors), those in repeat–rich regions (likely mapping artifacts), and variants occurring exclusively in one amplicon if overlapping amplicons cover the variant were excluded. High confidence somatic variants passing the above criteria were then visually confirmed in Integrative Genomics Viewer (https://www. broadinstitute.org/igv/). We have previously confirmed that these filtering criteria identify prioritized high-confidence somatic variants that pass Sanger sequencing validation with >95% accuracy[101, 103, 220, 259, 260]. Copy number analysis from total amplicon read counts provided by the Coverage Analysis Plug-in (v4.0-r77897) was performed essentially as described using a validated approach[101-103, 106, 168]. Genes with a

log2 copy number estimate of <-1 or >0.6 were considered to have high level loss or gain, respectively. To prioritize potential driving alterations, we utilized Oncomine software tools (https://powertools. oncomine.com) to annotate called variants, which use pan-cancer next generation sequencing data to identify genes as oncogenes or tumor suppressors, based on overrepresentation of hot-spot or deleterious mutations, respectively. Variants in oncogenes are then considered gain-of-function if at a hot-spot and variants in tumor suppressors are considered loss-of-function if deleterious or at a hot-spot[101, 102]. Likewise, high-level copy number alterations were prioritized if they were concordant with the minimal common region analysis used to design the Oncomine Comprehensive Assay (e.g. high level copy number gain in a gene prioritized as amplified/deleted by minimal common region analysis.

A.3.4 Results

We collected four small-volume vitreous biopsies from four patients with a high suspicion for VRL. All (cases 101-104) were males in their 60s. Cases 101 and 102 were diagnosed with PCNSL prior to biopsy, and showed cytology-proven VRL on vitreous biopsy (DLBCL). Cases 103 and 104 underwent vitreous biopsy in both eyes after developing vitreous debris and subretinal infiltrates bilaterally (Case 103, **Figure 35A-H**), yet vitreous cytological analyses were negative. Two years later, Case 103 developed vision loss with right hemianopia (**Figure 36A**) and MRI (**Figure 36B**) revealed a lesion in the right optic nerve and chiasm. Since cytologic analysis of CSF fluid confirmed PCNSL (DLBCL), the patient's earlier ocular presentation was presumed bilateral VRL. Case 104 exhibited painless and chronic, bilateral vitreous debris for 2 years with a negative workup for uveitis or other systemic causes of inflammation. This patient also had a right parietal lobe lesion on MRI, suggestive of PCNSL with VRL.

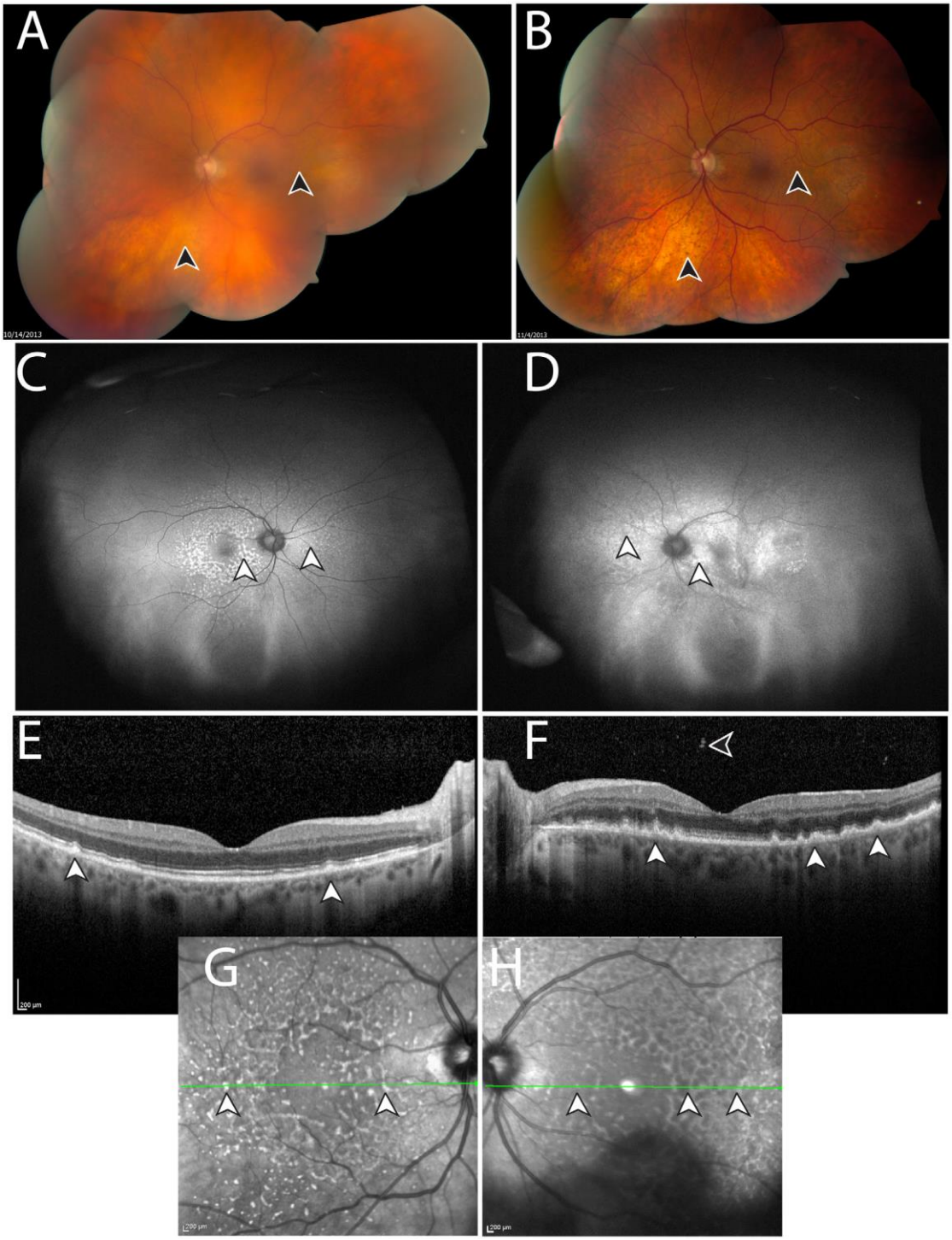


Figure 35 Manifestations of vitreoretinal lymphoma in Case 103. A. Montaged fundus photo of the left eye with vitreous debris prior to intraocular liquid biopsy and vitrectomy. Lymphoma cells are suspended in the vitreous, resulting in a "hazy" view, which obscures anatomic details of the

retina (arrowheads). **B.** Following intraocular liquid biopsy and vitrectomy, which did not detect malignant cells, the media of the left eye is clear and retinal details can be discerned, such as subretinal lipofuscin clumps, and sub-retinal pigment epithelium (RPE) deposits, which manifest in a yellow and dark stippled, leopard-like pattern (arrowheads). Ultra-wide field fundus autofluorescence of the right **C.** and left **D.** eye, shows stippled hyper-autofluorescence corresponding to the lymphomatous sub-RPE deposits (arrowheads). Optical coherence tomography of the right **E.** and left **F.** eye shows nodular hyperreflective lymphomatous lesions at the RPE level (arrowheads). Prior to biopsy of the left eye (F), lymphoma cells can be seen in the posterior vitreous. Insets **G, H.** represent near infrared reflectance imaging of the right (G) and left (H) eyes, which highlight the leopard-like pattern of the sub-RPE lymphomatous macular infiltrates. Green lines and arrowheads of insets (G, H) correspond to the cross sectional plane of the OCT images in (E) and (F). Similar to (A), autofluorescence (D), OCT (F), and near infrared reflectance imaging (H) in the left eye appear blurry compared to the right eye due to the presence of lymphoma cells in the vitreous. Except for (B), images were obtained following biopsy and vitrectomy in the right eye (C, E, G) but prior to these interventions in the left eye (A, D, F, H). During this time, visual acuity was within normal range.

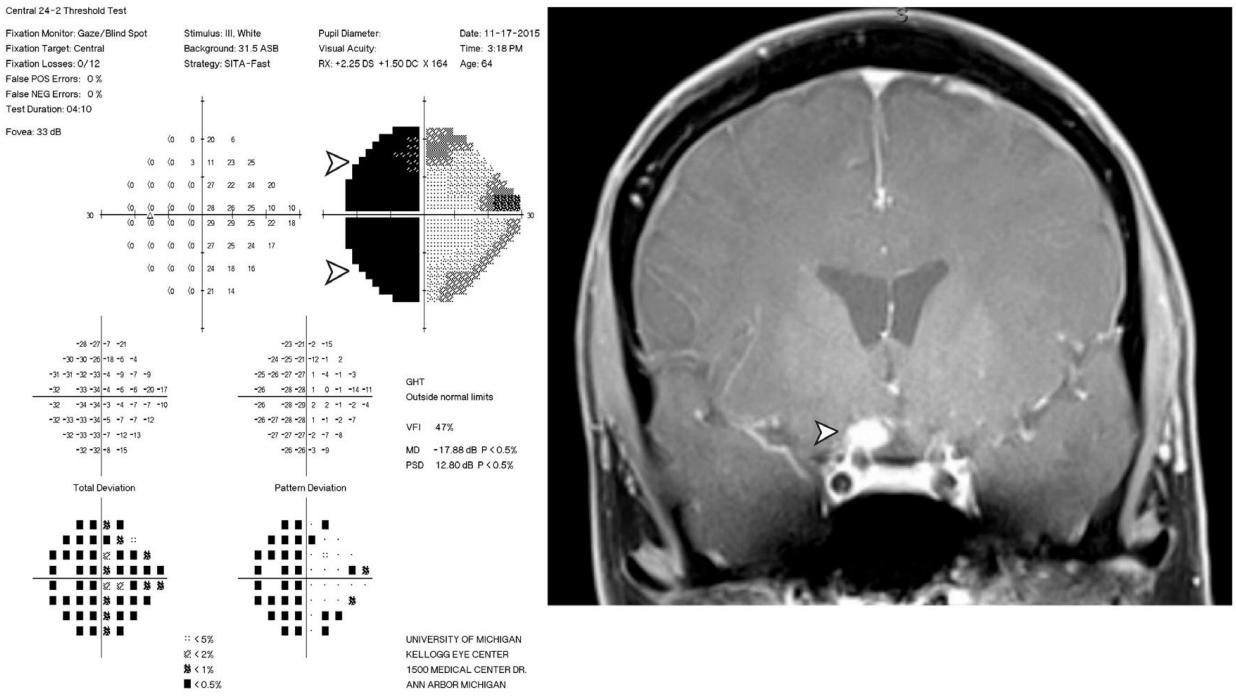


Figure 36 Functional and structural cause of vision loss in Case 103. Two years following negative cytology results from intraocular liquid biopsies in each eye, the patient developed severe vision loss in the left eye. **A**. Humphrey 24-2 visual field testing report of the left eye revealed a dense temporal hemianopia, denoted by black area (arrowheads), which signified lack of sensitivity to light stimulus in the temporal half of visual field. **B**. Magnetic resonance imaging with gadolinium contrast revealed an enhancing lesion abutting the right optic chiasm and optic nerve. Subsequent cytological analysis

of cerebrospinal fluid confirmed the diagnosis of primary CNS lymphoma, diffuse large-B cell subtype.

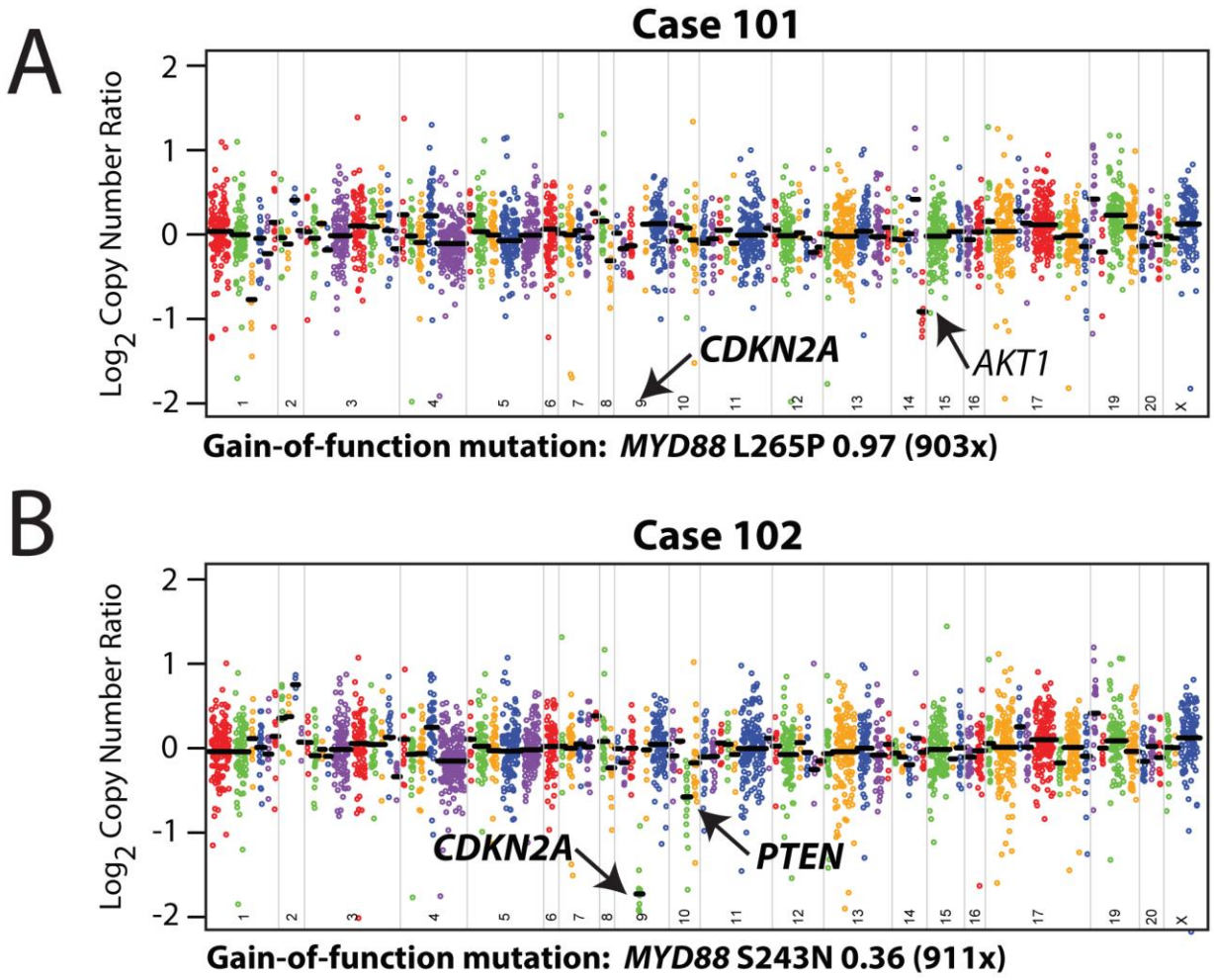
Sufficient genomic DNA was recovered from each vitreous biopsy (**Table 1**). NGS was performed using the DNA component of the Oncomine Comprehensive Assay, a panel comprised of 3,435 amplicons targeting 126 genes, a modified version of which is being used in the NCIMATCH trial[113], and sequencing on the Ion Torrent Proton sequencer[168]. Targeted genes were selected based on pansolid tumor NGS and copy number profiling data analysis to prioritize somatic, recurrently altered oncogenes, tumors suppressors and genes present in high level copy number alterations, filtered by available or investigational therapeutic targets. All four samples provided informative data, and sample statistics are listed in **Table 1**.

Case	DNA (ng)	Mapped Reads	On Target	Depth	Uniformity
101	412.5	2,636,783	98.86%	832.6	93.95%
102	247	2,778,318	99.06%	872.3	93.76%
103	5.8	3,548,338	99.04%	1,128	94.73%
104	26	4,377,192	98.90%	1,144	62.77%
Average	172.8	3,335,157	98.97%	994	86.30%
SD	193.6	801,889	0.10%	164	15.69%

Table 14. Summary of NGS Statistics. Means and corresponding standard deviations (SD) are given for the amount of genomic DNA isolated from each intraocular liquid biopsy, number of mapped reads, on target reads, coverage depth, and uniformity in next generation sequencing summary table for each case.

We identified one non-synonymous point mutation each in Cases 101-103 and one high level deletion in each sample (**Figure 37A-D**). Three of the four VRLs (75%) harbored prioritized gain-of-function mutations in the Toll/interleukin-1 receptor (TIR) domain of *MYD88*. MYD88 is an adaptor protein that binds to the intracellular domains of Toll-like receptors (TLRs) and interleukin-1 receptor on B-cells and macrophages, which stimulates NF-κB signaling pathway,

and is involved in innate immunity[224]. These included two p.L265P (cases 101 homozygous and 103 heterozygous) and one p.S243N (case 102 heterozygous) mutations all at presumed clonal, variant frequency (**Figure 37A-C**). We also assessed somatic copy number alterations from the same NGS data using a validated approach[106], with case-specific gene (and amplicon) level copy number plots (**Figure 37A-D**). We identified few copy number alterations in our VRL samples, however we identified prioritized, high level copy loss of CDKN2A in all samples. High level copy loss was observed in cases 101, 102, and 104, while case 103 also harbored a high level loss given the presumed tumor content of 30% (based on presumably heterozygous *MYD88* p.L265P at 0.15 variant allele frequency). We also detected a one-copy deletion in *PTEN* in case 102. Lastly, although not prioritized, we identified a focal one-copy loss in AKT1 in case 101 and low level gain of chromosome 19 was observed in case 103 (targeted genes included *STK1*, *MLL4*/*KMT2D*, and *ARHGAP43*).


A
Case 101
Log2 Copy Number Ratio
2
1
0
-1
-2
CDKN2A
AKT1
Gain-of-function mutation: MYD88 L265P 0.97 (903x)
B
Case 102
Log2 Copy Number Ratio
CDKN2A
PTEN
Gain-of-function mutation: MYD88 S243N 0.36 (911x)

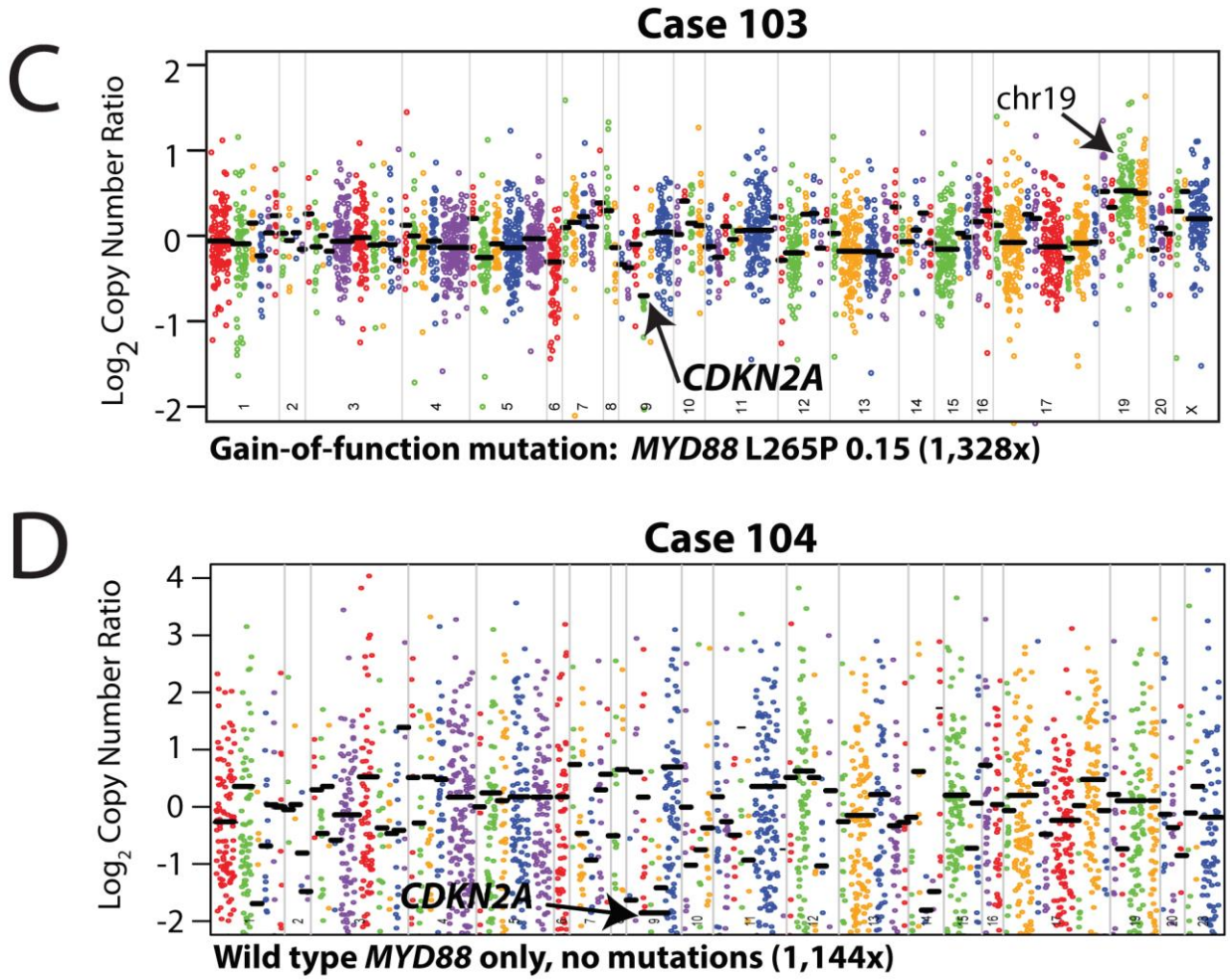


Figure 37 Mutation and copy-number analysis of vitreoretinal lymphomas from next generation sequencing data. Copy-number profiles, with prioritized alterations in *MYD88* (including point mutation/variant, variant fraction in the sample, and overall coverage depth [sum of the number of variant-containing reads and the number of reads without the variant]) are listed below each plot, for each of the three vitreoretinal lymphomas: **A.** Case 101, **B.** Case 102, **C.** Case 103, and **D.** Case 104. For each sequenced vitreoretinal lymphoma case, GC-content corrected, normalized read counts per amplicon were divided by those from a composite normal sample, yielding a tumor-to-normal copy-number ratio for each amplicon. Log2 tumor-to-normal copy-number ratios per amplicon are plotted (with each individual amplicon represented by a single dot, and each individual gene indicated by different colors), with gene-level copy-number estimates (black bars) determined by taking the weighted mean of the per-probe copy-number ratios. Prioritized high-level copy number alterations are indicated in bold. Log2 copy-number ratios for *CDKN2A* for Case 101 (A) are off the scale. Annotated, but not prioritized, one copy number losses in *AKT1* (A, Case 101), and low level copy number gains in genes on chromosome (chr) 19 (C, Case 103) are noted by arrows.

A.3.5 Discussion

We demonstrate the feasibility of targeted NGS on intraocular liquid biopsies using minute volumes (as little as 500 microliters) of non-diluted and diluted vitreous. Our approach does not compromise the volumes needed for cytology-based and other diagnostics (e.g. flow cytometry). Our NGS approach may overcome difficulties related to high vitreous viscosity, poor cellular preservation, low cellularity, false-negative and false-positive PCR-based results, which often leads to delays in diagnosis and therapy (**Figure 38**)[261]. For instance, in case 103, presumably false negative cytologic interpretations of vitreous biopsies in both eyes led to a two-year delay in treatment, when PCNSL cells were finally detected in the CSF. Despite repeated false-negative findings in Cases 103 and 104, our NGS strategy detected actionable genomic alterations in as little as 5.8 ng of tumor DNA from the original biopsy. *MYD88* GOF mutations were found in 3 of 4 samples (75%) and high level copy number loss of *CDKN2A* were found in all 4 samples (100%). All *MYD88* mutations occurred in the TIR domain, and 2 of 4 samples (50%) harbored the p.L265P point mutation. We detected *MYD88* p.S243N mutation (1 of 4 samples, 25%), high level *CDKN2A* loss (4 of 4 samples, 100%), and low level *PTEN* loss (1 of 4 samples, 25%) for the first time in VRL. Nearly all of these alterations are potentially targetable. For instance, currently recruiting clinical trials in DLBCLs assessing TLR inhibitors require the *MYD88* p.L265P mutation as an entry criterion (https://clinicaltrials.gov/ct2/results?term=myd88; accessed 8/8/16). Both p.L265P and p.S243N *MYD88* mutations show high levels of NF-κB transactivation, which can potentially be targeted by the Bruton's kinase inhibitor ibrutinib, and IRAK1/4 antagonists[229, 262, 263]. *MYD88* mutations other than p.L265P make up a quarter of

all *MYD88* mutations in patients with DLBCLs and unlike previous limited studies that solely evaluate for the p.L265P mutation (and one study that reported p.P258L)[235, 236, 264], our comprehensive NGS-based approach reveals other potentially actionable, and diagnostic, GOF *MYD88* alterations in VRL. High level *CDKN2A* loss is considered potentially targetable, as ilorasertib (an inhibitor of Aurora, VEGF, and PDGF tyrosine kinase families; https://clinicaltrials. gov/ct2/show/NCT02540876; accessed 8/8/16) and palbociclib (a CDK4/6 inhibitor; https://clinicaltrials. gov/ct2/show/NCT02693535; accessed 8/8/16) are being tested in advanced *CDKN2A*-deficient tumors. Likewise, although observed at only a single copy loss in our study, *PTEN*-deficient tumors are being targeted through PI3K-beta inhibition (https://clinicaltrials.gov/ct2/show/ NCT02465060), as well as similar strategies targeting other components of the PTEN-PI3K-AKT pathway (e.g. https://clinicaltrials.gov/ct2/show/NCT02761694 and others). Although not prioritized, an annotated one-copy loss in the oncogene *AKT1* was observed in case 101. While oncogenic *AKT1* GOF mutations and amplifications have been linked to a variety of cancers [265], a recent report has found recurrent one-copy losses of *AKT1* in DLBCLs[266]. How a one-copy loss of a known oncogene like *AKT1* might be related to DLBCL/VRL tumorigenesis requires further investigation.

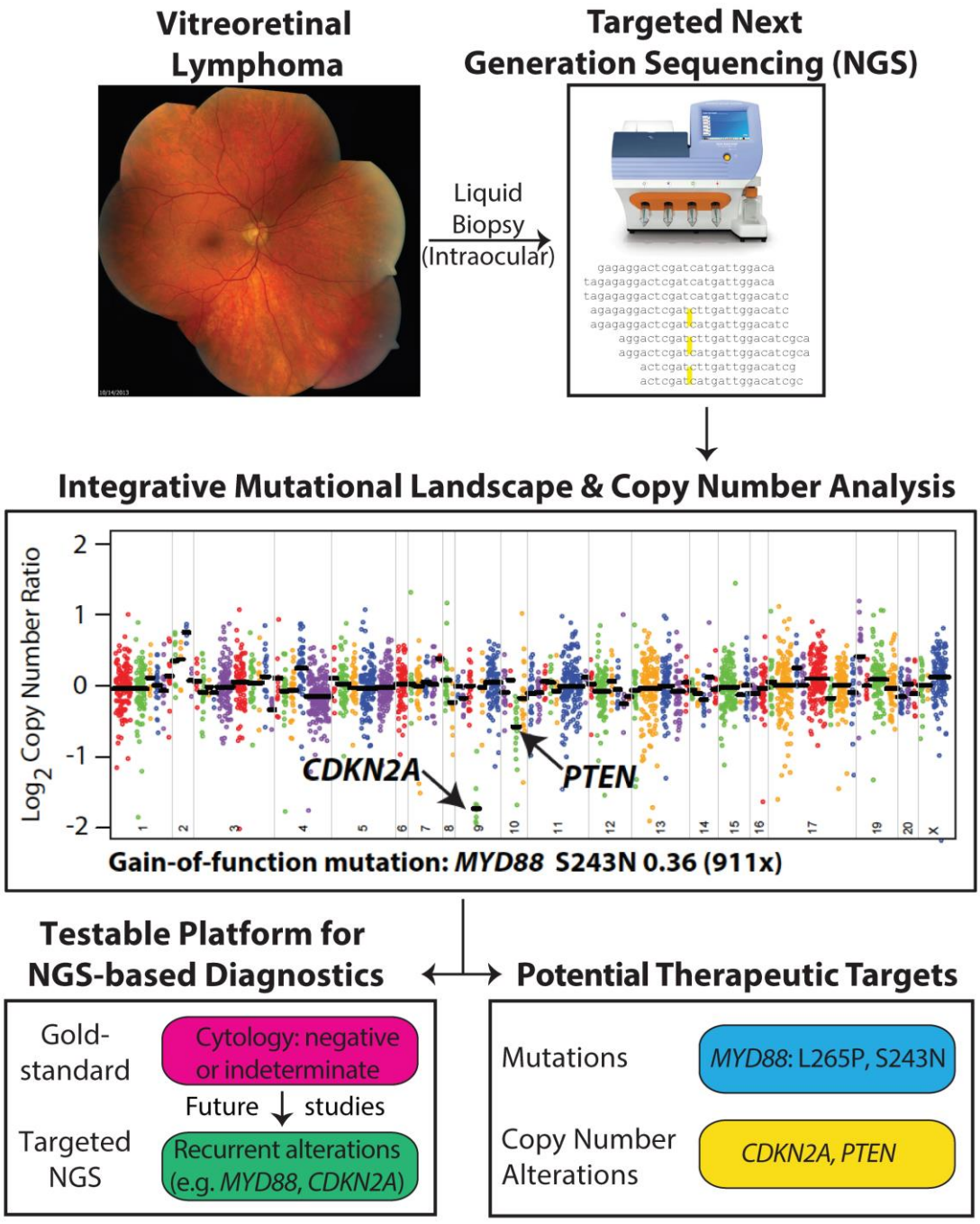


Figure 38 Workflow of determining VRL driver and potentially actionable genomic alterations. Genomic DNA from intraocular liquid biopsies is subjected to targeted next generation sequencing using a cancer gene panel. Bioinformatics analysis yields candidate point mutations and copy number alterations that potentially drive tumor growth and development in

vitreoretinal lymphomas. Potentially actionable therapeutic targets are prioritized and reported.

There are limitations in our study. While our analyzed samples represent > 1% of the number of VRLs that occur annually in the U.S., our sample size of 4 is modest. Future studies should investigate a larger cohort of VRL samples. While VRL is considered a subtype of PCNSL, similarities in their respective genetic landscapes have not yet been explicitly tested. A future investigation should focus on comparing the genetic alterations between PCNSL and VRL, especially in paired sample sets in patients with lymphoma cells in the eye (vitreous or choroid) and elsewhere in the CNS (cerebrospinal fluid, brain, spinal cord). Finally, while the modified, NCI-MATCH panel we employ is designed to detect alterations in solid tumors and lymphoma, a custom NGS panel enriched for potential non-Hodgkin lymphoma targets may be able to capture alterations not covered in our panel. As improved therapies are urgently needed for VRLs due to dismal survival rates, high rates of relapse as well as local and vision-related toxicities of present treatments (e.g. radiotherapy), our results demonstrate the utility of an NGS-based approach to nominate precision therapeutic approaches for VRLs. The route of administration of therapy employed would depend on how currently available targeted agents (FDA-approved for other cancers or in clinical trials) are delivered, and may be systemic (oral, intravenous, intrathecal, etc). Our NGS-based strategy is not intended to replace standard diagnostic methods, but to complement it. In cases in which VRL is confirmed by standard-of-care cytology-based diagnostic methods, our NGS-based strategy would be an adjunctive approach to characterize actionable targets and nominate potential therapeutic strategies. In our previous work, we confirmed that the filtering criteria employed here identify prioritized high-confidence somatic variants in tumoral

DNA that pass Sanger sequencing validation, the main comparator technique, with >95% accuracy[102]. This approach could potentially be used for detecting intraocular recurrence and monitoring treatment response, which currently remains difficult. If our results are confirmed in larger cohort analyses and are employed as an adjunctive strategy for VRL, we recognize that NGS-based testing may not be feasible for every clinical site. We envision that after collection of undiluted vitreous for cytology and other point-of-care diagnostic tests, diluted samples could be flash frozen and sent to centers with expertise in genomic DNA collection from vitreous, NGS, and bioinformatics analyses.

To our knowledge, this is the first NGS study to comprehensively profile VRLs. This precision medicine approach may have potential diagnostic and therapeutic applications in characterizing small-volume, liquid biopsies in order to nominate rationally driven clinical trials for ocular diseases with few therapeutic options.

A.3.6 Acknowledgements

This work was supported by the National Eye Institute (K12EY022299), National Cancer Institute University of Michigan Comprehensive Cancer Center Support Grant (P30CA046592), The Leonard G. Miller Ophthalmic Research Fund at the Kellogg Eye Center, Barbara Dunn Research Fund, Beatrice & Reymont Paul Foundation and March Hoops to Beat Blindness to RCR. RCR is the Leslie H. and Abigail S. Wexner Emerging Scholar and SAT is the A. Alfred Taubman Emerging Scholar, both of the A. Alfred Taubman Medical Research Institute, which supported, in part, this study. None of the sponsors participated in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

BIBLIOGRAPHY

1. Society, A.C., *Cancer Facts and Figures 2019*, in *Atlanta: American Cancer Society*. 2019.
2. Howlader N, N.A., Krapcho M, Miller D, Brest A, Yu M, Ruhl J, Tatalovich Z, Mariotto A, Lewis DR, Chen HS, Feuer EJ, Cronin KA (eds). *SEER Cancer Statistics Review, 1975-2016*. 2019; Available from: https://seer.cancer.gov/csr/1975_2016/.
3. Mariotto, A.B., et al., *Projections of the cost of cancer care in the United States: 2010-2020.* J Natl Cancer Inst, 2011. **103**(2): p. 117-28.
4. Warren, J.L., et al., *Evaluation of trends in the cost of initial cancer treatment.* J Natl Cancer Inst, 2008. **100**(12): p. 888-97.
5. Bradley, C.J., et al., *Productivity costs of cancer mortality in the United States: 2000-2020.* J Natl Cancer Inst, 2008. **100**(24): p. 1763-70.
6. Curtin, S.C., *Trends in Cancer and Heart Disease Death Rates Among Adults Aged 45–64: United States, 1999–2017.* Natl Vital Stat Rep, 2019. **68**(5): p. 1-8.
7. Banerjee, P.P., et al., *Androgen action in prostate function and disease.* Am J Clin Exp Urol, 2018. **6**(2): p. 62-77.
8. Yaghjyan, L. and G.A. Colditz, *Estrogens in the breast tissue: a systematic review.* Cancer Causes Control, 2011. **22**(4): p. 529-40.
9. Levin, E.R., *Elusive extranuclear estrogen receptors in breast cancer.* Clin Cancer Res, 2012. **18**(1): p. 6-8.
10. Klinge, C.M., *Estrogen receptor interaction with estrogen response elements.* Nucleic Acids Res, 2001. **29**(14): p. 2905-19.
11. Tan, M.H., et al., *Androgen receptor: structure, role in prostate cancer and drug discovery.* Acta Pharmacol Sin, 2015. **36**(1): p. 3-23.
12. Katzenellenbogen, B.S., et al., *Molecular mechanisms of estrogen action: selective ligands and receptor pharmacology.* J Steroid Biochem Mol Biol, 2000. **74**(5): p. 279-85.
13. Lindzey, J., et al., *Molecular mechanisms of androgen action.* Vitam Horm, 1994. **49**: p. 383-432.
14. Huang, B., M. Warner, and J.A. Gustafsson, *Estrogen receptors in breast carcinogenesis and endocrine therapy.* Mol Cell Endocrinol, 2015. **418 Pt 3**: p. 240-4.
15. Häggström, M., Richfield, D., *Diagram of the pathways of human steroidogenesis.* WikiJournal of Medicine, 2014. **1**(1).
16. Johnston, S.R. and B. Yeo, *The optimal duration of adjuvant endocrine therapy for early stage breast cancer--with what drugs and for how long?* Curr Oncol Rep, 2014. **16**(1): p. 358.
17. Gluck, S., *Extending the clinical benefit of endocrine therapy for women with hormone receptor-positive metastatic breast cancer: differentiating mechanisms of action.* Clin Breast Cancer, 2014. **14**(2): p. 75-84.
18. Reinbolt, R.E., et al., *Endocrine therapy in breast cancer: the neoadjuvant, adjuvant, and metastatic approach.* Semin Oncol Nurs, 2015. **31**(2): p. 146-55.
19. Lumachi, F., et al., *Endocrine therapy of breast cancer.* Curr Med Chem, 2011. **18**(4): p. 513-22.
20. Fugh-Berman, A. and S. Epstein, *Tamoxifen: disease prevention or disease substitution?* Lancet, 1992. **340**(8828): p. 1143-5.
21. McDonnell, D.P. and S.E. Wardell, *The molecular mechanisms underlying the pharmacological*

actions of ER modulators: implications for new drug discovery in breast cancer. Curr Opin Pharmacol, 2010. **10**(6): p. 620-8.
22. Carlson, R.W., *The history and mechanism of action of fulvestrant.* Clin Breast Cancer, 2005. **6 Suppl 1**: p. S5-8.
23. Paoletti, C., et al., *Circulating Biomarkers and Resistance to Endocrine Therapy in Metastatic Breast Cancers: Correlative Results from AZD9496 Oral SERD Phase I Trial.* Clin Cancer Res, 2018. **24**(23): p. 5860-5872.
24. Hamilton, E.P., et al., *A First-in-Human Study of the New Oral Selective Estrogen Receptor Degrader AZD9496 for ER(+)/HER2(-) Advanced Breast Cancer.* Clin Cancer Res, 2018. **24**(15): p. 3510-3518.
25. Garner, C., *Uses of GnRH agonists.* J Obstet Gynecol Neonatal Nurs, 1994. **23**(7): p. 563-70.
26. Litwin, M.S. and H.J. Tan, *The Diagnosis and Treatment of Prostate Cancer: A Review.* JAMA, 2017. **317**(24): p. 2532-2542.
27. Tarlatzis, B.C. and E.M. Kolibianakis, *GnRH agonists vs antagonists.* Best Pract Res Clin Obstet Gynaecol, 2007. **21**(1): p. 57-65.
28. Schally, A.V., et al., *Hypothalamic hormones and cancer.* Front Neuroendocrinol, 2001. **22**(4): p. 248-91.
29. van Poppel, H. and S. Nilsson, *Testosterone surge: rationale for gonadotropin-releasing hormone blockers?* Urology, 2008. **71**(6): p. 1001-6.
30. Masiello, D., et al., *Bicalutamide functions as an androgen receptor antagonist by assembly of a transcriptionally inactive receptor.* J Biol Chem, 2002. **277**(29): p. 26321-6.
31. Tran, C., et al., *Development of a second-generation antiandrogen for treatment of advanced prostate cancer.* Science, 2009. **324**(5928): p. 787-90.
32. Scher, H.I., et al., *Antitumour activity of MDV3100 in castration-resistant prostate cancer: a phase 1-2 study.* Lancet, 2010. **375**(9724): p. 1437-46.
33. Chi, K.N., et al., *Apalutamide for Metastatic, Castration-Sensitive Prostate Cancer.* N Engl J Med, 2019. **381**(1): p. 13-24.
34. Davis, I.D., et al., *Enzalutamide with Standard First-Line Therapy in Metastatic Prostate Cancer.* N Engl J Med, 2019. **381**(2): p. 121-131.
35. Potter, G.A., et al., *Novel steroidal inhibitors of human cytochrome P45017 alpha (17 alpha-hydroxylase-C17,20-lyase): potential agents for the treatment of prostatic cancer.* J Med Chem, 1995. **38**(13): p. 2463-71.
36. Fizazi, K., et al., *Abiraterone acetate plus prednisone in patients with newly diagnosed high-risk metastatic castration-sensitive prostate cancer (LATITUDE): final overall survival analysis of a randomised, double-blind, phase 3 trial.* Lancet Oncol, 2019. **20**(5): p. 686-700.
37. Fizazi, K., et al., *Abiraterone plus Prednisone in Metastatic, Castration-Sensitive Prostate Cancer.* N Engl J Med, 2017. **377**(4): p. 352-360.
38. Cancer Genome Atlas, N., *Comprehensive molecular portraits of human breast tumours.* Nature, 2012. **490**(7418): p. 61-70.
39. Cancer Genome Atlas Research, N., *The Molecular Taxonomy of Primary Prostate Cancer.* Cell, 2015. **163**(4): p. 1011-25.
40. Sanchez-Vega, F., et al., *Oncogenic Signaling Pathways in The Cancer Genome Atlas.* Cell, 2018. **173**(2): p. 321-337 e10.
41. Hoadley, K.A., et al., *Cell-of-Origin Patterns Dominate the Molecular Classification of 10,000 Tumors from 33 Types of Cancer.* Cell, 2018. **173**(2): p. 291-304 e6.
42. Ciriello, G., et al., *Comprehensive Molecular Portraits of Invasive Lobular Breast Cancer.* Cell, 2015. **163**(2): p. 506-19.
43. Lefebvre, C., et al., *Mutational Profile of Metastatic Breast Cancers: A Retrospective Analysis.* PLoS Med, 2016. **13**(12): p. e1002201.
44. Grasso, C.S., et al., *The mutational landscape of lethal castration-resistant prostate cancer.*

Nature, 2012. **487**(7406): p. 239-43.
45. Abida, W., et al., *Genomic correlates of clinical outcome in advanced prostate cancer.* Proc Natl Acad Sci U S A, 2019. **116**(23): p. 11428-11436.
46. Zehir, A., et al., *Mutational landscape of metastatic cancer revealed from prospective clinical sequencing of 10,000 patients.* Nat Med, 2017. **23**(6): p. 703-713.
47. Robinson, D.R., et al., *Integrative clinical genomics of metastatic cancer.* Nature, 2017. **548**(7667): p. 297-303.
48. Robinson, D.R., et al., *Activating ESR1 mutations in hormone-resistant metastatic breast cancer.* Nat Genet, 2013. **45**(12): p. 1446-51.
49. Paoletti, C., et al., *Comprehensive Mutation and Copy Number Profiling in Archived Circulating Breast Cancer Tumor Cells Documents Heterogeneous Resistance Mechanisms.* Cancer Res, 2018. **78**(4): p. 1110-1122.
50. Watson, P.A., V.K. Arora, and C.L. Sawyers, *Emerging mechanisms of resistance to androgen receptor inhibitors in prostate cancer.* Nat Rev Cancer, 2015. **15**(12): p. 701-11.
51. Visakorpi, T., et al., *In vivo amplification of the androgen receptor gene and progression of human prostate cancer.* Nat Genet, 1995. **9**(4): p. 401-6.
52. Tan, J., et al., *Dehydroepiandrosterone activates mutant androgen receptors expressed in the androgen-dependent human prostate cancer xenograft CWR22 and LNCaP cells.* Mol Endocrinol, 1997. **11**(4): p. 450-9.
53. Balbas, M.D., et al., *Overcoming mutation-based resistance to antiandrogens with rational drug design.* Elife, 2013. **2**: p. e00499.
54. Massarweh, S., et al., *Tamoxifen resistance in breast tumors is driven by growth factor receptor signaling with repression of classic estrogen receptor genomic function.* Cancer Res, 2008. **68**(3): p. 826-33.
55. Osipo, C., et al., *Paradoxical action of fulvestrant in estradiol-induced regression of tamoxifen-stimulated breast cancer.* J Natl Cancer Inst, 2003. **95**(21): p. 1597-608.
56. Howell, A., et al., *Response after withdrawal of tamoxifen and progestogens in advanced breast cancer.* Ann Oncol, 1992. **3**(8): p. 611-7.
57. Lei, J.T., et al., *Functional Annotation of ESR1 Gene Fusions in Estrogen Receptor-Positive Breast Cancer.* Cell Rep, 2018. **24**(6): p. 1434-1444 e7.
58. Hartmaier, R.J., et al., *Recurrent hyperactive ESR1 fusion proteins in endocrine therapy-resistant breast cancer.* Ann Oncol, 2018. **29**(4): p. 872-880.
59. Veeraraghavan, J., et al., *Recurrent ESR1-CCDC170 rearrangements in an aggressive subset of oestrogen receptor-positive breast cancers.* Nat Commun, 2014. **5**: p. 4577.
60. Matissek, K.J., et al., *Expressed Gene Fusions as Frequent Drivers of Poor Outcomes in Hormone Receptor-Positive Breast Cancer.* Cancer Discov, 2018. **8**(3): p. 336-353.
61. Quigley, D.A., et al., *Genomic Hallmarks and Structural Variation in Metastatic Prostate Cancer.* Cell, 2018. **174**(3): p. 758-769 e9.
62. Viswanathan, S.R., et al., *Structural Alterations Driving Castration-Resistant Prostate Cancer Revealed by Linked-Read Genome Sequencing.* Cell, 2018. **174**(2): p. 433-447 e19.
63. Beije, N., et al., *Estrogen receptor mutations and splice variants determined in liquid biopsies from metastatic breast cancer patients.* Mol Oncol, 2018. **12**(1): p. 48-57.
64. Beltran, H., et al., *Molecular characterization of neuroendocrine prostate cancer and identification of new drug targets.* Cancer Discov, 2011. **1**(6): p. 487-95.
65. Tan, H.L., et al., *Rb loss is characteristic of prostatic small cell neuroendocrine carcinoma.* Clin Cancer Res, 2014. **20**(4): p. 890-903.
66. Razavi, P., et al., *The Genomic Landscape of Endocrine-Resistant Advanced Breast Cancers.* Cancer Cell, 2018. **34**(3): p. 427-438 e6.
67. Tomlins, S.A., et al., *Distinct classes of chromosomal rearrangements create oncogenic ETS gene fusions in prostate cancer.* Nature, 2007. **448**(7153): p. 595-9.

68. Tomlins, S.A., et al., *Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer.* Science, 2005. **310**(5748): p. 644-8.
69. Tomlins, S.A., et al., *Role of the TMPRSS2-ERG gene fusion in prostate cancer.* Neoplasia, 2008. **10**(2): p. 177-88.
70. Bernardo, G.M. and R.A. Keri, *FOXA1: a transcription factor with parallel functions in development and cancer.* Biosci Rep, 2012. **32**(2): p. 113-30.
71. Parolia, A., et al., *Distinct structural classes of activating FOXA1 alterations in advanced prostate cancer.* Nature, 2019. **571**(7765): p. 413-418.
72. Haber, D.A. and V.E. Velculescu, *Blood-based analyses of cancer: circulating tumor cells and circulating tumor DNA.* Cancer Discov, 2014. **4**(6): p. 650-61.
73. Cristofanilli, M., et al., *Circulating tumor cells, disease progression, and survival in metastatic breast cancer.* N Engl J Med, 2004. **351**(8): p. 781-91.
74. Tomlins, S.A., et al., *Urine TMPRSS2:ERG Plus PCA3 for Individualized Prostate Cancer Risk Assessment.* Eur Urol, 2016. **70**(1): p. 45-53.
75. Alix-Panabieres, C. and K. Pantel, *Clinical Applications of Circulating Tumor Cells and Circulating Tumor DNA as Liquid Biopsy.* Cancer Discov, 2016. **6**(5): p. 479-91.
76. Paoletti, C., J. Smerage, and D.F. Hayes, *Circulating tumor cells as a marker of prognosis.* Princip Prac Oncol, 2012. **26**: p. 1-8.
77. Paoletti, C. and D.F. Hayes, *Circulating Tumor Cells.* Adv Exp Med Biol, 2016. **882**: p. 235-58.
78. Parsons, H.A., J.A. Beaver, and B.H. Park, *Circulating Plasma Tumor DNA.* Adv Exp Med Biol, 2016. **882**: p. 259-76.
79. Jankowitz, R.C., et al., *New Strategies in Metastatic Hormone Receptor-Positive Breast Cancer: Searching for Biomarkers to Tailor Endocrine and Other Targeted Therapies.* Clin Cancer Res, 2017. **23**(5): p. 1126-1131.
80. Cristofanilli, M., *Circulating tumor cells, disease progression, and survival in metastatic breast cancer.* N. Engl. J. Med., 2004. **351**: p. 781-791.
81. Smerage, J.B., et al., *Circulating Tumor Cells and Response to Chemotherapy in Metastatic Breast Cancer: SWOG S0500.* J Clin Oncol, 2014.
82. Jamal-Hanjani, M., et al., *Tracking the Evolution of Non-Small-Cell Lung Cancer.* N Engl J Med, 2017. **376**(22): p. 2109-2121.
83. Abbosh, C., et al., *Phylogenetic ctDNA analysis depicts early-stage lung cancer evolution.* Nature, 2017. **545**(7655): p. 446-451.
84. Roychowdhury, S., et al., *Personalized oncology through integrative high-throughput sequencing: a pilot study.* Sci Transl Med, 2011. **3**(111): p. 111ra121.
85. Nardone, A., et al., *The changing role of ER in endocrine resistance.* Breast, 2015. **24 Suppl 2**: p. S60-6.
86. Toy, W., et al., *ESR1 ligand-binding domain mutations in hormone-resistant breast cancer.* Nat Genet, 2013. **45**(12): p. 1439-45.
87. Li, S., et al., *Endocrine-therapy-resistant ESR1 variants revealed by genomic characterization of breast-cancer-derived xenografts.* Cell Rep, 2013. **4**(6): p. 1116-30.
88. Zhang, Q.X., et al., *An estrogen receptor mutant with strong hormone-independent activity from a metastatic breast cancer.* Cancer Res, 1997. **57**(7): p. 1244-9.
89. Merenbakh-Lamin, K., et al., *D538G mutation in estrogen receptor-alpha: A novel mechanism for acquired endocrine resistance in breast cancer.* Cancer Res, 2013. **73**(23): p. 6856-64.
90. Chu, D., et al., *ESR1 Mutations in Circulating Plasma Tumor DNA from Metastatic Breast Cancer Patients.* Clin Cancer Res, 2016. **22**(4): p. 993-9.
91. Shaw, J.A., et al., *Mutation Analysis of Cell-Free DNA and Single Circulating Tumor Cells in Metastatic Breast Cancer Patients with High Circulating Tumor Cell Counts.* Clin Cancer Res, 2017. **23**(1): p. 88-96.
92. Mu, Z., et al., *Detection and Characterization of Circulating Tumor Associated Cells in*

Metastatic Breast Cancer. Int J Mol Sci, 2016. **17**(10).
93. Carter, L., et al., *Molecular analysis of circulating tumor cells identifies distinct copy-number profiles in patients with chemosensitive and chemorefractory small-cell lung cancer.* Nat Med, 2017. **23**(1): p. 114-119.
94. Paolillo, C., et al., *Detection of Activating Estrogen Receptor Gene (ESR1) Mutations in Single Circulating Tumor Cells.* Clin Cancer Res, 2017.
95. Gulbahce, N., et al., *Quantitative Whole Genome Sequencing of Circulating Tumor Cells Enables Personalized Combination Therapy of Metastatic Cancer.* Cancer Res, 2017. **77**(16): p. 4530-4541.
96. Kidess-Sigal, E., et al., *Enumeration and targeted analysis of KRAS, BRAF and PIK3CA mutations in CTCs captured by a label-free platform: Comparison to ctDNA and tissue in metastatic colorectal cancer.* Oncotarget, 2016. **7**(51): p. 85349-85364.
97. Heitzer, E., et al., *Complex tumor genomes inferred from single circulating tumor cells by array-CGH and next-generation sequencing.* Cancer Res, 2013. **73**(10): p. 2965-75.
98. Paoletti, C., et al., *Development of circulating tumor cell-endocrine therapy index in patients with hormone receptor-positive breast cancer.* Clin Cancer Res, 2015. **21**(11): p. 2487-98.
99. Peeters, D.J., et al., *Semiautomated isolation and molecular characterisation of single or highly purified tumour cells from CellSearch enriched blood samples using dielectrophoretic cell sorting.* Br J Cancer, 2013. **108**(6): p. 1358-67.
100. Polzer, B., et al., *Molecular profiling of single circulating tumor cells with diagnostic intention.* EMBO Mol Med, 2014. **6**(11): p. 1371-86.
101. Cani, A.K., et al., *Next-Gen Sequencing Exposes Frequent MED12 Mutations and Actionable Therapeutic Targets in Phyllodes Tumors.* Mol Cancer Res, 2015. **13**(4): p. 613-9.
102. Hovelson, D.H., et al., *Development and validation of a scalable next-generation sequencing system for assessing relevant somatic variants in solid tumors.* Neoplasia, 2015. **17**(4): p. 385-99.
103. Warrick, J.I., et al., *Tumor evolution and progression in multifocal and paired non-invasive/invasive urothelial carcinoma.* Virchows Arch, 2015. **466**(3): p. 297-311.
104. Salvianti, F., et al., *Feasibility of a workflow for the molecular characterization of single cells by next generation sequencing.* Biomol Detect Quantif, 2015. **5**: p. 23-9.
105. McDaniel, A.S., et al., *Next-Generation Sequencing of Tubal Intraepithelial Carcinomas.* JAMA Oncol, 2015.
106. Grasso, C., et al., *Assessing copy number alterations in targeted, amplicon-based next-generation sequencing data.* J Mol Diagn, 2015. **17**(1): p. 53-63.
107. Li, H., et al., *The Sequence Alignment/Map format and SAMtools.* Bioinformatics, 2009. **25**(16): p. 2078-9.
108. Koboldt, D.C., et al., *VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing.* Genome Res, 2012. **22**(3): p. 568-76.
109. Wang, K., M. Li, and H. Hakonarson, *ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data.* Nucleic Acids Res, 2010. **38**(16): p. e164.
110. Lonigro, R.J., et al., *Detection of somatic copy number alterations in cancer using targeted exome capture sequencing.* Neoplasia, 2011. **13**(11): p. 1019-25.
111. Beaver, J.A., et al., *Detection of cancer DNA in plasma of patients with early-stage breast cancer.* Clin Cancer Res, 2014. **20**(10): p. 2643-50.
112. Johnson, M.D., et al., *Pharmacological characterization of 4-hydroxy-N-desmethyl tamoxifen, a novel active metabolite of tamoxifen.* Breast Cancer Res Treat, 2004. **85**(2): p. 151-9.
113. Conley, B.A. and J.H. Doroshow, *Molecular analysis for therapy choice: NCI MATCH.* Semin Oncol, 2014. **41**(3): p. 297-9.
114. Hovelson, D.H., et al., *Rapid, ultra low coverage copy number profiling of cell-free DNA as a precision oncology screening strategy.* Oncotarget, 2017. **8**(52): p. 89848-89866.
115. Gorges, T.M., et al., *Accession of Tumor Heterogeneity by Multiplex Transcriptome Profiling of*

Single Circulating Tumor Cells. Clin Chem, 2016. **62**(11): p. 1504-1515.
116. Schiavon, G., et al., *Analysis of ESR1 mutation in circulating tumor DNA demonstrates evolution during therapy for metastatic breast cancer.* Sci Transl Med, 2015. **7**(313): p. 313ra182.
117. Fribbens, C., et al., *Plasma ESR1 Mutations and the Treatment of Estrogen Receptor-Positive Advanced Breast Cancer.* J Clin Oncol, 2016. **34**(25): p. 2961-8.
118. Ellsworth, D.L., et al., *Single-cell sequencing and tumorigenesis: improved understanding of tumor evolution and metastasis.* Clin Transl Med, 2017. **6**(1): p. 15.
119. Pestrin, M., et al., *Heterogeneity of PIK3CA mutational status at the single cell level in circulating tumor cells from metastatic breast cancer patients.* Mol Oncol, 2015. **9**(4): p. 749-57.
120. Paolillo, C., et al., *Detection of Activating Estrogen Receptor Gene (ESR1) Mutations in Single Circulating Tumor Cells.* Clinical Cancer Research, 2017. **Published onlineFirst on July 5, 2017**.
121. Paoletti, C., et al., *Heterogeneous estrogen receptor expression in circulating tumor cells suggests diverse mechanisms of fulvestrant resistance.* Mol Oncol, 2016.
122. Babayan, A., et al., *Heterogeneity of estrogen receptor expression in circulating tumor cells from metastatic breast cancer patients.* PLoS One, 2013. **8**(9): p. e75038.
123. Riethdorf, S., et al., *Detection and HER2 expression of circulating tumor cells: prospective monitoring in breast cancer patients treated in the neoadjuvant GeparQuattro trial.* Clin Cancer Res, 2010. **16**(9): p. 2634-45.
124. Gasch, C., et al., *Frequent detection of PIK3CA mutations in single circulating tumor cells of patients suffering from HER2-negative metastatic breast cancer.* Mol Oncol, 2016. **10**(8): p. 1330-43.
125. Allard, W.J., et al., *Tumor cells circulate in the peripheral blood of all major carcinomas but not in healthy subjects or patients with nonmalignant diseases.* Clin Cancer Res, 2004. **10**(20): p. 6897-904.
126. Pantel, K., et al., *Circulating epithelial cells in patients with benign colon diseases.* Clin Chem, 2012. **58**(5): p. 936-40.
127. Hodgkinson, C.L., et al., *Tumorigenicity and genetic profiling of circulating tumor cells in small-cell lung cancer.* Nat Med, 2014. **20**(8): p. 897-903.
128. Wang, Z., et al., *Ex vivo expansion of circulating lung tumor cells based on one-step microfluidics-based immunomagnetic isolation.* Analyst, 2016. **141**(12): p. 3621-5.
129. Bidard, F.C., et al., *Clinical validity of circulating tumour cells in patients with metastatic breast cancer: a pooled analysis of individual patient data.* Lancet Oncol, 2014. **15**(4): p. 406-14.
130. Jeselsohn, R., et al., *ESR1 mutations-a mechanism for acquired endocrine resistance in breast cancer.* Nat Rev Clin Oncol, 2015.
131. Montano, M.M., et al., *The carboxy-terminal F domain of the human estrogen receptor: role in the transcriptional activity of the receptor and the effectiveness of antiestrogens as estrogen antagonists.* Mol Endocrinol, 1995. **9**(7): p. 814-25.
132. Gundem, G., et al., *The evolutionary history of lethal metastatic prostate cancer.* Nature, 2015. **520**(7547): p. 353-7.
133. Babayan, A., et al., *Comparative study of whole genome amplification and next generation sequencing performance of single cancer cells.* Oncotarget, 2017. **8**(34): p. 56066-56080.
134. Spoerke, J.M., et al., *Heterogeneity and clinical significance of ESR1 mutations in ER-positive metastatic breast cancer patients receiving fulvestrant.* Nat Commun, 2016. **7**: p. 11579.
135. Wasson, J.H., et al., *A structured literature review of treatment for localized prostate cancer. Prostate Disease Patient Outcome Research Team.* Arch Fam Med, 1993. **2**(5): p. 487-93.
136. Stamey, T.A., et al., *Prostate-specific antigen as a serum marker for adenocarcinoma of the prostate.* N Engl J Med, 1987. **317**(15): p. 909-16.
137. Siddiqui, M.M., et al., *Comparison of MR/ultrasound fusion-guided biopsy with ultrasound-guided biopsy for the diagnosis of prostate cancer.* JAMA, 2015. **313**(4): p. 390-7.

138. Filson, C.P., et al., *Prostate cancer detection with magnetic resonance-ultrasound fusion biopsy: The role of systematic and targeted biopsies.* Cancer, 2016. **122**(6): p. 884-92.
139. Salami, S.S., et al., *Biologic Significance of Magnetic Resonance Imaging Invisibility in Localized Prostate Cancer.* Jco Precision Oncology, 2019. **3**: p. 1-12.
140. Johnson, D.C., et al., *Detection of Individual Prostate Cancer Foci via Multiparametric Magnetic Resonance Imaging.* Eur Urol, 2019. **75**(5): p. 712-720.
141. Raaijmakers, R., et al., *Complication rates and risk factors of 5802 transrectal ultrasound-guided sextant biopsies of the prostate within a population-based screening program.* Urology, 2002. **60**(5): p. 826-30.
142. Sieweke, M.H. and M.J. Bissell, *The tumor-promoting effect of wounding: a possible role for TGF-beta-induced stromal alterations.* Crit Rev Oncog, 1994. **5**(2-3): p. 297-311.
143. Stuelten, C.H., et al., *Acute wounds accelerate tumorigenesis by a T cell-dependent mechanism.* Cancer Res, 2008. **68**(18): p. 7278-82.
144. Loeb, S., et al., *Overdiagnosis and overtreatment of prostate cancer.* Eur Urol, 2014. **65**(6): p. 1046-55.
145. Thompson, I.M., et al., *Assessing prostate cancer risk: results from the Prostate Cancer Prevention Trial.* J Natl Cancer Inst, 2006. **98**(8): p. 529-34.
146. Catalona, W.J., et al., *A multicenter study of [-2]pro-prostate specific antigen combined with prostate specific antigen and free prostate specific antigen for prostate cancer detection in the 2.0 to 10.0 ng/ml prostate specific antigen range.* J Urol, 2011. **185**(5): p. 1650-5.
147. Vickers, A., et al., *Reducing unnecessary biopsy during prostate cancer screening using a four-kallikrein panel: an independent replication.* J Clin Oncol, 2010. **28**(15): p. 2493-8.
148. Fenton, J.J., et al., *Prostate-Specific Antigen-Based Screening for Prostate Cancer: Evidence Report and Systematic Review for the US Preventive Services Task Force.* JAMA, 2018. **319**(18): p. 1914-1931.
149. Eggener, S.E., et al., *A multi-institutional evaluation of active surveillance for low risk prostate cancer.* J Urol, 2013. **189**(1 Suppl): p. S19-25; discussion S25.
150. Groskopf, J., et al., *APTIMA PCA3 molecular urine test: development of a method to aid in the diagnosis of prostate cancer.* Clin Chem, 2006. **52**(6): p. 1089-95.
151. Mehra, R., et al., *Heterogeneity of TMPRSS2 gene rearrangements in multifocal prostate adenocarcinoma: molecular evidence for an independent group of diseases.* Cancer Res, 2007. **67**(17): p. 7991-5.
152. Boutros, P.C., et al., *Spatial genomic heterogeneity within localized, multifocal prostate cancer.* Nat Genet, 2015. **47**(7): p. 736-45.
153. Palapattu, G.S., et al., *Molecular Profiling to Determine Clonality of Serial Magnetic Resonance Imaging/Ultrasound Fusion Biopsies from Men on Active Surveillance for Low-Risk Prostate Cancer.* Clin Cancer Res, 2017. **23**(4): p. 985-991.
154. Van Neste, L., et al., *Detection of High-grade Prostate Cancer Using a Urinary Molecular Biomarker-Based Risk Score.* Eur Urol, 2016. **70**(5): p. 740-748.
155. McKiernan, J., et al., *A Novel Urine Exosome Gene Expression Assay to Predict High-grade Prostate Cancer at Initial Biopsy.* JAMA Oncol, 2016. **2**(7): p. 882-9.
156. Roychowdhury, S. and A.M. Chinnaiyan, *Translating cancer genomes and transcriptomes for precision oncology.* CA Cancer J Clin, 2016. **66**(1): p. 75-88.
157. Bostrom, P.J., et al., *Genomic Predictors of Outcome in Prostate Cancer.* Eur Urol, 2015. **68**(6): p. 1033-44.
158. Salami, S.S., et al., *Transcriptomic heterogeneity in multifocal prostate cancer.* JCI Insight, 2018. **3**(21).
159. Prensner, J.R., et al., *The long noncoding RNA SChLAP1 promotes aggressive prostate cancer and antagonizes the SWI/SNF complex.* Nat Genet, 2013. **45**(11): p. 1392-8.
160. Zhang, Y., et al., *Analysis of the androgen receptor-regulated lncRNA landscape identifies a role*

for ARLNC1 in prostate cancer progression. Nat Genet, 2018. **50**(6): p. 814-824.
161. Iyer, M.K., et al., *The landscape of long noncoding RNAs in the human transcriptome.* Nat Genet, 2015. **47**(3): p. 199-208.
162. Cuzick, J., et al., *Prognostic value of an RNA expression signature derived from cell cycle proliferation genes in patients with prostate cancer: a retrospective study.* Lancet Oncol, 2011. **12**(3): p. 245-55.
163. Erho, N., et al., *Discovery and validation of a prostate cancer genomic classifier that predicts early metastasis following radical prostatectomy.* PLoS One, 2013. **8**(6): p. e66855.
164. Knezevic, D., et al., *Analytical validation of the Oncotype DX prostate cancer assay - a clinical RT-PCR assay optimized for prostate needle biopsies.* BMC Genomics, 2013. **14**: p. 690.
165. Klein, E.A., et al., *A 17-gene assay to predict prostate cancer aggressiveness in the context of Gleason grade heterogeneity, tumor multifocality, and biopsy undersampling.* Eur Urol, 2014. **66**(3): p. 550-60.
166. Hovelson, D.H., et al., *Targeted DNA and RNA Sequencing of Paired Urothelial and Squamous Bladder Cancers Reveals Discordant Genomic and Transcriptomic Events and Unique Therapeutic Implications.* Eur Urol, 2018. **74**(6): p. 741-753.
167. Cani, A.K., et al., *Next-Gen Sequencing Exposes Frequent MED12 Mutations and Actionable Therapeutic Targets in Phyllodes Tumors.* Molecular Cancer Research, 2015. **13**(4): p. 613-619.
168. Cani, A.K., et al., *Comprehensive genomic profiling of orbital and ocular adnexal lymphomas identifies frequent alterations in MYD88 and chromatin modifiers: new routes to targeted therapies.* Mod Pathol, 2016. **29**(7): p. 685-97.
169. Cani, A.K., et al., *Next generation sequencing of vitreoretinal lymphomas from small-volume intraocular liquid biopsies: new routes to targeted therapies.* Oncotarget, 2017. **8**(5): p. 7989-7998.
170. Storebjerg, T.M., et al., *Prevalence of the HOXB13 G84E mutation in Danish men undergoing radical prostatectomy and its correlations with prostate cancer risk and aggressiveness.* BJU Int, 2016. **118**(4): p. 646-53.
171. Kuhn, M., et al., *Caret: Classification and Regression Training. R package version 6.0-80.* https://CRAN.R-project.org/package=caret, 2018.
172. Genuer, R., J.M. Poggi, and C. Tuleau-Malot, *VSURF: An R Package for Variable Selection Using Random Forests.* The R Journal, R Foundation for Statistical Computing,, 2015. **7**(2): p. 19-33.
173. Hendriks, R.J., et al., *Comparative analysis of prostate cancer specific biomarkers PCA3 and ERG in whole urine, urinary sediments and exosomes.* Clin Chem Lab Med, 2016. **54**(3): p. 483-92.
174. Auprich, M., et al., *Contemporary role of prostate cancer antigen 3 in the management of prostate cancer.* Eur Urol, 2011. **60**(5): p. 1045-54.
175. Boormans, J.L., et al., *Identification of TDRD1 as a direct target gene of ERG in primary prostate cancer.* Int J Cancer, 2013. **133**(2): p. 335-45.
176. Gronberg, H., et al., *Prostate cancer screening in men aged 50-69 years (STHLM3): a prospective population-based diagnostic study.* Lancet Oncol, 2015. **16**(16): p. 1667-76.
177. Ewing, C.M., et al., *Germline mutations in HOXB13 and prostate-cancer risk.* N Engl J Med, 2012. **366**(2): p. 141-9.
178. Smith, S.C., et al., *HOXB13 G84E-related familial prostate cancers: a clinical, histologic, and molecular survey.* Am J Surg Pathol, 2014. **38**(5): p. 615-26.
179. Niknafs, Y.S., et al., *MiPanda: A Resource for Analyzing and Visualizing Next-Generation Sequencing Transcriptomics Data.* Neoplasia, 2018. **20**(11): p. 1144-1149.
180. Zapparoli, G.V., et al., *Quantitative threefold allele-specific PCR (QuanTAS-PCR) for highly sensitive JAK2 V617F mutant allele detection.* BMC Cancer, 2013. **13**: p. 206.
181. Connell, S.P., et al., *A Four-Group Urine Risk Classifier for Predicting Outcome in Prostate*

Cancer Patients. BJU Int, 2019.
182. Markopoulos, C., et al., *Multigene assays in early breast cancer: Insights from recent phase 3 studies.* Eur J Surg Oncol, 2019.
183. Kohaar, I., G. Petrovics, and S. Srivastava, *A Rich Array of Prostate Cancer Molecular Biomarkers: Opportunities and Challenges.* Int J Mol Sci, 2019. **20**(8).
184. Antonarakis, E.S., et al., *AR-V7 and resistance to enzalutamide and abiraterone in prostate cancer.* N Engl J Med, 2014. **371**(11): p. 1028-38.
185. Turner, N.C., et al., *Palbociclib in Hormone-Receptor-Positive Advanced Breast Cancer.* N Engl J Med, 2015. **373**(3): p. 209-19.
186. Cameron, D., et al., *11 years' follow-up of trastuzumab after adjuvant chemotherapy in HER2-positive early breast cancer: final analysis of the HERceptin Adjuvant (HERA) trial.* Lancet, 2017. **389**(10075): p. 1195-1205.
187. Andre, F., et al., *Alpelisib for PIK3CA-Mutated, Hormone Receptor-Positive Advanced Breast Cancer.* N Engl J Med, 2019. **380**(20): p. 1929-1940.
188. Robson, M., et al., *Olaparib for Metastatic Breast Cancer in Patients with a Germline BRCA Mutation.* N Engl J Med, 2017. **377**(6): p. 523-533.
189. Ploussard, G. and A. de la Taille, *Urine biomarkers in prostate cancer.* Nat Rev Urol, 2010. **7**(2): p. 101-9.
190. Ramakrishnan Geethakumari, P., et al., *PARP Inhibitors in Prostate Cancer.* Curr Treat Options Oncol, 2017. **18**(6): p. 37.
191. Gerlinger, M., et al., *Intratumor heterogeneity and branched evolution revealed by multiregion sequencing.* N Engl J Med, 2012. **366**(10): p. 883-892.
192. Mittendorf, E.A., et al., *PD-L1 expression in triple-negative breast cancer.* Cancer Immunol Res, 2014. **2**(4): p. 361-70.
193. Bielski, C.M., et al., *Widespread Selection for Oncogenic Mutant Allele Imbalance in Cancer.* Cancer Cell, 2018. **34**(5): p. 852-862 e4.
194. Tanenbaum, D.M., Wang, Y., Sigler, P.B., *ESTROGEN RECEPTOR ALPHA LIGAND-BINDING DOMAIN COMPLEXED TO ESTRADIOL*. 1998: Protein Data Bank.
195. Chan, T.A., et al., *Development of tumor mutation burden as an immunotherapy biomarker: utility for the oncology clinic.* Ann Oncol, 2019. **30**(1): p. 44-56.
196. Ho, A.Y., et al., *A phase 2 clinical trialassessing theefficacy and safety of pembrolizumab and radiotherapy in patients with metastatic triple-negative breast cancer.* Cancer, 2019.
197. Zaretsky, J.M., et al., *Mutations Associated with Acquired Resistance to PD-1 Blockade in Melanoma.* N Engl J Med, 2016. **375**(9): p. 819-29.
198. Yoon, H.J., et al., *Sensitive capture of circulating tumour cells by functionalized graphene oxide nanosheets.* Nat Nanotechnol, 2013. **8**(10): p. 735-41.
199. Kim, T.H., et al., *A temporary indwelling intravascular aphaeretic system for in vivo enrichment of circulating tumor cells.* Nat Commun, 2019. **10**(1): p. 1478.
200. Lambros, M.B., et al., *Single-Cell Analyses of Prostate Cancer Liquid Biopsies Acquired by Apheresis.* Clin Cancer Res, 2018. **24**(22): p. 5635-5644.
201. Hazar-Rethinam, M., et al., *Convergent Therapeutic Strategies to Overcome the Heterogeneity of Acquired Resistance in BRAF(V600E) Colorectal Cancer.* Cancer Discov, 2018. **8**(4): p. 417-427.
202. Adalsteinsson, V.A., et al., *Scalable whole-exome sequencing of cell-free DNA reveals high concordance with metastatic tumors.* Nat Commun, 2017. **8**(1): p. 1324.
203. Dinan, M.A., et al., *Proposal for Value-Based, Tiered Reimbursement for Tumor Biomarker Tests to Promote Innovation and Evidence Generation.* JCO Precision Oncology, 2019(3): p. 1-10.
204. Margo, C.E. and Z.D. Mulla, *Malignant tumors of the orbit. Analysis of the Florida Cancer Registry.* Ophthalmology, 1998. **105**(1): p. 185-90.
205. Spraul, C.W. and H.E. Grossniklaus, *Analysis of 24,444 surgical specimens accessioned over 55 years in an ophthalmic pathology laboratory.* Int Ophthalmol, 1997. **21**(5): p. 283-304.

206. Fung, C.Y., et al., *Ocular adnexal lymphoma: clinical behavior of distinct World Health Organization classification subtypes.* Int J Radiat Oncol Biol Phys, 2003. **57**(5): p. 1382-91.
207. White, W.L., et al., *Ocular adnexal lymphoma. A clinicopathologic study with identification of lymphomas of mucosa-associated lymphoid tissue type.* Ophthalmology, 1995. **102**(12): p. 1994-2006.
208. Aronow, M.E., et al., *Ocular adnexal lymphoma: assessment of a tumor-node-metastasis staging system.* Ophthalmology, 2013. **120**(9): p. 1915-9.
209. Morin, R.D., et al., *Mutational and structural analysis of diffuse large B-cell lymphoma using whole-genome sequencing.* Blood, 2013. **122**(7): p. 1256-65.
210. Morin, R.D., et al., *Frequent mutation of histone-modifying genes in non-Hodgkin lymphoma.* Nature, 2011. **476**(7360): p. 298-303.
211. Fend, F., A.J. Ferreri, and S.E. Coupland, *How we diagnose and treat vitreoretinal lymphoma.* Br J Haematol, 2016. **173**(5): p. 680-92.
212. Mondello, P., M. Mian, and F. Bertoni, *Primary central nervous system lymphoma: Novel precision therapies.* Crit Rev Oncol Hematol, 2019. **141**: p. 139-145.
213. Stafford, S.L., et al., *Orbital lymphoma: radiotherapy outcome and complications.* Radiother Oncol, 2001. **59**(2): p. 139-44.
214. Moslehi, R., et al., *Rapidly increasing incidence of ocular non-hodgkin lymphoma.* J Natl Cancer Inst, 2006. **98**(13): p. 936-9.
215. Kim, W.S., et al., *Genome-wide array-based comparative genomic hybridization of ocular marginal zone B cell lymphoma: comparison with pulmonary and nodal marginal zone B cell lymphoma.* Genes Chromosomes Cancer, 2007. **46**(8): p. 776-83.
216. Matteucci, C., et al., *Typical genomic imbalances in primary MALT lymphoma of the orbit.* J Pathol, 2003. **200**(5): p. 656-60.
217. Ruiz, A., et al., *Extranodal marginal zone B-cell lymphomas of the ocular adnexa: multiparameter analysis of 34 cases including interphase molecular cytogenetics and PCR for Chlamydia psittaci.* Am J Surg Pathol, 2007. **31**(5): p. 792-802.
218. Schiby, G., et al., *Orbital marginal zone lymphomas: an immunohistochemical, polymerase chain reaction, and fluorescence in situ hybridization study.* Hum Pathol, 2007. **38**(3): p. 435-42.
219. Tanimoto, K., et al., *Fluorescence in situ hybridization (FISH) analysis of primary ocular adnexal MALT lymphoma.* BMC Cancer, 2006. **6**: p. 249.
220. McDaniel, A.S., et al., *HRAS mutations are frequent in inverted urothelial neoplasms.* Hum Pathol, 2014. **45**(9): p. 1957-65.
221. Xu, X.L., et al., *Retinoblastoma has properties of a cone precursor tumor and depends upon cone-specific MDM2 signaling.* Cell, 2009. **137**(6): p. 1018-31.
222. Xu, X.L., et al., *Rb suppresses human cone-precursor-derived retinoblastoma tumours.* Nature, 2014. **514**(7522): p. 385-8.
223. Zhang, J., et al., *A novel retinoblastoma therapy from genomic and epigenetic analyses.* Nature, 2012. **481**(7381): p. 329-34.
224. Wang, J.Q., et al., *Toll-Like Receptors and Cancer: MYD88 Mutation and Inflammation.* Front Immunol, 2014. **5**: p. 367.
225. Bodor, C., et al., *EZH2 mutations are frequent and represent an early event in follicular lymphoma.* Blood, 2013. **122**(18): p. 3165-8.
226. Bodor, C., et al., *EZH2 Y641 mutations in follicular lymphoma.* Leukemia, 2011. **25**(4): p. 726-9.
227. Morin, R.D., et al., *Somatic mutations altering EZH2 (Tyr641) in follicular and diffuse large B-cell lymphomas of germinal-center origin.* Nat Genet, 2010. **42**(2): p. 181-5.
228. Treon, S.P., et al., *MYD88 L265P somatic mutation in Waldenstrom's macroglobulinemia.* N Engl J Med, 2012. **367**(9): p. 826-33.
229. Ngo, V.N., et al., *Oncogenically active MYD88 mutations in human lymphoma.* Nature, 2011. **470**(7332): p. 115-9.

230. Schmidt, J., et al., *MYD88 L265P and CXCR4 mutations in lymphoplasmacytic lymphoma identify cases with high disease activity.* Br J Haematol, 2015. **169**(6): p. 795-803.
231. Gachard, N., et al., *IGHV gene features and MYD88 L265P mutation separate the three marginal zone lymphoma entities and Waldenstrom macroglobulinemia/lymphoplasmacytic lymphomas.* Leukemia, 2013. **27**(1): p. 183-9.
232. Parry, M., et al., *Genetics and Prognostication in Splenic Marginal Zone Lymphoma: Revelations from Deep Sequencing.* Clin Cancer Res, 2015.
233. Martinez-Lopez, A., et al., *MYD88 (L265P) somatic mutation in marginal zone B-cell lymphoma.* Am J Surg Pathol, 2015. **39**(5): p. 644-51.
234. Nakamura, T., et al., *Recurrent mutations of CD79B and MYD88 are the hallmark of primary central nervous system lymphomas.* Neuropathol Appl Neurobiol, 2015.
235. Bonzheim, I., et al., *High frequency of MYD88 mutations in vitreoretinal B-cell lymphoma: a valuable tool to improve diagnostic yield of vitreous aspirates.* Blood, 2015. **126**(1): p. 76-9.
236. Pulido, J.S., et al., *MyD-88 L265P mutations are present in some cases of vitreoretinal lymphoma.* Retina, 2015. **35**(4): p. 624-7.
237. Li, Z.M., et al., *MYD88 somatic mutations in MALT lymphomas.* Br J Haematol, 2012. **158**(5): p. 662-4.
238. Zhu, D., et al., *Molecular and genomic aberrations in Chlamydophila psittaci negative ocular adnexal marginal zone lymphomas.* Am J Hematol, 2013. **88**(9): p. 730-5.
239. Liu, F., et al., *Mutation analysis of NF-kappaB signal pathway-related genes in ocular MALT lymphoma.* Int J Clin Exp Pathol, 2012. **5**(5): p. 436-41.
240. Martinez-Climent, J.A., *The origin and targeting of mucosa-associated lymphoid tissue lymphomas.* Curr Opin Hematol, 2014. **21**(4): p. 309-19.
241. Lee, J.L., et al., *Extranodal marginal zone B-cell lymphomas of mucosa-associated lymphoid tissue-type of the orbit and ocular adnexa.* Ann Hematol, 2005. **84**(1): p. 13-8.
242. Rao, R.C. and Y. Dou, *Hijacked in cancer: the KMT2 (MLL) family of methyltransferases.* Nat Rev Cancer, 2015. **15**(6): p. 334-46.
243. Kandoth, C., et al., *Mutational landscape and significance across 12 major cancer types.* Nature, 2013. **502**(7471): p. 333-9.
244. Pasqualucci, L., et al., *Analysis of the coding genome of diffuse large B-cell lymphoma.* Nat Genet, 2011. **43**(9): p. 830-7.
245. Martinez-Glez, V. and P. Lapunzina, *Sotos syndrome is associated with leukemia/lymphoma.* Am J Med Genet A, 2007. **143A**(11): p. 1244-5.
246. Liggins, A.P., et al., *Serologic detection of diffuse large B-cell lymphoma-associated antigens.* Int J Cancer, 2004. **110**(4): p. 563-9.
247. Takeuchi, I., et al., *The potential of copy number gains and losses, detected by array-based comparative genomic hybridization, for computational differential diagnosis of B-cell lymphomas and genetic regions involved in lymphomagenesis.* Haematologica, 2009. **94**(1): p. 61-9.
248. Smedby, K.E., et al., *GWAS of follicular lymphoma reveals allelic heterogeneity at 6p21.32 and suggests shared genetic susceptibility with diffuse large B-cell lymphoma.* PLoS Genet, 2011. **7**(4): p. e1001378.
249. Li, H., et al., *Mutations in linker histone genes HIST1H1 B, C, D, and E; OCT2 (POU2F2); IRF8; and ARID1A underlying the pathogenesis of follicular lymphoma.* Blood, 2014. **123**(10): p. 1487-98.
250. Love, C., et al., *The genetic landscape of mutations in Burkitt lymphoma.* Nat Genet, 2012. **44**(12): p. 1321-5.
251. Zhang, J., et al., *Genetic heterogeneity of diffuse large B-cell lymphoma.* Proc Natl Acad Sci U S A, 2013. **110**(4): p. 1398-403.
252. Li, H., et al., *Mutations in linker histone genes HIST1H1 B, C, D and E, OCT2 (POU2F2), IRF8 and ARID1A underlying the pathogenesis of follicular lymphoma.* Blood, 2014.

253. McCabe, M.T., et al., *EZH2 inhibition as a therapeutic strategy for lymphoma with EZH2-activating mutations.* Nature, 2012. **492**(7427): p. 108-12.
254. Knutson, S.K., et al., *A selective inhibitor of EZH2 blocks H3K27 methylation and kills mutant lymphoma cells.* Nat Chem Biol, 2012. **8**(11): p. 890-6.
255. Bradley, W.D., et al., *EZH2 inhibitor efficacy in non-Hodgkin's lymphoma does not require suppression of H3K27 monomethylation.* Chem Biol, 2014. **21**(11): p. 1463-75.
256. Bitler, B.G., et al., *Synthetic lethality by targeting EZH2 methyltransferase activity in ARID1A-mutated cancers.* Nat Med, 2015. **21**(3): p. 231-8.
257. Chan, C.C., et al., *Primary vitreoretinal lymphoma: a report from an International Primary Central Nervous System Lymphoma Collaborative Group symposium.* Oncologist, 2011. **16**(11): p. 1589-99.
258. Davis, J.L., *Intraocular lymphoma: a clinical perspective.* Eye (Lond), 2013. **27**(2): p. 153-62.
259. McDaniel, A.S., et al., *Genomic Profiling of Penile Squamous Cell Carcinoma Reveals New Opportunities for Targeted Therapy.* Cancer Res, 2015. **75**(24): p. 5219-27.
260. McDaniel, A.S., et al., *Next-Generation Sequencing of Tubal Intraepithelial Carcinomas.* JAMA Oncol, 2015. **1**(8): p. 1128-32.
261. Wang, Y., et al., *Molecular biomarkers for the diagnosis of primary vitreoretinal lymphoma.* Int J Mol Sci, 2011. **12**(9): p. 5684-97.
262. Rhyasen, G.W. and D.T. Starczynowski, *IRAK signalling in cancer.* Br J Cancer, 2015. **112**(2): p. 232-7.
263. Treon, S.P., L. Xu, and Z. Hunter, *MYD88 Mutations and Response to Ibrutinib in Waldenstrom's Macroglobulinemia.* N Engl J Med, 2015. **373**(6): p. 584-6.
264. Raja, H., et al., *Prevalence of Myd88 L265p Mutation in Histologically Proven, Diffuse Large B-Cell Vitreoretinal Lymphoma.* Retina, 2016. **36**(3): p. 624-8.
265. Carpten, J.D., et al., *A transforming mutation in the pleckstrin homology domain of AKT1 in cancer.* Nature, 2007. **448**(7152): p. 439-44.
266. Cui, W., et al., *Frequent copy number variations of PI3K/AKT pathway and aberrant protein expressions of PI3K subunits are associated with inferior survival in diffuse large B cell lymphoma.* J Transl Med, 2014. **12**: p. 10.