GENOMIC LANDSCAPE PROFILING OF BREAST CANCER IN PTEN HAMARTOMA TUMOR SYNDROME

by

TAKAE MIZUKAMI

Submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Department of Molecular Medicine

CASE WESTERN RESERVE UNIVERISTY

January 2024

CASE WESTERN RESERVE UNIVERSITY

SCHOOL OF GRADUATE STUDIES

We hereby approve the dissertation of

Takae Mizukami

Candidate for the degree of Doctor of Philosophy.

Committee Chair:

Jonathan Smith, PhD

Committee Member:

Ruth Keri, PhD

Committee Member:

Thomas LaFramboise, PhD

Dissertation Mentor:

Charis Eng, MD, PhD

Date of Defense **December 6th, 2023**

*We also certify that written approval has been obtained

for any proprietary material contained therein.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Genomic Landscape Profiling of Breast Cancer

in *PTEN* Hamartoma Tumor Syndrome (PHTS)

Abstract

by

TAKAE MIZUKAMI

Germline *PTEN* variants (PTEN hamartoma tumor syndrome, PHTS) confer up to 85-91% lifetime risk of female breast cancer (BC). BCs arising in PHTS are clinically distinct from sporadic BCs, including younger age of onset, multifocality, and an increased risk of second primary BCs. Yet, there is no previous investigation into the underlining genomic landscape of this entity. We sought to address the hypothesis that BCs arising in PHTS have a distinct genomic landscape compared to sporadic counterparts. We performed and analyzed exome sequencing from 44 breast tissues from women with germline *PTEN* variants who developed BC. The control cohort comprised of 497 sporadic BCs from The Cancer Genome Atlas (TCGA) dataset. We demonstrate that PHTS-derived BCs have a distinct somatic mutational landscape compared to the sporadic counterparts, namely second somatic hits in *PTEN*, distinct mutational signatures and increased genomic instability. The PHTS group had a significantly higher frequency of somatic *PTEN* variants compared to TCGA (22.7% versus 5.6%; odds ratio [OR] 4.93; 95% confidence interval [CI] 2.21 to 10.98; p<0.001), and a lower mutational frequency in *PIK3CA* (22.7% versus 33.4%; OR 0.59; 95% CI 0.28 to 1.22; p=0.15). Somatic variants

in *PTEN* and *PIK3CA* were mutually exclusive in PHTS (p=0.01), but not in TCGA. There appear to be two genomically distinct groups of PHTS BCs, depending on the pathogenicity of underlying germline *PTEN* variants.

We then examined copy number variations (CNV) and expression data to further characterize the somatic landscape of PHTS-derived BC. To examine CNV, we used the same exome sequencing data from 44 PHTS-derived BC cases described above and an expanded TCGA cohort with 558 women with sporadic BCs. We demonstrated that PHTS-derived BCs have several distinct CNV peaks compared to TCGA. Furthermore, RNA sequencing data revealed that PHTS-derived BCs have distinct inferred immunologic cell types, which points toward cancer immune evasion. Gene expression data also revealed previously described Tier-1 PHTS-derived BC, with pathogenic germline *PTEN* variants, appears to have vitamin E degradation as a key pathway in tumorigenesis. In conclusion, our findings have important implications for the personalized management of *PTEN*-related BCs, which are currently treated similarly to the sporadic BCs. Germline *PTEN* alterations are expected to be identified more frequently as an important underlying cause of BCs, especially in the context of more accessible genetic testing. More effective and targeted treatment strategies are warranted for this unique group of women affected by BCs.

CHAPTER I: BACKGROUND AND SIGNIFICANCE

Sporadic breast cancer - Epidemiology, staging, and survival

Breast cancer (BC) is the most common type of cancer among women worldwide. In 2020, 2.3 million new cases of BC and about 685,000 deaths from this disease have been reported globally[1]. In the United States alone, 287,850 new cases and 43,250 deaths are estimated to have occurred in the period of 2017-2019[2]. The lifetime risk of developing BC is approximately 12.9% among women in the United States, based on the Surveillance, Epidemiology, and End Results (SEER) 2017-2019 data[2]. BC is a heterogeneous disease with complex causes[3]. The majority of cases are sporadic, which do not occur in the setting of hereditary cancer syndromes. Environmental, lifestyle, or other unknown factors are usually attributable to sporadic cases[4].

The survival rate of BC varies, depending on the extent of the disease burden and affected organs at the time of initial diagnosis. The most commonly used clinical staging system is set by the American Joint Committee on Cancer (AJCC), which is an anatomical staging called the TNM staging system, based on primary tumor size (T stage), the number of axillary lymph nodes involved with malignancy (N stage), and the presence of distant metastases (M stage)[5]. Based on this system, clinicians describe the extent of disease from stage 0 (in situ carcinoma) to stage I and stage II (early and localized disease), stage III (locally advanced) and stage IV (distant metastasis). More recently, the staging system has been refined to include molecular markers as well as tumor grade (AJCC 8th

edition)[6]. The clinical staging is useful for many reasons but the information is most importantly, prognostic.

In general, the survival rate decreases as the clinical stage advances. When BC has not spread beyond the breast, thus localized disease (stage I and II), the 5-year survival is 99-100%. When the cancer spreads beyond the breast but is still limited to the locoregional lymph nodes (stage II and III), the 5-year survival rate is approximately 86%. Finally, when the cancer cells spread to distant organs beyond the breast as well as the locoregional lymph nodes (stage IV), the 5-year survival rate goes down to approximately 30%[2]. The disease course of each BC case is complex and heterogeneous, and other factors including histologic grade and clinical subtypes based on tumor markers affect survival rates[7;8].

Clinical subtypes of breast cancer

BC arises in the epithelial components of the ducts or lobules of the breast. When cancerous epithelial cells are confined within the duct (intraductal) or lobule (intralobular) without invading thorough the basement membrane, the cancer is non-invasive and called ductal carcinoma *in situ* (DCIS) and lobular carcinoma *in situ* (LCIS), respectively. BC is further classified into clinical subtypes based on tumor markers detected by immunohistochemistry (IHC) and fluorescence *in situ* hybridization (FISH)[9]. Three important tumor markers for breast cancer are estrogen receptor (ER), progesterone receptor (PR) and

human epidermal growth factor receptor 2 (HER2)[10]. Based on the IHC expression and FISH, it is important to recognize four BC subtypes in the clinical setting: ER+/PR+/HER2-, ER+/PR+/HER2+, ER-/PR-/HER2+, and ER-/PR-/HER2-.

Intrinsic subtype classification, on the other hand, is based on gene expression profiling, and groups BCs into five subtypes: luminal A, luminal B, HER2 over-expressed, basal-like and normal-like tumors[9;11]. Luminal A and B tend to correlate with ER+ BC, HER2 over-expressed with HER2+ BC, and basal subtype with triple negative breast cancer (TNBC)[9].

Current treatment approaches for sporadic breast cancer

Treatment strategies are determined based on multiple factors, including staging, tumor characteristics including tumor markers (ER, PR, HER2), tumor grade, proliferation rate, and each patient's functional status. The main modalities for breast cancer treatment include local management with surgery and radiation, and systemic therapies including hormonal therapy, chemotherapy, immune therapy, and targeted therapy. For early stage (stage I and II) and many cases of locally advanced breast cancers (stage III), treatment is given with curative intent in combination with surgery, radiation, and systemic therapies (either neoadjuvant or adjuvant). Systemic therapy options mainly depend on the tumor marker results, such as hormonal therapy for ER+ BC and HER2-directed therapy for HER2-positive BC.

When BC metastasizes to distant organs, such as bone, liver, lung, and/or brain, the disease is not considered curable, and the main focus will be to palliate symptoms and potentially prolong overall survival while keeping good quality of life. Depending on the sites of metastasis and degree of disease burden, a combination of modalities may be used including chemotherapy, radiation therapy and rarely surgery. Since many metastatic BC cases are not surgical, systemic therapy is mainly used when appropriate. In addition to conventional chemotherapy and hormonal therapy, a number of targeted therapies have become available and continue to evolve. Trastuzumab, a HER2 tagged therapy with monoclonal antibody, is the first systemic treatment which was developed to specifically treat BC with HER2 amplification, a genomic alteration giving cancer cells survival advantage leading to aggressive phenotype. This approach specifically blocks HER2 signaling and is thus highly effective for BC with HER2 amplification, which can be detected by immunohistochemistry (IHC) and confirmed with fluorescence in situ hybridization (FISH). Since the FDA approval of trastuzumab, a number of other anti-HER2 therapies, with subset conjugated to drug, have been developed and approved, including pertuzumab, margetuximab, lapatinib, neratinib, tucatinib, ado-trastuzumab emtansine (T-DM1), trastuzumab emtansine, and fam-trastuzumab deruxtecan-nxki.

For certain cases, targeted therapies are used such as mammalian target of rapamycin (mTOR) inhibitors, cyclin-dependent kinase (CDK) inhibitors, and phosphatidylinositol-4, 5-bisphosphate 3-kinase, catalytic subunit (PIK3CA) inhibitors. These therapeutic agents inhibit specific oncogenic pathways, thus

more targeted for certain type of breast cancers. The mTOR inhibitor is a protein kinase inhibitor of the mTOR protein, which is one of the key molecules in the phosphatidylinositol-3-kinase (PI3K) pathway[12]. The U.S. Food and Drug Administration (FDA) has approved everolimus in combination with exemestane, a hormonal therapy, for the treatment of postmenopausal women with hormone receptor (HR)-positive, HER2-negative BCs whose tumors failed an aromatase inhibitor (letrozole or anastrozole) in the metastatic setting[13;14]. Another important agent targeting the PI3K pathway is alpelisib, a PIK3CA inhibitor. *PIK3CA* is an oncogene, which drives many cases of sporadic BCs when somatic gain-of-function mutations are present. Alpelicib was FDA approved in combination with fulvestrant for postmenopausal women and men with HR-positive, HER2-negative, somatically *PIK3CA*-mutated breast cancer after failing an endocrine regimen in the metastatic setting[15]. Furthermore, alpelisib was also approved as a systemic therapy for adult and pediatric patients with severe PIK3CA-related overgrowth syndrome (PROS)[16].

The CDK family plays an important role in the cell cycle by phosphorylating retinoblastoma (RB) protein, thus promoting progression of cell division. By selectively targeting CDK 4 and 6, CDK 4/5 inhibitors inactivate these kinases and thus dephosphorylate RB, leading to cell cycle arrest[17]. FDA approved ribociclib[18], palbociclib[19], and abemaciclib[20] in combination with an aromatase inhibitor for women with HR-positive, HER2-negative advanced or metastatic BC. Abemaciclib with endocrine therapy (tamoxifen or an aromatase inhibitor) was further approved by the FDA for women with HR-positive, HER2-

negative, node-positive, early BC with high risk of recurrence and a Ki-67 score greater than 20% in the adjuvant setting[21].

More recently, approved targeted therapies include sacituzumab govitecan-hziy, an antibody-drug conjugate which targets TROP-2. This drug was first approved by the FDA for triple negative breast cancer (TNBC) in the metastatic setting after two or more prior systemic therapies[22]. This agent has also been approved for BC beyond TNBC, more specifically for unresectable locally advanced or metastatic, hormone receptor (HR)-positive, HER2-negative BC which failed endocrine therapy and at least two additional systemic therapies[23].

In January of 2023, the FDA approved elacestrant for ER-positive, HER2-negative, ESR-1 somatically mutated advanced or metastatic BC after failing one endocrine regimen[24]. Elacestrant is an oral selective estrogen receptor degrader and targets endogenous estrogens. The Phase III EMERALD trial stratified eligible patients by ESR-1 mutation status, which showed significant improvement in median progression free survival (PFS) in the ESR-1 mutation group[24]. Similarly to alpelisib, this is another example of targeted therapy which was FDA approved based on somatic mutational status.

Lastly, immunotherapy, which targets the PD-1/PD-L1 pathway has been used to treat selected cases of BC. Agents which are used in this setting include pembrolizumab for metastatic and early-stage, high-risk TNBC[25;26] and dostarlimab for advanced tumors of any type, including BC, with DNA mismatch repair deficiency (dMMR) [27], for which biomarkers such as PD-1/PD-L1 positivity

and dMMR are used as biomarkers indicating potential efficacy. Although immunotherapy has been only recently introduced as part of treatment strategies for BC, there are hopes for immunotherapy as one of the potentially important modalities since BC, similarly to other solid tumors, is now well-recognized as an immune-mediated or immune-failure disease. Before immunotherapy becomes a promising treatment strategy for BC in general, there remain many challenges including low response rate[28], resistance to immune therapy[29] and immune-related adverse effects[30]. Furthermore, heterogenic immunogenicity is an inherent challenge in BC, which adds another layer of complexity for immunotherapy to be effective[31].

Genomic landscape of sporadic breast cancer

In sporadic BCs, acquired alterations occur in the DNA of the epithelial component of the breast, which confer survival advantage to the cancer initiating cells, leading to carcinogenesis. This is purely somatic in nature, meaning the disease-causing genomic changes are only found in the cancer cells of the affected organ. Cancer-promoting genomic alterations either promote uncontrolled cell proliferation (oncogene) or inhibit suppression of tumor development (tumor suppressor gene).

The genomic landscape of sporadic BCs has been well-documented and characterized in the literature. The Cancer Genome Atlas (TCGA) Network reported a comprehensive molecular analysis of human breast tumors, which

included 825 BC samples to examine the somatic landscape of sporadic BCs[32]. This landmark study revealed that there are characteristic somatic mutational patterns based on clinical subtypes. In the luminal subtypes (in general, estrogen receptor positive [ER+] and human epidermal growth factor 2 negative [HER2-]), the most frequently mutated gene was *PIK3CA* (45-29%), followed by *TP53* (29-12%), *GATA3* (14-16%), and *MAP3K1* (5-13%). In the HER2-enriched subtype and basal-like subtypes, the most frequently mutated gene was *TP53* (72-80%), followed by *PIK3CA* (9-39%) and *MLL3* (5-7%).

Since *PIK3CA* mutations are the most common somatic mutation type found in the sporadic BC, alpelisib, an FDA-approved PIK3CA targeted therapy, has become one of the important treatment strategies for BC with somatic alterations in *PIK3CA* in the metastatic setting[33].

Hereditary breast cancer syndromes

The cause of each BC is complex and often multifactorial. Sporadic BCs tend to arise from environmental factors, such as exposure to carcinogens including tobacco use, or can be hormonally driven. Fundamentally, however, cancer is a genomic disease with molecular mechanisms leading to its development and progression initiating at the genomic level. Approximately 5-10% of all BCs are associated with hereditary cancer syndromes, which predispose affected women to breast cancer development at younger ages[4]. Hereditary BC arises in the setting of germline mutations in cancer susceptibility

genes, including *BRCA1*, *BRCA2*, *PTEN*, *TP53*, *STK11*, and *CDH1*[34]. Cancers arising in this context tend to be more aggressive, multifocal and occur at a younger age than sporadic counterparts.

For example, hereditary breast and ovarian cancer (HBOC) syndrome is a well-known clinical entity caused by germline pathogenic mutations in *BRCA1* and *BRCA2* genes. The BC risk by age 80 years has been reported to be 72% for *BRCA1* and 69% for *BRCA2* mutation carriers with BC incidences increasing rapidly until ages 30 to 40 years for *BRCA1* and until ages 40 to 50 years for *BRCA2* mutation carriers[35]. Women with HBOC are also at higher risk for contralateral BC and ovarian cancer, while men are at a higher risk for prostate cancer. Both men and women with HBOC have increased risk for pancreatic cancer[36].

Similarly, pathogenic, or likely-pathogenic germline mutations in *TP53*, *PTEN*, *STK11*, *CDH1* cause Li-Fraumeni syndrome, PTEN hamartoma tumor syndrome, Peutz-Jeghers syndrome and hereditary diffuse gastric cancer (HDGC) syndrome, respectively. These entities also predispose carriers to multiple types of cancers other than BC, warranting personalized screening and preventative strategies.

Assessment and management of hereditary cancer syndromes

Genetic evaluation is a crucial part of comprehensive cancer care for management and prevention of hereditary cancer. During cancer risk

assessment, the goal is to identify patients with a personal history of cancer who may be predisposed to additional primary cancers. Furthermore, it is important to identify individuals without a personal history of cancer but who may be at an increased risk of developing cancer based on their family history. Once we identify individuals with higher risk for certain cancers, cancer surveillance, risk-reduction and therapeutic strategies can be personalized and implemented.

An individual should be evaluated for hereditary BC syndromes based on the personal and/or family history of certain types of cancers. In general, clinicians should start to suspect hereditary BC syndromes when the BC diagnosis occurs at a younger age (≤ 50 years old), more than one primary BC is found in one individual, TNBC at any age, male BC, a known pathogenic or likely-pathogenic mutation in a BC susceptibility gene in the family, and three or more close relatives on the paternal or maternal side of the family are diagnosed with BC[36]. Once a hereditary cancer syndrome is suspected, the individual should proceed with genetic testing, guided by genetic counseling, to identify pathogenic or likely pathogenic mutations in BC susceptibility genes. Other cancer types, including ovarian cancer, pancreatic cancer, and prostate cancer, in the family can provide further information supporting the possibility that a hereditary BC syndrome may be present. Detailed testing criteria for hereditary BC syndromes are outlined in the National Comprehensive Cancer Network (NCCN) guidelines, which are updated regularly[36].

Since absolute risk of developing breast cancer in hereditary cancer syndromes is high, individuals carrying pathogenic or likely pathogenic mutations

in BC susceptibility genes are recommended to undergo increased cancer surveillance and prophylactic surgeries[36]. Information on germline mutation status is useful to choose targeted therapies in some cases of BC. Olaparib, a Poly (ADP-ribose) polymerase (PARP) inhibitor, has been approved by the FDA to treat metastatic BC with germline *BRCA1/2* mutation[37]. The indication was later expanded to be used in the adjuvant setting to treat HER2 negative, high-risk early BC with germline *BRCA1/2* mutations[38].

There are currently no germline mutation-directed BC therapies for individuals with pathogenic or likely pathogenic mutations in *PTEN*, *TP53*, *STK11*, or *CDH1*.

Evaluation of PTEN hamartoma tumor syndrome (PHTS)

Before the discovery of *PTEN* and its recognition as a cancer susceptibility gene, Cowden syndrome (CS) was first described in 1963 in a 20-years old female who initially presented with a right-sided breast lesion[39]. Other notable features included scrotal tongue, oral papillomatosis, multiple thyroid adenomas, and extensive fibrocystic disease of the breast. CS was then further characterized to include other phenotypes that can impact multiple organ systems, including dermatologic manifestations and multiple hamartomas[40;41]. Importantly, CS was a clinical diagnosis based on established clinical diagnostic criteria[42], until the susceptibility gene for CS was mapped to 10q22.23 in 1996[43]. In 1997, *PTEN*, a tumor suppressor gene, was identified as the molecular cause

of CS[44]. In 1999, the term “PTEN hamartoma tumor syndrome (PHTS)” was coined as a genetic diagnosis, encompassing all the clinical syndromes associated with germline *PTEN* mutations, including CS, Bannayan-Riley-Ruvalcaba syndrome, and other *PTEN*-related overgrowth syndromes[45].

Since then, the phenotypic and molecular characterization of PHTS have been codified in national practice guidelines. In the US, the National Comprehensive Cancer Network (NCCN) guideline outlines the testing criteria for PHTS as well as clinical diagnostic criteria for CS. Major criteria include personal history of BC, endometrial cancer, follicular thyroid cancer, multiple GI hamartomas or ganglioneuroma, macrocephaly, and mucocutaneous lesions. Minor criteria include personal history of autism spectrum disorder (ASD), colon cancer, greater than 3 lesions of esophageal glycogenic acanthoses, lipomas, intellectual disability (IQ less than 75), papillary or follicular variant of papillary thyroid cancer, thyroid structural lesions (e.g., adenoma, nodules, goiter), renal cell carcinoma, a single GI hamartoma/ganglioneuroma, testicular lipomatosis and vascular anomalies[36]. To qualify for revised clinical diagnostic criteria for CS, three or more major criteria (one must include macrocephaly, Lhermitte-Duclos disease, or GI hamartomas), or two major and three minor criteria need to be met. Also, individuals who have close relatives with known pathogenic or likely-pathogenic germline mutations in *PTEN* should be tested.

Family history may not be informative when clinical features of PHTS are absent or there are no known pathogenic or likely-pathogenic mutations in the family. The frequency of *de novo PTEN* mutations is reported to be 10.7% and

maximally 47.6%[46]. When clinical suspicion of PHTS is made, it is useful to utilize the Cleveland Clinic PTEN risk calculator, which provides the pretest probability of detecting a germline *PTEN* variant[47]. This online tool provides a Cleveland Clinic (CC) score, which guides clinicians to make an informed decision whether proceeding with genetic evaluation of the *PTEN* gene is warranted. Since the CC score is also a surrogate of phenotypic burden, more extensive genetic testing including exome sequencing may be beneficial in the research setting to identify other genetic etiologies which are related to non-*PTEN* genes such as *SDHx* and *WWP1*[48;49]. This is particularly relevant for individuals who have a high CC score, indicating a high phenotypic burden, and who yet test negative for germline *PTEN* alterations.

Prevalence, clinical burden, and breast cancer predisposition of PTEN hamartoma tumor syndrome (PHTS)

PTEN hamartoma tumor syndrome (PHTS) is one of the hereditary cancer syndromes, which causes heritable predisposition to multiple types of cancer including BC[50]. In women with PHTS, BC represents the malignancy with the highest lifetime risk of up to 85-91%, compared to 12.9% in the general population[50-54]. The average age of diagnosis is between 38 and 46 years[50], compared to 63 years old in the general population[37]. Furthermore, women with PHTS have a significantly higher incidence of second primary malignant neoplasms, particularly BC[55].

PTEN, encoding phosphatase and tensin homolog (MIM: 158350), is a tumor suppressor gene[56] and is among the most commonly somatically altered genes in diverse sporadic malignancies including BC[50;57]. PHTS is a molecular diagnosis that encompasses individuals harboring a germline *PTEN* mutation, which causes heritable predisposition to multiple types of cancer including breast, thyroid, kidney, endometrial and colon cancers, and melanoma[50]. Among all the PHTS component cancers, BC has the highest lifetime risk, followed by thyroid cancer (approximately 35%), renal cell cancer (approximately 34%) and endometrial cancer (approximately 28%)[50;58].

It has been estimated that the prevalence of Cowden syndrome (one of the clinical manifestations of PHTS) is 1 in 200,000[59]. However, this appears to be an underestimation, as the incidence of PHTS is expected to increase as clinical genetic testing becomes more widely accessible in the clinic.

PTEN tumor suppressor gene

PTEN is a tumor suppressor gene located on the long arm of chromosome 10 and codes for the PTEN protein[56]. Like all tumor suppressor genes, a mutation anywhere along the length of *PTEN* could lead to loss-of-function and can be potentially disease-causing. Indeed, germline loss-of-function *PTEN* mutations have been found across the span of the gene. Recurrent mutations, however, tend to cluster in exons 5, 7 and 8 (phosphatase domain). The most frequently found germline mutations at amino acid residue arginine at position

130 (R130X/G/L/Q), occur in the phosphatase core motif within exon 5[60]. There are two reasons for this: one is biological whereby the phosphatase domain houses the most frequent mutations[61]; and the R residues are encoded by the triplet CGn where CpGs deaminate and thus are prone to mutations[62].

The *PTEN* gene is composed of nine exons, encoding the PTEN protein made up of a total of 403 amino acids[56]. The PTEN protein has a PIP-binding domain (PBD), a phosphatase domain, a C2 membrane binding domain, a C-terminal tail with a PDZ-binding motif. PTEN has a protein phosphatase activity as well as a lipid phosphatase activity[63;64]. The PTEN protein antagonizes the action of PI3K via its lipid phosphatase activity[65]. This downregulates the downstream pathway including AKT/mTOR[66]. When PTEN loses function or is dysfunctional, the AKT/mTOR pathway will be upregulated, leading to increased cell growth and dysregulated cell cycle[67].

PTEN signaling and overgrowth pathway in PHTS

PTEN was first recognized as a tumor suppressor gene based on somatic *PTEN* loss in brain, breast and prostate cancers and later discovered as an underlying genetic cause of Cowden syndrome, a subtype of PHTS, in 1997[44;56]. *PTEN* is a negative regulator of the phosphatidylinositol 3-kinase (PI3K) activity, an important oncogene which promotes cell growth and survival, occurring in multiple human cancers including BC[68;69]. PI3K has a 110 kDa catalytic subunit encoded by *PIK3CA* gene and an 85 kDa regulatory subunit encoded by *PIK3R1*[70;71]. Through its catalytic region, PI3K phosphorylates the third position

of the phosphatidylinositol 4,5-biphospate (PIP2), a plasma membrane bound phospholipid, to phosphatidylinositol 3,4,5-biphospate (PIP3), which leads to the activation of the AKT serine/threonine kinase (AKT) and mammalian target of rapamycin (mTOR) cell proliferation pathway[72]. Conversely, as a tumor suppressor, PTEN antagonizes PI3K activity by dephosphorylating PIP3 to PIP2, thus dampening AKT/mTOR signaling.

Once PI3K is activated, AKT is recruited to the plasma membrane, leading to phosphorylation of up to a hundred substrates, leading to multiple cellular functions including anti-apoptotic effects, cell proliferation, and increase in the glucose transporter GLUT4 translocation through the plasma membrane to increase glucose catabolism[72]. Downstream of PI3K activation is mTORC1, which activates multiple substrates, leading to cell proliferation by inactivating the 4E-binding protein 1 (4EBP1), which inhibits eIF4E. The eIF4E is the cap-binding protein which binds to mRNA and allows translation initiation by recruiting ribosomes to the mRNA[73].

With defective PTEN functions leading to upregulation of PI3K/AKT/mTOR pathway, PHTS clinically presents as an overgrowth syndrome, which inevitably predisposes affected individuals to malignancies including BC[74].

Non-canonical roles of PTEN

Beyond the well-established canonical downstream PI3K/AKT/mTOR pathway, PTEN has multiple non-canonical roles, eg, in the nucleus, mainly involved in DNA repair, maintaining genomic integrity, chromatin remodeling and

stabilization, the apoptotic process, and cell cycle regulation[75-77]. PTEN shuttles from the cytoplasm into the nucleus despite the absence of a classic nuclear localization signal (NLS) region. There are many proposed mechanisms of how the PTEN protein localizes into the nucleus. One of the mechanisms is by simple diffusion through nuclear pores[78]. Another mechanism for PTEN translocation into the nucleus is through association with growth factor receptor bound protein 2 (GRB2) during oxidative stress[79]. Other postulated means of the nuclear localization of PTEN include transport via RAN GTPase, major vault protein (MVP), monoubiquitylation of PTEN, and a non-classical NLS[66].

PTEN associates with chromatin within the nucleus. PTEN physically interacts with the centromere at the CENP-C centromeric protein by its C-terminus, which prevents breakage at the centromere and thus prevents aneuploidy and keeps chromatin integrity[77;80]. It also has been shown that histone H1 interacts with PTEN via the C-terminal domain, facilitating chromatin condensation[76]. During mitosis, PTEN binds and co-localizes with the mitotic kinesin motor EG5, ensuring the spindle fiber to be properly aligned[81].

Once inside the nucleus, PTEN induces RAD51, which is important in DNA homologous recombination (HR) repair[77]. PTEN further directly associates with p53, a tumor suppressor, which leads to stabilization of p53 and facilitation of cell cycle arrest[82]. Once phosphorylated, PTEN binds to DNA damage sites along with ATM and gamma H2AX to facilitate non-homologous end joining DNA repair[80]. Studies have also shown PTEN interacts with Mitotic Check-point Complexes (MCC) including Cdc20, Mad1, and Mad2[83;84]. Furthermore, PTEN

controls DNA replication by binding to the minichromosome maintenance complex component2 (MCM2). Under replicative stress, PTEN dephosphorylates MCM2 at serine 41 (S41), stalling replication fork progression, thus ensuring a proper mitosis process[85].

Nuclear PTEN also regulates excessive cell growth by prohibiting the apoptotic process. For example, human glioblastoma cell lines which show the presence of PTEN predominantly in the nucleus display nucleus condensation, a marker of apoptosis, upon exposure to TNF-α[86]. Furthermore, nuclear PTEN induces cell cycle arrest at the G0-G1 phase transition point by downregulating cyclin D1. By arresting the progression of the cell cycle, nuclear PTEN contributes to tumor suppression in addition to PI3K/AKT/mTOR inhibition in the cytoplasm[66].

Clinical management of PTEN hamartoma tumor syndrome (PHTS)

Once CS/PHTS is identified, enhanced surveillance is recommended for breast, thyroid, colon, uterus, kidney, and skin. Women with PHTS should undergo high-risk breast cancer screening with mammogram and breast MRI beginning at age 30, and endometrial cancer screening with transvaginal ultrasound and/or biopsy starting at age 30. Hysterectomy should be discussed after completion of childbearing. For both genders, thyroid cancer screening with ultrasound should be started at age 7. The current guideline recommends starting whole-body skin check at the time of PHTS diagnosis and then annually. However, in a recent prospective study, skin cancers were not observed in the

pediatric population (age <=18 years). Thus, future guidelines should recommend a baseline skin exam at the time of diagnosis and another skin examination during puberty for children[87]. Renal ultrasound should start at age 40, and colon cancer screening with colonoscopy at age 35. Children, especially those with personal history of developmental delay, should have appropriate autism spectrum disorder evaluation and subsequent management by a neurodevelopmental specialist. Detailed screening recommendations are outlined in the NCCN guidelines, which are routinely updated[36].

Historically, PHTS-derived BCs have been treated similarly to sporadic BCs, according to the standard of care in the absence of somatic landscape data. Although targeting *PTEN*-associated alterations for therapeutic purposes has tremendous challenges[60;88], more targeted, personalized treatment strategies to effectively treat PHTS-derived BCs are warranted.

CHAPTER II: EXOME SEQUENCING REVEALS A DISTINCT SOMATIC GENOMIC LANDSCAPE IN BREAST CANCER FROM WOMEN WITH GERMLINE *PTEN* VARIANTS.

Introduction

PTEN hamartoma tumor syndrome (PHTS) causes heritable predisposition to multiple types of cancer including breast, thyroid, kidney, endometrial and colon cancers, and melanoma[50]. BC has the highest lifetime risk, which is up to 85-91%, with an average age of diagnosis between 38 and 46 years[50-54]. In contrast, the lifetime risk of developing BC in the general population is estimated to be 12.9% with an average age of onset at approximately 61 years of age[2].

Clinically, BC arising in PHTS appear distinct from sporadic BC without underling germline mutations in BC susceptibility genes including *PTEN*. This is characterized by younger onset, multifocality and secondary BCs[55]. Yet, very little is known regarding the molecular mechanism of tumorigenesis in PHTS-derived BCs. As a result, they are treated similarly to sporadic BCs, according to the standard of care in the absence of somatic landscape data. In order to elucidate distinct underlying biological and molecular processes, we analyzed exome sequencing data from 44 primary BCs arising in the setting of germline *PTEN* variants (PHTS) and compared them to 497 sporadic BC samples from the Cancer Genome Atlas (TCGA).

Subjects and methods

Research participants

Approved by the Cleveland Clinic's Institutional Review Boards (IRB), written informed consents were obtained from each individual enrolled under the study protocol. Among 6934 research participants prospectively accrued from September 1, 2005 to September 10, 2020, we identified 3066 female participants with a personal history of breast cancer (BC). Of these, 130 had germline *PTEN* variants. We then identified 44 women with appropriate consents for acquisition of biospecimens and whose tissues representing BC were available for sequencing.

Germline *PTEN* mutation and deletion/duplication status was confirmed by clinical genetic testing and verified by polymerase chain reaction (PCR) and/or multiplex ligation-dependent probe amplification (MLPA). *PTEN* pathogenicity predictions were derived from genetic test reports from orthogonal testing in CLIA-certified facilities, ClinVar database classifications, and/or the ClinGen gene-specific criteria for *PTEN* variant curation[89]. For unreported variants, Franklin by Genoox, an online variant interpretation tool based on machine learning and the American College of Medical Genetics and Genomics (ACMG) criteria, was used in conjunction with expert opinions at the Genomic Medicine Institute at the Cleveland Clinic (Cleveland, OH, USA). We classified variants in the *PTEN* promoter region as mutation positive if the clinical presentations were associated with Cowden syndrome or if the variants are known to affect PTEN functions[47;50;90;91]. We further classified germline *PTEN* variants into two groups; 1) Tier 1 variants: classified as pathogenic or likely pathogenic by ClinVar or

Genoox, or either large deletions, nonsense, or frameshift mutations which are predicted to be damaging either by impairing PTEN function and transcript stability; and 2) Tier 2 variants: classified as variants of uncertain significance or likely benign by ClinVar or Genoox, or they have never been reported with sufficient supporting evidence. Tier 2 variants were initially classified as mutations or determined as potentially deleterious based on clinical presentations and functional evidence at the time of the individuals' enrollment into the study. We further defined Tier 1 derived tumors as TIER-1 and Tier 2 derived tumors as TIER-2. The baseline Cleveland Clinic score (CC score), which is a semi-quantitative surrogate of age-related PHTS phenotypic burden[47], was extracted from the Cleveland Clinic Genomic Medicine Institute's relational database.

Original formalin-fixed, paraffin-embedded (FFPE) samples representing primary BC were obtained from respective healthcare institutions where the pathology specimens were collected. Two of the 44 cases had exposure to systemic chemotherapy prior to sample collection from which the source DNA originated. Clinical information including tumor type, grade, stage, and hormone receptor status was obtained by reviewing surgical pathology reports and pertinent medical records (**Table 1**, **2** in the Results section).

DNA extraction

DNA was extracted from the FFPE samples using QIAamp® DNA FFPE Tissue kit (Qiagen, Maryland, USA). Briefly, tissues from FFPE blocks were

deparaffinated with xylene and crude DNA was precipitated with 100% ethanol. Following complete proteolysis of the samples with Proteinase K at 56 degrees Celsius, DNA was extracted and purified using the column method according to the manufacturer's protocol with slight reagent volume modifications. For matched germline samples, we obtained blood-derived genomic DNA originating from whole blood from the PHTS individuals from the Genomic Medicine Biorepository of the Genomic Medicine Institute at the Cleveland Clinic (Cleveland, OH, USA) following standard procedures.

Processing of extracted DNA samples

Samples with sufficient DNA yields and quality were subjected to exome sequencing. The DNA concentration was measured with the Qubit™ Fluorometer dsDNA HS (High Sensitivity) Assay kit (Thermo Fisher Scientific, Waltham, Massachusetts, USA). While the ideal DNA concentration for sequencing library preparation was considered to be 30-40 ng/µL, the range of DNA concentrations of submitted samples was 9.6-68.4 ng/µL and 19.0-98.4 ng/µL, and the range of sample volumes submitted was 30-45 µL and 30-60 µL for tumor and normal samples, respectively.

Exome sequencing

The tumor-blood DNA pairs were sent to the Broad Institute Genomic Services (Cambridge, MA, USA) for next generation sequencing (NGS). The Broad Institute created libraries from the submitted DNA samples and used the

Illumina HiSeq platform to generate NGS data. Of the 44 tumor-normal samples, 28 were processed with the Illumina Somatic Exome protocol and the remaining 16 with the TWIST Somatic Exome protocol (pair-end sequencing with read length range of 67bp to 140 bp). The Illumina Somatic Exome protocol had target depths of 20x and 50x for the normal and tumor samples, respectively. For the TWIST Somatic Exome protocol, the target depth was x100 for both normal and tumor samples. The raw data were quality controlled, aligned and sorted through a standard NGS pipeline at the Broad Institute. Reads were aligned to the reference human genome GRCh37/hg19 using the BWA-ALN aligner (version 0.5.9)[92]. Local realignment, duplicate removal and base quality score recalibration were performed using the Genome Analysis Toolkit and Picard per the Broad Institute standard protocol[93]. The processed sequencing data, derived from both tumor and blood samples, were delivered as binary alignment map (BAM) files.

Sporadic breast cancer cohort

The control cohort data were derived from The Cancer Genome Atlas (TCGA) breast cancer dataset from the Genomic Data Commons (GDC). BC cases with available exome sequencing data were selected. Cases with germline mutations in known cancer susceptibility genes were identified based on previously published data[94] and excluded. Pertinent clinical information of the selected cases was obtained from Nationwide Children's Hospital dataset, which is publicly available from the GDC portal (universally unique identifier [UUID]

8162d394-8b64-4da2-9f5b-d164c54b9608). The final control cohort comprised of 497 women with no known germline mutations in cancer susceptibility genes and, thus, who developed sporadic BC (estrogen receptor [ER] positive/human epidermal growth factor 2 [HER2] negative: n=308; ER+/HER2+: n=80; ER-/HER2+: n=23; triple negative breast cancer [TNBC]: n=86). The original input files (BAMs) of tumor and matched normal samples, aligned to reference human genome GRCh38/hg38, were downloaded from the GDC archive website for bioinformatics analyses.

Somatic variant calling

We analyzed the BAM files using our in-house bioinformatics pipeline, which is composed of the following steps: 1) variant calling for somatic single-nucleotide variants (SNVs) and small insertions and deletions (indels); 2) variant annotation; 3) data visualization; 4) downstream analyses. To optimize recall and precision in the variant calling step, we used three variant callers: 1) The Genome Analysis toolkit (GATK) Mutect2 (version 4.1.9.0)[93]; 2) Strelka (version 2.9.10)[95]; and 3) VarDict (version 1.8.2)[96]. For each variant caller, we employed the default settings. Outputs from these variant callers were combined by SomaticCombiner[97], which generates a consensus set of somatic variants as a single variant call format (VCF) file. Somatic variants which passed all the SomaticCombiner filters with a read depth of at least 20x for normal (blood) and 20x for tumor samples were included for visual inspection using the Integrative Genomics Viewer (IGV) (version 2.6.3). The cut-off of 20 (Phred-scaled quality score) was used for both mean base call quality and mean mapping quality.

Following the filtering step, the detected variants were annotated with GATK4/Funcotator (version 4.1.9.0) according to the GATK best practice workflow of the Broad Institute, which generates mutation annotation format (MAF) files.

Variant filtration and selection

For targeted analysis, we aggregated lists of genes associated with BC from the TCGA publication[32], NCCN Genetic/Familial High-Risk Assessment: Breast and Ovarian guidelines (version 1.2022-August 11, 2021) and 22 previously reported gold standard (GS) genes for BC[98]. We further searched for preliminary BC susceptibility genes and targetable BC-associated genes from the literature and compiled a list to examine mutational status in these genes[48;98-116]. A total of 84 BC-associated genes were included in the final list (**Table 3**). Variants in all BC-associated genes were manually reviewed using the IGV with the original BAM files as input. The determination whether the variant reviewed should pass or fail was made with modified criteria based on the standard operating procedure (SOP) for somatic variant refinement of tumor-normal pair sequencing data[117]. During visual inspection, SNVs and indels were allowed to pass if the variant allele frequency (VAF) was greater than 5% and they were not classified as benign or likely benign. Variants with VAFs between 2% and 5% were retained if they were previously reported as pathogenic or likely pathogenic. An online variant interpretation tool, Franklin by Genoox was used to assist with variant classification. To remove variants found at relatively high frequencies in

the general population (minor allele frequency [MAF] greater than 0.001 [0.1%]), the allele frequency information for female samples from a large population database was used (field “gnomAD_exome_AF_female” from the genome Aggregation Database, gnomAD version 2.1, Funcotator Data Source version 1.7.20200429s). Sex determination for this gnomAD dataset is based on both X heterozygosity and Y coverage and it contains a wide range of ancestries.

Identification of fequently mutated genes and gene interactions

In order to identify the most frequently mutated BC-associated genes, we used the maftools package (version 2.10.0)[118] in R to summarize and visualize our dataset. R version used was R-4.1.2 for windows. Samples containing variants in BC-associated genes which passed the manual review on IGV were included in the analyses. We used metafor package in R (version 3.4.0)[119] to identify which frequently altered genes are statistically significantly different in frequency between PHTS and sporadic BC groups from TCGA. Forest plots were created using GraphPad Prism version 9.0 (GraphPad Software, San Diego, California, USA). The subset analyses were done to compare TIER-1 (n=31) and TIER-2 (n=13) against TCGA BCs (n=497). To investigate the presence of gene sets which were co-occurring or mutually exclusive in PHTS-all (TIER-1 and TIER-2 BCs combined), we used the ‘somaticInteractions’ function in maftools to create a pair-wise correlation matrix plot based on Fisher’s exact test to identify significant pairs of genes.

Driver gene identification

To uncover cancer driver genes, we performed a cancer-type agnostic gene ranking analysis using cDriver (version 0.4.2)[98]. The input files included both synonymous and non-synonymous variants which passed all the SomaticCombiner filters, with read depth of greater than 20x for both tumor and normal, and with mean base call quality and mean mapping quality above 20 (Phred-scaled quality). An optimized VAF cutoff of 2% was used for this analysis since we observed that known pathogenic variants in genes such as *PIK3CA* (MIM: 114480), although rare, can occur with VAF lower than 5%. To generate purity and ploidy information, FACETS (version 0.6.1), an R package, was used[120].

Mutational signature identification

To characterize the genome-wide mutational landscape, we investigated mutational signatures including apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like (APOBEC) enrichment, and single base substitution (SBS) signatures. Variants which passed all the SomaticCombiner filters with minimal read depths of 20x for both normal and tumor samples, and a VAF greater than 5% were included. We used the BSgenome package (version 1.62.0) to create a matrix of nucleotide substitutions which served as input for APOBEC enrichment and SBS signature analyses. APOBEC enrichment scores were estimated using a previously described method[121]. We then used maftools 'plotApobecDiff' function to create a plot showing the difference in mutational load of tCw (W=A or

T) motif as well as differentially mutated genes between APOBEC enriched and non-APOBEC enriched samples. For SBS signature analyses, we used non-negative matrix factorization, NMF package in R (version 0.24.0)[122], to measure the goodness of fit in order to determine the number of best match clusters. Then, we used maftools 'extractSignatures' function to extract mutational signatures to compare against known SBS signatures in the Catalogue of Somatic Mutations in Cancer (COSMIC) legacy and SBS signature database. Finally, we used maftools 'plotSignatures' function to plot best match SBS signatures detected in our dataset.

Tumor mutational burden estimation

Tumor mutational burden (TMB) was calculated by taking the total number of non-synonymous variants divided by a capture size of 30. This capture size is based on the size of the coding region, for which 30Mb was employed. This approximation is a generally accepted assumption that approximately 1% of the whole human genome (an approximate total of 3 billion base pairs) is protein-coding[123]. Non-synonymous variants which passed all the SomaticCombiner filters, had minimal read depths of 20x for both normal and tumor samples, and had VAF greater than 5% were included. The cut-off of 20 (Phred-scaled quality) was used for both mean base call quality and mean mapping quality. The TMB was calculated for each sample which was then log2 transformed for optimum scatterplot visualization. Kruskal-Wallis test was used to perform an overall comparison of all eight BC groups (PHTS-all, TIER-1, TIER-2, all sporadic BC,

ER+/HER2-, ER+/HER2+, ER-/HER2+, and TNBC). For post-hoc, pairwise comparisons, we employed Mann-Whitney test with adjusted p-values <0.05 to be considered statistically significant.

TCGA data analysis

We applied the same variant calling algorithm to the raw TCGA sporadic BC dataset to make a head-to-head comparison with our PHTS series data.

Sample size estimation

We performed sample size calculations to determine the minimum number of cases we need to show statistically significant genomic differences between the PHTS and TCGA sporadic BC groups. In order to detect characteristic driver mutations at the variant level, we used the two proportions derived from the somatic *PTEN* mutation rate in the preliminary PHTS group with 29 samples (21.0%) and that of sporadic luminal subtypes in literature (4.0%)[32;124]. We estimated that 30 samples from PHTS and 250 samples from TCGA should be sufficient to achieve a power of 81.0% with an alpha of 0.05 (two-sided) to detect a significant difference.

Statistical analysis

To compare clinical characteristics and somatic mutational landscapes between PHTS and TCGA BCs, a two-tailed Chi-squared or Fisher's exact test was applied to categorical variables. For continuous variables, analysis of

variance was used. A logistic regression model was used for a binary outcome with continuous independent variables. Statistical analyses were performed using GraphPad Prism version 9.0 (GraphPad Software, San Diego California, USA), except for statistical analyses incorporated in maftools (version 2.10.0)[118]. P-values <0.05 were considered statistically significant unless stated otherwise.

Results

Notable differences in the clinical characteristics and somatic variant spectra between PHTS-derived BCs and sporadic counterparts from TCGA

In all combined PHTS group (PHTS-all, n=44), the majority of cases had ductal histology (86.4%), were ER positive (84.1%), progesterone receptor (PR) positive (81.8%), and HER2 negative (84.1%). The most common grade was 2 (intermediate grade, 52.3%) and the majority of cases were early stage, namely stages 0, I and II (95.5%, including cases whose stages are unknown and presumed to be early-stage). The median age of diagnosis was 49.5 years (range, 34-85 years) and 47 years (range, 34-75 years) in PHTS and Tier 1, respectively, both of which were significantly younger than that of TCGA (median, 56 years; range, 34-85 years; p=0.003 [PHTS vs. TCGA] and p<0.001 [Tier-1 vs. TCGA]; **Table 1**). Compared to TCGA, PHTS-all displayed a significantly higher proportion of ductal histology (86.4% vs 66.8%; p=0.007), and of earlier clinical stages (95.5% vs 75.3%; p=0.001; **Table 1**).

Table 1: Clinical characteristics of women with PHTS-derived breast cancers compared to those with sporadic breast cancer from TCGA

Characteristics	PHTS (n=44)	Tier 1 (n=31)	Tier 2 (n=13)	TCGA (n=497)	P-value [a]
Median age of diagnosis – yr.	49.5	47.0	53.0	56	0.003 [b]
Diagnosis age range – yr.	34 - 85	34-75	40-85	34 - 85	[<0.001] [b]
IQR (95% CI of median)	14.5 (45 to 54)	16 (40-56)	17.5 (46-63.5)	17.5 (55 to 59)	(>0.99)[b]
Age of diagnosis – No. (%)					
=< 40 yr.	9 (20.5)	8 (25.8)	1 (7.7)	35 (7.0)	
41-49 yr.	13 (29.5)	10 (32.3)	3 (23.1)	111 (22.3)	
50-59 yr.	13 (29.5)	9 (29.0)	4 (30.8)	139 (28.0)	
>= 60 yr.	9 (20.5)	4 (12.9)	5 (38.5)	212 (42,7)	
Tumor Type – No. (%)					
Ductal	38 (86.4)	27 (87.1)	11 (84.6)	332 (66.8)	0.007 [c]
Lobular	1 (2.3)	1 (3.2)	0 (0)	117 (23.5)	[0.02] [c]
Mixed histology	5 (11.4)	3 (9.7)	2 (15.4)	15 (3.0)	(0.24) [c]
Other/Unknown	0 (0)	0 (0)	0 (0)	33 (6.6)	
Stage - No. (%)					
0	1 (2.3)	1 (3.2)	0 (0)	0 (0)	0.001[d]
I	26 (59.1)	18 (58.1)	8 (61.5)	88 (17.7)	[0.3] [d]
II	11 (25.0)	7 (22.6)	4 (30.8)	286 (57.5)	(0.5) [d]
III	1 (2.3)	1 (3.2)	0 (0)	119 (23.9)	
IV	1 (2.3)	1 (3.2)	0 (0)	4 (0.8)	
Unknown	4 (9.1)	3 (9.7)	1 (7.7)	0	
ER status – No. (%)					
Positive	37 (84.1)	26 (83.9)	11 (84.6)	388 (78.1)	0.25 [e]
Negative	6 (13.6)	4 (12.9)	2 (15.4)	109 (21.9)	[0.37] [e]
Unknown	1 (2.3)	1 (3.2)	0 (0)	0 (0)	(0.74) [e]
PR status – No. (%)					
Positive	36 (81.8)	26 (83.9)	10 (76.9)	338 (68.0)	0.03 [e]
Negative	7 (15.9)	4 (12.9)	3 (23.1)	159 (32.0)	[0.03] [e]
Unknown	1 (2.3)	1 (3.2)	0 (0)	0 (0)	(0.77) [e]
HER2 status – No. (%)					
Positive	5 (11.4)	3 (9.7)	2 (15.4)	101 (20.3)	0.33 [f]
Negative	37 (84.1)	26 (83.9)	11 (84.6)	396 (79.7)	[0.49] [f]
Unknown	2 (4.5)	2 (6.5)	0 (0)	0 (0)	(>0.99) [f]
Grade – No. (%)					
I	12 (27.3)	7 (22.6)	5 (38.5)	Unknown	
II	23 (52.3)	17 (54.8)	6 (46.2)	Unknown	
III	9 (20.5)	7 (22.6)	2 (15.4)	Unknown	

Abbreviations: PHTS = all PTEN hamartoma tumor syndrome breast cancer cases combined, Tier 1 = breast cancer cases arising from germline PTEN variants classified as pathogenic or likely pathogenic, Tier 2 = breast cancer cases arising from germline

PTEN variants classified as variants of unknown significance or likely benign, TCGA = The Cancer Genome Atlas breast cancer cohort (all subtypes combined), IQR = interquartile range, ER = estrogen receptor, PR = progesterone receptor, HER2 = human epidermal growth factor receptor 2

[a] *Comparison between PHTS and TCGA is shown, followed by Tier-1 compared to TCGA (in the brackets) and Tier-2 compared to TCGA (in the parentheses). Differences were considered statistically significant with p<0.01 (two-tailed, Bonferroni corrected) unless stated otherwise*

[b] *While the PHTS group is normally distributed, TCGA cohort was not. Thus, a nonparametric test (the Kruskal-Wallis) was performed. Adjusted p-value <0.05 is considered statistically significant.*

[c] *Fisher's exact test was performed between pure ductal vs non-pure ductal (lobular, mixed, other, unknowns) to meet the criteria of all expected values greater than 1 and at least 20% of the expected values greater than 5.*

[d] *Fisher's exact test was performed between early stage (0, I, II, unknown) vs advanced (III and IV) since Chi-squared test calculations on 2x6 require all expected values to be greater than one.*

[e] *Fisher's exact test was performed between positive marker status (ER or PR positive and unknown) vs negative marker status (ER negative or PR negative) to meet the criteria of all expected values greater than 1 and at least 20% of the expected values are greater than 5. Unknown cases were considered ER and PR positive.*

[f] *Fisher's exact test was performed between HER2 positive vs. HER2 negative. One unknown case was considered positive (for a case of ductal carcinoma in situ) and the other, negative (for a case HER2-directed treatment history is not documented)*

Somatic variants in PTEN as the predominant, somatic oncogenic drivers in PHTS-derived BCs, especially in TIER-1.

In PHTS-all BCs, *PTEN* and *PIK3CA* were the most frequently somatically-mutated BC-associated genes, each affecting 22.7% of the samples, followed by *MAP3K1* (MIM: 600982 [13.6%]) , *TP53* (MIM: 114480 [11.4%]), *GATA3* (MIM: 131320 [9.1%]), and *TBX3* (MIM: 601621 [9.1%]; **Figure 1A** and **Table 4**). All the somatic hits in *PTEN* were distinct from the germline *PTEN* variants (**Table 2, 4**), representing second hits to *PTEN*, while the underlying germline *PTEN* variants represent the first hits per Knudson's two-hit hypothesis[125]. Most of these somatic variants have been previously reported and considered pathogenic or likely pathogenic (**Table 4**). Two individuals had two somatic hits in *PTEN*. The relative risk for the presence of somatic *PTEN* variants was 4.03 (95% CI, 2.1 to 7.75; p<0.001) in PHTS-all BCs compared to TCGA. The relative risk for the presence of somatic *PTEN* variants was 5.15 in TIER-1 BCs (95% CI, 2.7 to 9.9; p<0.01) while it was 1.36 in TIER-2 BCs (95% CI, 0.2 to 9.3; p=0.75) compared to TCGA. The majority of somatic variants detected in *PIK3CA* were pathogenic hotspot mutations (**Table 4**). Somatic variants in *PTEN* and *PIK3CA* were found to be mutually exclusive in PHTS-all BCs and this was statistically significant (p=0.01; **Figure 1B**).

Table 2 Clinical characteristics of PHTS-related breast cancer series

Case ID	Histologic subtype	Dx Age (yr.)	Stage	ER	HER2	CC score
PHTS-A	IDC & DCIS	58	IA	+	-	14
PHTS-B	IDC & DCIS	39	IA	+	-	25
PHTS-C	IDC & DCIS	45	IA	+	-	34
PHTS-D	IDC & DCIS	45	IA	+	-	20
PHTS-E	IDC & DCIS	43	IA	+	-	42
PHTS-F	IDC & ILC	56	IIA	+	-	46
PHTS-G	IDC & DCIS	37	IIB	+	-	43
PHTS-H	IDC & DCIS	49	IA	N/A	N/A	22
PHTS-I	IDC & DCIS	52	IB	+	-	13
PHTS-J	IDC	37	IIB	-	+	37
PHTS-K	IDC with lobular features	60	IA	+	-	16
PHTS-L	IDC & DCIS	54	IA	+	-	14

PHTS-M	Invasive mucinous Ca & DCIS	46	IIA	+	-	7
PHTS-N	IDC & DCIS	65	IA	+	-	36
PHTS-O	IDC & DCIS	44	IA	+	-	18
PHTS-P[a]	IDC & DCIS	40	IIB [b]	-	-	2
PHTS-Q	IDC with lobular features	58	IA	+	-	16
PHTS-R	IDC & DCIS	62	IA	+	+	8
PHTS-S	IDC with focal infiltrating lobular dif. & DCIS	42	IIA	+	-	1
PHTS-T	IDC & DCIS	56	IA	+	-	51
PHTS-U	IDC & DCIS	63	IA	+	-	28
PHTS-V	IDC & DCIS	43	IA	-	+	41
PHTS-W	Invasive mucinous Ca	52	Unk	+	+	2
PHTS-X	IDC	56	IIA	+	-	1
PHTS-Y	ILC & LCI	55	Unk	+	-	35

PHTS-Z	Papillary Ca (invasive and in situ) & DCIS	50	IA	+	-	18
PHTS-AA	IDC & DCIS	34	IA	+	-	20
PHTS-AB	IDC & DCIS	61	IIA	+	-	25
PHTS-AC	IDC	75	IIA	+	-	1
PHTS-AD	IDC	79	IA	+	-	8
PHTS-AE	IDC & DCIS	85	IA	+	-	7
PHTS-AF	IDC & DCIS	45	IA	+	-	9
PHTS-AG	IDC & DCIS	39	Unk	+	-	28
PHTS-AH	IDC	49	IIIA	+	-	8
PHTS-AI	IDC, DCIS & LCIS	38	IA	+	-	32
PHTS-AJ	IDC	46	IIA	+	-	36
PHTS-AK	DCIS	50	0	+	N/A	32
PHTS-AL[a]	IDC	60	Unk	-	-	28
PHTS-AM	IDC & DCIS	47	IIA	+	-	3

PHTS-AN	IDC & DCIS	38	IB	-	-	32
PHTS-AO	IDC & DCIS	54	IA	+	-	28
PHTS-AP	IDC & DCIS	40	IA	+	-	24
PHTS-AQ	IDC & DCIS	53	IA	-	-	4
PHTS-AR	IDC & DCIS	47	IV	+	+	8

Table 2 Clinical characteristics of PHTS-related breast cancer series (Continued)

Case ID	Germline *PTEN* variants	Genoox classification	Supporting evidence	Final variant priority (Tier)
PHTS-A	c.238A>T p.(Lys80*)	Likely pathogenic	Wang and Jiang 2008[1] Ngeow et al. 2012[2]	1
PHTS-B	c.1003C>T p.(Arg335*)	Pathogenic	See ClinVar: VCV000007833.40	1
PHTS-C	c.(801+1_802-1)_(1026+1_1027-1)del	N/A	Zhou et al. 2003[3]	1
PHTS-D	c.1026+1G>A	Pathogenic	See ClinVar: VCV000183722.9	1
PHTS-E	c.210-4_210-1del	Pathogenic	See ClinVar: VCV000427615.2	1
PHTS-F	c.697C>T p.(Arg233*)	Pathogenic	See ClinVar: VCV000007813.20	1
PHTS-G	c.389G>A p.(Arg130Gln)	Pathogenic	See ClinVar: VCV000007829.31	1
PHTS-H	c.71A>G p.(Asp24Gly)	Pathogenic	See ClinVar: VCV000208723.9	1
PHTS-I	c.-1195_1184del	N/A	Wang et al. 2011[4] Tan et al. 2012[5]	2
PHTS-J	c.549_550dup p.(Asn184Argfs*16)	Pathogenic	See ClinVar: VCV000484620.3	1
PHTS-K	c.235G>A p.(Ala79Thr)	VUS	Jaini et al. 2020[6]	2
PHTS-L	c.517C>T p.(Arg173Cys)	Pathogenic	See ClinVar: VCV000189500.14	1
PHTS-M	c.328C>G p.(Gln110Glu)	Likely pathogenic	See ClinVar: VCV000428229.8	1
PHTS-N	c.974T>C p.(Leu325Pro)	VUS	See ClinVar: VCV000428231.3 Mighell et al. 2020[7]	2
PHTS-O	c.697C>T p.(Arg233*)	Pathogenic	See ClinVar: VCV000007813.20	1

PHTS-P[a]	c.-834C>T	N/A	Zhou et al. 2003[3] Teresi et al. 2007[8] Tan et al. 2012[5]	2
PHTS-Q	c.518G>A p.(Arg173His)	Pathogenic	See ClinVar: VCV000376032.17	1
PHTS-R	c.892C>G p.(Gln298Glu)	VUS	See ClinVar: VCV000127695.18	2
PHTS-S	c.-1170C>T	N/A	Wang et al. 2011[4] Tan et al. 2012[5]	2
PHTS-T	c.(0+1_1-1)_(1212+1_1213-1)del	N/A	Jaini et al. 2020[6]	1
PHTS-U	c.315dup p.(Glu106*)	Pathogenic	See ClinVar: VCV000468681.1	1
PHTS-V	c.401T>G p.(Met134Arg)	Pathogenic	See ClinVar: VCV000431829.2	1
PHTS-W	c.1061C>A p.(Pro354Gln)	VUS	See ClinVar: VCV000143020.27	2
PHTS-X	c.-1066_-1037del	N/A	Tan et al. 2012[5]	2
PHTS-Y	c.697C>T p.(Arg233*)	Pathogenic	See ClinVar: VCV000007813.20	1
PHTS-Z	c.1003C>T p.(Arg335*)	Pathogenic	See ClinVar: VCV000007833.40	1
PHTS-AA	c.732_733insGG p.(Gln245Glyfs*12)	Likely pathogenic	Wang and Jiang 2008[1] Ngeow et al. 2012[2]	1
PHTS-AB	c.696dup p.(Arg233Thrfs*10)	N/A	Wang and Jiang 2008[1] Ngeow et al. 2012[2]	1
PHTS-AC	c.123_130del p.(Leu42Argfs*7)	N/A	Wang and Jiang 2008[1] Ngeow et al. 2012[2]	1
PHTS-AD	c.-1170C>T	N/A	Wang et al. 2011[4] Tan et al. 2012[5]	2
PHTS-AE	c.-1170C>T	N/A	Wang et al. 2011[4] Tan et al. 2012[5]	2

PHTS-AF	c.-1170C>T	N/A	Wang et al. 2011[4] Tan et al. 2012[5]	2
PHTS-AG	c.522T>A p.(Tyr174*)	Pathogenic	See ClinVar: VCV000486227.3	1
PHTS-AH	c.(0+1_1-1)_(79+1_80-1)del	N/A	Zhou et al. 2003[3]	1
PHTS-AI	c.635-9A>G p.(Asn212Thrfs*12)[d]	VUS	Chen, et al. 2017[9] Wang and Jiang 2008[1] Ngeow et al. 2012[2]	1
PHTS-AJ	c.892del p.(Gln298Lysfs*9)	Pathogenic	See ClinVar: VCV000189432.1	1
PHTS-AK	c.(79+1_80-1)_(164+1_165-1)del	N/A	Jaini et al. 2020[6] Zhou et al. 2003[3]	1
PHTS-AL[a]	c.209+1G>T	Pathogenic	See ClinVar: VCV000229705.6	1
PHTS-AM	c.-890C>T	N/A	Tan et al. 2012[5]	2
PHTS-AN	c.389G>A p.(Arg130Gln)	Pathogenic	See ClinVar: VCV000007829.31	1
PHTS-AO	c.1003C>T p.(Arg335*)	Pathogenic	See ClinVar: VCV000007833.40	1
PHTS-AP	c.1003C>T p.(Arg335*)	Pathogenic	See ClinVar: VCV000007833.40	1
PHTS-AQ	c.114T>G p.(Pro38=)	VUS	See ClinVar: VCV000184068.10	2
PHTS-AR	c.620G>T p.(Ser207Ile)	Likely pathogenic	See Franklin[c]	1

The transcript accession number for PTEN germline variants (RefSeq ID): NM_000314.3
Clinical stage is based on medical records or determined by pathology reports using the American Joint Committee on Cancer (AJCC)7th edition breast cancer staging criteria.
Abbreviations: Ca = Carcinoma, DCIS = Ductal carcinoma in situ, Dx = Diagnosis, GMI= Genomic Medicine Institute at the Cleveland Clinic, Cleveland, OH, IDC = Invasive ductal carcinoma, ILC = Invasive lobular carcinoma, LCIS = Lobular carcinomain situ, Unk = Information unknown, VUS = variant of unknown significance, yr. = years

[a] *Exposure to neoadjuvant chemotherapy*

[b] *Staging based on path report before neoadjuvant chemotherapy. Staging after neo-adjuvant is reported to be IIA (yT1bN1aMx)*

[c] https://franklin.genoox.com/clinical-db/variant/snp/chr10-89712002-G-T (last accessed on February, 23, 2022).
[d] The transcript change and predicted protein truncation were confirmed molecularly using Rapid Amplification of cDNA Ends (RACE)
ClinVar and Genoox websites access date: February 23, 2022.

Table 3: Aggregated breast cancer (BC) associated genes

Gene Name	Chrom	Coordinates (GRCh37/hg19)	
		Start	*End*
ABRAXAS1	4	84380669	84406253
AFF2	X	147582136	148082193
AKT1	14	105235685	105262085
AKT3	1	243651534	244006584
APC	5	112043194	112181936
AR	X	66763862	66950461
ARID1A	1	27022505	27108595
ATM	11	108093793	108239829
BARD1	2	215590369	215674407
BRAF	7	140413127	140624729
BRCA1	17	41196311	41277381
BRCA2	13	32889644	32974405
BRIP1	17	59756499	59940889
CASP8	2	202098165	202152434
CBFB	16	67063051	67134961
CCND1	11	69455923	69469242
CCND2	12	4382927	4414519
CCND3	6	41902670	42016632
CCNE1	19	30302897	30315219
CD274	9	5450541	5470554
CDH1	16	68771194	68869440
CDK4	12	58141509	58146093
CDK6	7	92234234	92465887
CDKN1B	12	12870301	12875303
CDKN2A	9	21967750	21995323
CDKN2B	9	22002901	22009312
CHEK2	22	29083730	29137822
CTCF	16	67596428	67673080
EGFR	7	55086709	55279321
ERBB2	17	37844346	37884911
ERBB3	12	56473948	56497289
ESR1	6	151977806	152424409
ESR2	14	64693424	64805331
FANCM	14	45605132	45670093
FBXW7	4	153241695	153457244
FGFR1	8	38268660	38326153
FGFR2	10	123237843	123357972
FOXA1	14	38058756	38064454

GATA3	10	8096650	8117161
IGF1	12	102789651	102874341
JAK2	9	4984389	5129948
KLLN	10	89618914	89623290
KMT2A	11	118307206	118397547
KMT2C	7	151832009	152133088
KRAS	12	25358179	25403863
MAP2K1	15	66679249	66783882
MAP2K4	17	11924193	12047145
MAP3K1	5	56111375	56191979
MET	7	116312249	116438431
MRE11	11	94148735	94227010
MUTYH	1	45794913	45806112
MYB	6	135502445	135540310
NBN	8	90945558	90996895
NCOR1	17	15932470	16118848
NF1	17	29421944	29704695
NOTCH1	9	139388884	139440500
NOTCH4	6	3505440	3534689
NTHL1	16	2089820	2097835
PALB2	16	23614485	23652631
PGR	11	100900354	101000544
PIK3CA	3	178866144	178957881
PIK3CB	3	138371539	138553770
PIK3R1	5	67511583	67597649
PTEN	10	89623381	89731687
PTPN22	1	114356432	114414381
PTPRD	9	8314245	10613002
RAD50	5	131892668	131982041
RAD51C	17	56769962	56812972
RAD51D	17	33419239	33446879
RB1	13	48877886	49056026
RECQL	12	21621843	21654569
RINT1	7	105172647	105208124
RPTOR	17	78518637	78940168
RUNX1	21	36160097	36421599
SDHB	1	17345216	17380527
SDHD	11	111957596	111966518
SEC23B	20	18488191	18542059
SF3B1	2	198283519	198299817
STK11	19	1205776	1228430
TBL1XR1	3	176737131	176915270
TBX3	12	115108059	115121980

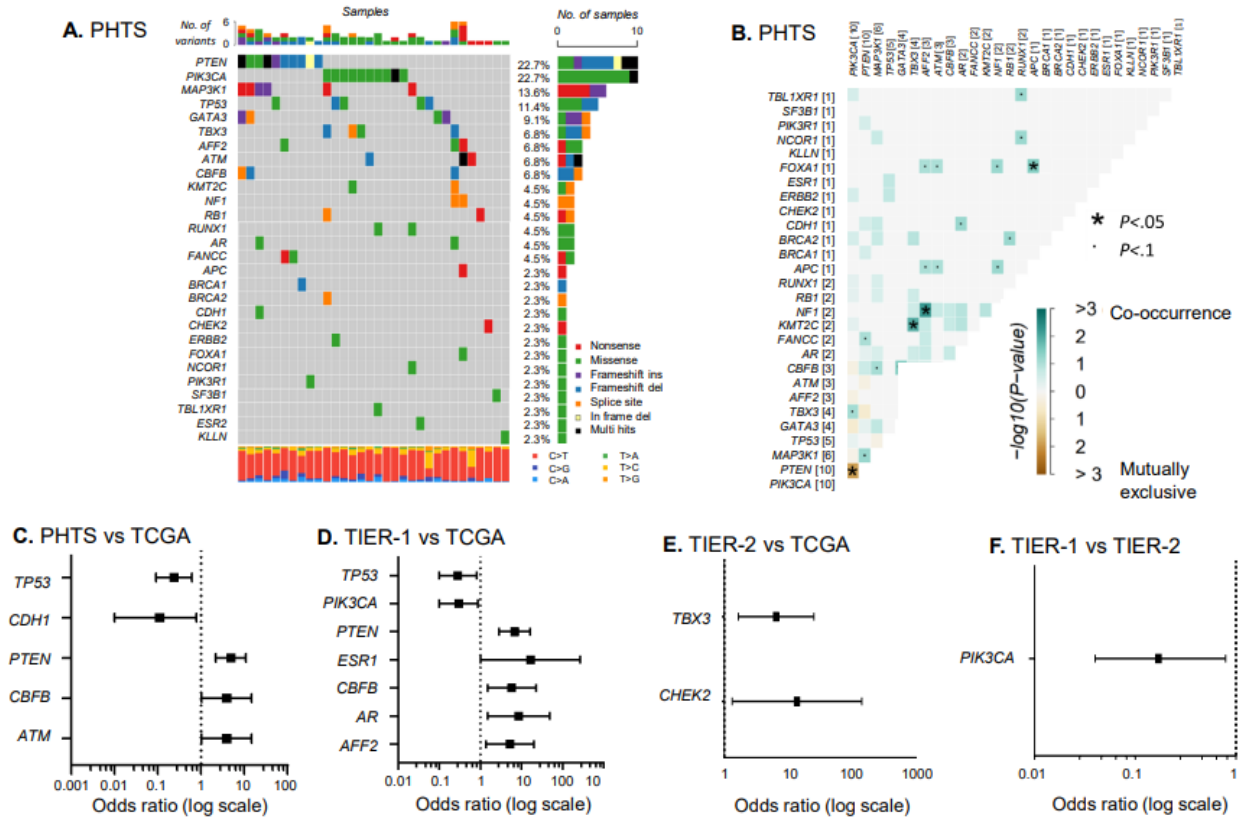
TP53	17	7571738	7590808
WWP1	8	87354775	87480732
XRCC2	7	152341860	152373226

Table 4: Somatic variants in PTEN and PIK3CA detected in the PHTS series

Sample	HUGO symbol	Interpretation (Genoox)	cDNA change	Protein change
PHTS-AB	*PTEN*	Likely pathogenic	c.283_289del	p.(Pro95Serfs*2)
PHTS-AE	*PTEN*	Pathogenic	c.855_856del	p.(Glu285Aspfs*12)
PHTS-AE	*PTEN*	Likely pathogenic	c.1026G>C	p.(Lys342Asn)
PHTS-AK	*PTEN*	Likely pathogenic	c.513_521del	p.(Gln171_Tyr174delinsHis)
PHTS-E	*PTEN*	Likely pathogenic	c.674_675del	p.(Tyr225Phefs*17)
PHTS-F	*PTEN*	Likely pathogenic	c.535dup	p.(Ser179Lysfs*11)
PHTS-G	*PTEN*	Not previously reported	c.768_784del	p.(Phe257Glnfs*35)
PHTS-J	*PTEN*	Likely pathogenic	c.609_619del	p.(Ile203Metfs*36)
PHTS-L	*PTEN*	Pathogenic	c.741dup	p.(Pro248Thrfs*5)
PHTS-L	*PTEN*	Pathogenic	c.448G>A	p.(Glu150Lys)
PHTS-Q	*PTEN*	Pathogenic	c.737C>T	p.(Pro246Leu)
PHTS-Z	*PTEN*	Pathogenic	c.464A>G	p.(Tyr155Cys)
PHTS-AC	*PIK3CA*	Pathogenic	c.3140A>G	p.(His1047Arg)
PHTS-AD	*PIK3CA*	Pathogenic	c.1258T>C	p.(Cys420Arg)
PHTS-AF	*PIK3CA*	Pathogenic	c.3140A>G	p.(His1047Arg)
PHTS-AF	*PIK3CA*	Pathogenic	c.2191C>T	p.(Gln731*)
PHTS-AH	*PIK3CA*	Pathogenic	c.1633G>A	p.(Glu545Lys)
PHTS-AM	*PIK3CA*	Pathogenic	c.3140A>T	p.(His1047Leu)
PHTS-AO	*PIK3CA*	Likely pathogenic	c.1093G>A	p.(Glu365Lys)
PHTS-AQ	*PIK3CA*	Pathogenic	c.1258T>C	p.(Cys420Arg)
PHTS-AR	*PIK3CA*	Pathogenic	c.3140A>G	p.(His1047Arg)
PHTS-K	*PIK3CA*	Pathogenic	c.3140A>G	p.(His1047Arg)
PHTS-X	*PIK3CA*	Pathogenic	c.1633G>A	p.(Glu545Lys)

RefSeq IDs: NM_000314.4 (PTEN), NM_006218.2 (PIK3CA).

Figure 1: Somatic mutational landscape of PHTS-derived breast cancers (BCs) and comparison to sporadic BCs from TCGA



A. Oncoplot showing the distribution of non-synonymous somatic mutations in BC-associated genes among the PHTS-all samples (TIER-1 and TIER-2 BCs combined). Each column represents a sample and each row, a BC-associated gene with at least one somatic variant detected in any sample. The gene names are listed in order of the highest to the lowest mutational frequency, which is the proportion of all samples affected (out of 44 samples) shown as a percentage. The top bar plot shows the number of variants detected in each sample. The bar plot on the far right shows the number of samples harboring variants in each gene. The bottom bar plot shows transition and transversion patterns for each sample. Abbreviations: ins = insertion, del = deletion.

B. The correlation plot shows co-occurring and mutually exclusive gene pairs in PHTS-all (TIER-1 and TIER-2 BCs combined). Pair-wise Fisher's exact test (two-tailed) was performed to identify statistically significant pairs. The numbers in brackets represent the number of samples harboring non-synonymous variants in each gene. The plot shows variants in PTEN and PIK3CA occur in a mutually exclusive manner in PHTS-all BCs and this was statistically significant (p=0.01). The p-values for statistically significant gene pairs are: p=0.006 for NF1 and AFF2, p=0.01 for KMT2C and TBX3, p=0.01 for PTEN and PIK3CA, p=0.03 for FOXA1 and APC, and p=0.03 for GATA3 and CBFB.

C. The forest plot shows odds ratio (OR) to compare the PHTS-all series (n=44) and TCGA cohort (n=497) for statistically significantly mutated BC-associated genes. OR >1

indicates there were more somatic variants in the PHTS-all group (ATM, CBFB, and PTEN). OR <1 indicates there were more somatic variants in the TCGA cohort (CDH1 and TP53). All BC subtypes from both groups are included in this analysis. The horizontal bar represents 95% confidence interval (CI).

D. The forest plot shows OR to compare TIER-1 BCs (n=31) and TCGA (n=497) cohort for statistically significantly mutated BC-associated genes. OR >1 indicates there were more somatic variants in the TIER-1 BC group (AFF2, AR, CBFB, ESR1, and PTEN). OR <1 indicates there were more somatic variants in the TCGA cohort (TP53 and PIK3CA). All BC subtypes from both groups are included in this analysis. The horizontal bar represents 95% CI.

E. The forest plot shows odds ratio (OR) to compare TIER-2 BCs (n=13) and TCGA cohort (n=497) for statistically significantly mutated BC-associated genes. OR >1 indicates there were more somatic variants in the TIER-2 BC group (TBX3 and CHEK2). All BC subtypes from both groups are included in this analysis. The horizontal bar represents 95% CI
F. The forest plot shows odds ratio (OR) to compare TIER-1 series (n=31) and TIER-2 series (n=13) for statistically significantly mutated BC-associated genes. OR <1 indicates there were less somatic variants in the TIER-1 group (PIK3CA). The horizontal bar represents 95% CI.

Genomic data reveal TIER-1 and TIER-2 BCs are two genomically different types of tumors.

Subset analysis with TIER-1 BC cases (n=31) revealed a higher somatic mutational frequency in *PTEN* (29.0%) and a lower mutational frequency in *PIK3CA* (12.9%) compared to all the PHTS samples combined (PHTS-all, n=44). Compared to TIER-1 BCs, TIER-2 had a higher mutational frequency in *PIK3CA* (46.1%) and a lower frequency in *PTEN* (7.7%; **Table 5**). Based on the data above, TIER-1 and TIER-2 BCs appear to be two genomically different types of tumors. Thus, we performed TIER-1 and TIER-2 analyses separately, along with our overall PHTS series (PHTS-all).

CC score is a clinically useful tool to provide pretest probability of detecting a germline *PTEN* variant and is a useful surrogate of phenotypic burden[47]. We performed logistic regression to examine if the CC score was associated with the presence of somatic *PTEN* or *PIK3CA* variants in the PHTS groups. In PHTS-all BCs, no significant association was found between the CC score and the occurrence of somatic *PTEN* variants (odds ratio (OR) per 5-point increase in CC score=1.28; 95% CI, 0.98 to 1.72; p=0.08; **Figure 2A**). There was, however, a statistically significant relationship between a lower CC score and an increased likelihood of detecting somatic *PIK3CA* mutations (OR per 5-point increase in CC score=0.54; 95% CI, 0.31 to 0.78; p=0.007; **Figure 2B**). This association persisted in TIER-1 (**Figure 2C** and **2D**) but not in TIER-2 BCs (**Figure 2E** and **2F**).

For the entire sporadic BC cohort from TCGA (n=497), including all the four subtypes, *TP53* was identified as the most frequently mutated BC-associated gene, affecting 34.4% of the samples, followed by *PIK3CA* (33.4%), *CDH1* (MIM: 192090 [17.7 %]), *GATA3* (11.3%), and *KMT2C* (MIM: 606833 [10.3%]). *PIK3CA* was the most frequently mutated gene in ER+ subgroups (40.6% in ER+/HER2- and 36.3% in ER+/HER2+), while *TP53* was the most frequently mutated in ER- subgroups (73.9% in ER-/HER2+ and 82.6% in TNBC). Somatic variants in *PTEN* were only found at low frequencies (5.6-7.0%), with the highest of 7.0% in the TNBC subgroup and 5.6% in the combined sporadic BC cohort (**Table 5**).

We found that several BC-associated genes were significantly more somatically altered in PHTS-all BCs than in TCGA: *ATM* (MIM: 607585 [OR = 3.97; 95% CI, 1.03 to 15.23; p=0.04])*, CBFB* (MIM: 121360 [OR = 3.97; 95% CI, 1.03 to 15.23; p=0.04])*, and PTEN* (OR = 4.93; 95% CI, 2.21 to 10.98; p<0.001). In contrast, we found two significantly less altered BC-associated genes in PHTS-all BCs than in TCGA: *CDH1* (OR = 0.11; 95% CI, 0.01 to 0.80; p=0.03) and *TP53* (OR = 0.24; 95% CI, 0.09 to 0.63; p=0.04; **Figure 1C**). Compared to TCGA, TIER-1 BCs had five BC-associated genes which were more significantly altered: *AFF2* (MIM: 300806 [OR = 5.22; 95% CI, 1.36 to 20.03; p=0.02]), *AR* (MIM: 313700 [OR = 8.50; 95% CI, 1.49 to 48.34; p=0.02), *CBFB* (OR = 5.81; 95% CI, 1.49 to 22.66; p=0.01), *ESR1* (MIM: 114480 [OR = 16.53; 95% CI, 1.01 to 270.85; p=0.049]), and *PTEN* (OR = 6.85; 95% CI, 2.89 to 16.26; p<0.001), while two BC-associated genes were significantly less altered: *PIK3CA* (OR =

0.30; 95% CI, 0.10 to 0.86; p=0.03) and *TP53* (OR = 0.28; 95% CI, 0.10 to 0.82; p=0.02; **Figure 1D**). TIER-2 BCs had two BC-associated genes which were more significantly altered: *CHEK2* (MIM: 604373 [OR, 13.72; 95% CI, 1.33-141.67; p=0.03]) and *TBX3* (OR, 6.48; 95% CI, 1.66-25.22; p=0.007; **Figure 1E**) compared to TCGA. When TIER-1 and TIER-2 BCs were compared, *PIK3CA* was significantly less somatically mutated in TIER-1 compared to TIER-2 BCs (OR, 0.17; 95% CI, 0.04 to 0.79; p=0.02).

cDriver identified *PTEN* and *TP53* as the top ranked driver genes in PHTS-all and TCGA BCs, respectively (**Table 6, 7**). Other high-ranking BC-associated genes in PHTS-all BCs included *PIK3CA*, *MAP3K1*, *GATA3*, *CBFB*, and *TBX3*. While the top 10 genes ranked for TCGA samples were all known BC-associated genes, some non-BC-associated genes were ranked as top driver genes in PHTS-all BCs (*ZNF253* [MIM: 606954], *TCHH* [MIM: 190370], *STAP2* [MIM: 607881], *NPIPA5* , *ZNF676*, *MAGEC1* [MIM: 300223], *MDC1*[MIM: 607593]). *PTEN* remained to be the top driver gene but *PIK3CA* ranked number six in TIER-1 BCs (**Table 8**). On the other hand, *PIK3CA* was identified as the top driver gene but *PTEN* failed to be ranked within the top 14 in TIER-2 BCs (**Table 9**).

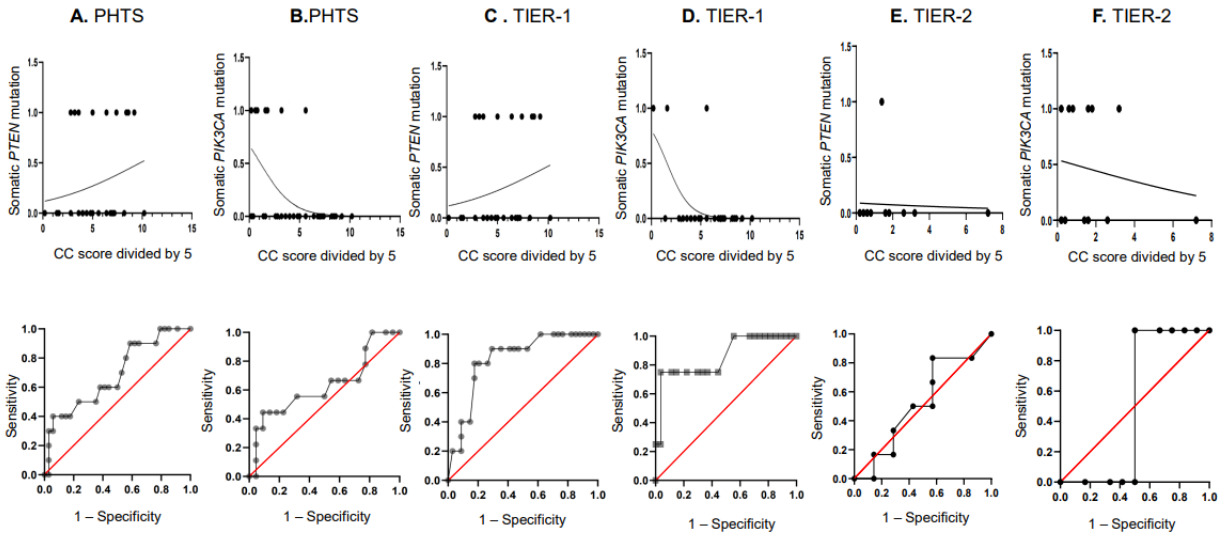
Table 5: Most frequently mutated breast cancer-associated genes in breast cancers derived from PHTS and TCGA

	PHTS			TCGA				
Genes	All-cases (n=44) No. (%)	TIER-1 (n=31) No. (%)	TIER-2 (n=13) No. (%)	All cases (n=497) No. (%)	ER+/HER2- (n=308) No. (%)	ER+/HER2+ (n=80) No. (%)	ER-/HER2+ (n=23) No. (%)	TNBC (n=86) No. (%)
PTEN	10 (22.7)	9 (29.0)	1 (7.7)	28 (5.6)	17 (5.5)	5 (6.3)	0 (0)	6 (7.0)
PIK3CA	10 (22.7)	4 (12.9)	6 (46.2)	166 (33.4)	125 (40.6)	29 (36.3)	2 (8.7)	10 (11.6)
MAP3K1	6 (13.6)	3 (9.7)	3 (23.1)	45 (9.1)	32 (10.4)	8 (10.0)	3 (13.0)	2 (2.3)
TP53	5 (11.4)	4 (12.9)	1 (7.7)	171 (34.4)	62 (20.1)	21 (26.3)	17 (73.9)	71 (82.6)
GATA3	4 (9.1)	4 (12.9)	0 (0)	56 (11.3)	48 (15.6)	8 (10.0)	0 (0)	0 (0)
TBX3	4 (9.1)	1 (3.2)	3 (23.1)	22 (4.4)	19 (6.2)	1 (1.3)	0 (0)	2 (2.3)
AFF2	3 (6.8)	3 (9.7)	0 (0)	10 (2.0)	3 (1.0)	4 (5.0)	1 (4.4)	2 (2.3)
ATM	3 (6.8)	2 (6.5)	1 (7.7)	9 (1.8)	5 (1.6)	2 (2.5)	0 (0)	2 (2.3)
CBFB	3 (6.8)	3 (9.7)	0 (0)	9 (1.8)	7 (2.3)	2 (2.5)	0 (0)	0 (0)
KMT2C	2 (4.5)	1 (3.2)	1 (7.7)	51 (10.3)	38 (12.3)	8 (10.0)	0 (0)	5 (5.8)
NF1	2 (4.5)	2 (6.5)	0 (0)	16 (3.2)	10 (3.3)	3 (3.8)	0 (0)	3 (3.5)
RB1	2 (4.5)	1 (3.2)	1 (7.7)	15 (3.0)	3 (1.0)	3 (3.8)	0 (0)	9 (10.5)
RUNX1	2 (4.5)	1 (3.2)	1 (7.7)	21 (4.2)	14 (4.6)	5 (6.3)	0 (0)	2 (2.3)
AR	2 (4.5)	2 (6.5)	0 (0)	4 (0.8)	3 (1.0)	1 (1.3)	0 (0)	0 (0)
FANCC	2 (4.5)	2 (6.5)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)

Abbreviations: PHTS = PTEN hamartoma tumor syndrome (germline PTEN variant positive), TCGA = The Cancer Genome Atlas, TIER-1 = breast cancers arising from germline PTEN variants classified as pathogenic or likely pathogenic, TIER-2 = breast cancers arising from germline PTEN variants classified as variants of unknown significance or likely benign,
ER = estrogen receptor, HER2 = human epidermal growth factor receptor 2, TNBC = triple negative breast cancer

The table shows a comparison of the most frequently mutated BC-associated genes between PHTS (all cases combined), TIER-1, TIER-2 and TCGA sporadic breast cancer groups. The genes are listed in descending order by the percentage of samples harboring somatic variants (single-nucleotide variants and indels) in each gene.

Figure 2: Logistic regression and receiver operating characteristic (ROC) curves for CC scores to predict somatic variants in PHTS, TIER-1 and TIER-2 breast cancers



A. The relationship between the CC score (per 5 point increase) and the presence of somatic PTEN variants in PHTS-all BCs. Upper panel shows a logistic regression curve with the CC score (per 5 point increase) on the x-axis and probability of detecting somatic PTEN variants on the y-axis. Odds ratio (OR) was 1.28 (95% confidence interval [CI], 0.98 to1.72; p=0.08). Lower panel shows ROC curve from the logistic regression model. The area under the curve was 0.69 (95% CI, 0.50 to 0.88; p=0.08).

B. The relationship between the CC score (per 5 point increase) and the presence of somatic PIK3CA variants in PHTS-all BCs. Upper panel shows a logistic regression curve with the CC score (per 5 point increase) on the x-axis and probability of detecting somatic PIK3CA variants on the y-axis. Odds ratio was 0.54 (95% CI, 0.31 to 0.78; p=0.007). Lower panel shows ROC curve from the logistic regression model. The area under the curve was 0.83 (95% CI, 0.70 to 0.96; p=0.002).

C. The relationship between the CC score (per 5 point increase) and the presence of somatic PTEN variants in TIER-1 BCs. Upper panel shows a logistic regression curve with the CC score (per 5 point increase) on the x-axis and probability of detecting somatic PTEN variants on the y-axis. Odds ratio was 1.23 (95% CI, 0.89 to 1.80; p=0.24). Lower panel shows ROC curve from the logistic regression model. The area under the curve was 0.63 (95% CI, 0.39 to 0.86; p=0.27).

D. The relationship between the CC score (per 5 point increase) and the presence of somatic PIK3CA variants in TIER-1 BCs. Upper panel shows a logistic regression curve with the CC score (per 5 point increase) on the x-axis and probability of detecting somatic PIK3CA variants on the y-axis. Odds ratio was 0.42 (95% CI, 0.15 to 0.79; p=0.03). Lower panel shows ROC curve from the logistic regression model. The area under the curve was 0.86 (95% CI, 0.64 to 1.00; p=0.02).

E. The relationship between the CC score (per 5 point increase) and the presence of somatic PTEN variants in TIER-2 BCs. Upper panel shows a logistic regression curve with the CC score (per 5 point increase) on the x-axis and probability of detecting somatic PTEN variants on the y-axis. Odds ratio (OR) was 0.89 (95% confidence interval [CI], -2.66 to 0.83; p=0.87). Lower panel shows ROC curve from the logistic regression model. The area under the curve was 0.50 (95% CI, 0.22 to 0.78; p>0.99).

F. The relationship between the CC score (per 5 point increase) and the presence of somatic PIK3CA variants in TIER-2 BCs. Upper panel shows a logistic regression curve with the CC score (per 5 point increase) on the x-axis and probability of detecting somatic PIK3CA variants on the y-axis. OR was 0.82 (95% CI, -1.12 to 0.43; p=0.57). Lower panel shows ROC curve from the logistic regression model. The area under the curve was 0.52 (95% CI, 0.20 to 0.85; p=0.89).

Table 6. cDriver ranking in PHTS series (TIER-1 and TIER-2 BCs combined)

A. Raw ranking by cDriver, involving all variants detected (Top 20 genes)

Rank	Gene	Bayes risk	Bayes driver	Risk Rank	Driver Rank	sample no. with mutation	total no. of mutation	CCF sum	Indels count
1	*PTEN*	0.047717	1	1	2	12	14	9.82506	8
2	*ZNF253*	0.037606	1	4	11	6	7	5.495057	0
3	*PIK3CA*	0.027754	1	11	5	10	13	8.505611	0
4	*ZNF676*	0.02793	1	10	7	7	10	6.295527	0
5	*STAP2*	0.039698	1	2	15	5	6	4.461452	0
6	*MAP3K1*	0.029022	1	6	13	7	7	4.893192	2
7	*MAGEC1*	0.027209	1	12	8	7	7	6.262691	0
8	*NPIPA5*	0.03938	1	3	17	4	8	3.733615	0
9	*TCHH*	0.023208	1	21	3	11	15	9.402182	0
10	*GATA3*	0.027985	1	9	16	5	5	4.212189	4
11	*MDC1*	0.026672	1	14	12	6	6	4.985487	0
12	*CBFB*	0.032082	0.9998	5	23	5	5	3.488581	2
13	*MUC7*	0.028688	0.9995	7	26	4	6	3.389296	0
14	*SZT2*	0.020738	1	24	14	10	13	4.773305	0
15	*ZNF626*	0.01944	1	29	10	6	6	5.500638	0
16	*QRICH2*	0.024409	0.9998	18	22	7	9	3.954906	0
17	*ERC2*	0.020164	0.9999	25	18	7	7	4.296698	0
18	*AL139011.2*[a]	0.025108	0.9993	17	28	4	4	3.453971	0
19	*ZNF695*	0.025277	0.9991	16	29	4	5	3.349708	0
20	*ZNF468*	0.019952	0.9998	27	21	5	6	4.146923	0

B. Driver ranking among BC-associated genes by cDriver (Top 6 genes)

Rank	Gene	Raw Ranking	Bayes risk	Bayes driver	Risk Rank	Driver Rank	sample no. with mutation	total no. of mutation	CCF sum	Indels count
1	*PTEN*	1	0.047717	1	1	2	12	14	9.82506	8
2	*PIK3CA*	3	0.027754	1	11	5	10	13	8.505611	0
3	*MAP3K1*	8	0.029022	1	6	13	7	7	4.893192	2
4	*GATA3*	10	0.027985	1	9	16	5	5	4.212189	4
5	*CBFB*	12	0.032082	0.9998	5	23	5	5	3.488581	2
6	*TBX3*	26	0.017271	0.9994	51	27	4	4	3.968849	3

Abbreviations: CCF = cancer cell fraction, no. = numbers

Cancer driver mutation ranking for all genes detected (A), calculated using cDriver. The ranking is based on the combined statistical analyses from two types of Bayes models (bayes.risk and bays.driver per cDriver software). The input data contained all variants with VAF greater than 2%, including 2 additional splice site variants in PTEN. BC-genes were extracted and ranking is created separately (B). Suggested final cut-off by cDriver was 67.5 in this dataset.

[a] *This gene is described as a novel protein and does not have an official HUGO gene name assigned at the time of this study.*

Table 7. cDriver ranking in the TCGA cohort

A. Raw ranking by cDriver, involving all variants detected (Top 20 genes)

Rank	Gene	Bayes risk	Bayes driver	Risk Rank	Driver Rank	sample no. with mutation	total no. of mutation	CCF sum	Indels count
1	*TP53*	0.275493	1	1	1	170	174	166.8041	35
2	*PIK3CA*	0.111935	1	3	2	167	194	158.4778	6
3	*GATA3*	0.150403	1	2	4	57	60	51.87531	46
4	*CDH1*	0.100637	1	4	3	89	90	85.46817	52
5	*MAP3K1*	0.044361	1	7	6	49	69	43.0575	38
6	*PTEN*	0.058625	1	5	12	28	29	26.3021	11
7	*TBX3*	0.035263	1	15	21	23	23	19.92758	16
8	*MAP2K4*	0.036308	1	12	39	16	16	14.91751	4
9	*FOXA1*	0.032755	1	24	31	20	22	15.77101	3
10	*RB1*	0.03726	1	11	53	14	16	12.25576	6
11	*CBFB*	0.045846	1	6	72	9	9	8.955335	3
12	*MIA2*	0.029526	1	30	56	21	22	11.57539	0
13	*RUNX1*	0.023984	1	66	20	22	24	20.18999	14
14	*KDM6A*	0.029507	1	31	57	13	14	11.49238	5
15	*CRNKL1*	0.032815	1	23	67	15	17	9.955869	0
16	*ZNF41*	0.031227	1	26	71	11	11	9.000675	0
17	*TINAG*	0.032402	1	25	75	11	13	7.438018	0
18	*MT-CO2*	0.027461	1	38	69	12	12	9.387406	0
19	*MT-ND2*	0.024537	1	58	49	14	16	13.09031	3
20	*AKT1*	0.024291	1	61	47	14	14	13.63171	0

B. Driver ranking among BC-associated genes by cDriver (Top 10 genes)

Rank	Gene	Raw Ranking	Bayes risk	Bayes driver	Risk Rank	Driver Rank	sample no. with mutation	total no. of mutation	CCF sum	Indels count
1	*TP53*	1	0.275493	1	1	1	170	174	166.8041	35
2	*PIK3CA*	2	0.111935	1	3	2	167	194	158.4778	6
3	*GATA3*	3	0.150403	1	2	4	57	60	51.87531	46
4	*CDH1*	4	0.100637	1	4	3	89	90	85.46817	52
5	*MAP3K1*	5	0.044361	1	7	6	49	69	43.0575	38
6	*PTEN*	6	0.058625	1	5	12	28	29	26.3021	11
7	*TBX3*	7	0.035263	1	15	21	23	23	19.92758	16
8	*MAP2K4*	8	0.036308	1	12	39	16	16	14.91751	4
9	*FOXA1*	9	0.032755	1	24	31	20	22	15.77101	3
10	*RB1*	10	0.03726	1	11	53	14	16	12.25576	6

Abbreviations: CCF = cancer cell fraction, no. = numbers

Cancer driver mutation ranking for all genes detected (A) in TCGA, calculated using cDriver. The ranking is based on the combined statistical analyses from two types of Bayes models (bayes.risk and bays.driver per cDriver software). The input data contained all variants with VAF greater than 2% (A). BC-genes were extracted and ranking is created separately (B). Suggested final cut-off by cDriver was 242.5 in this dataset

Table 8. cDriver ranking for the TIER-1 BC series

A. Raw ranking by cDriver, involving all variants detected (Top 20 genes)

Rank	Gene	Bayes risk	Bayes driver	Risk Rank	Driver Rank	sample no. with mutation	total no. of mutation	CCF sum	Indels count
1	*PTEN*	0.055157	1	1	1	11	12	8.82506	7
2	*MDC1*	0.031277	1	7	4	5	5	4.540357	0
3	*CBFB*	0.040124	0.9998	2	10	5	5	3.488581	2
4	*GATA3*	0.032593	0.9999	6	9	4	4	3.842848	3
5	*SZT2*	0.02555	1	12	5	8	10	4.536931	0
6	*ZNF626*	0.023058	1	15	3	5	5	5	0
7	*STAP2*	0.033917	0.9984	4	16	3	4	2.999914	0
8	*MAGEC1*	0.024195	0.9999	13	8	5	5	4.262691	0
9	*PTENP1*	0.027766	0.997	8	19	3	3	3	0
10	*OR2M5*	0.035165	0.987	3	26	3	3	2.376391	0
11	*ZNF253*	0.026195	0.9958	10	22	3	4	2.954371	0
12	*NCOA6*	0.022565	0.9988	18	15	4	4	3.549924	0
13	*SCN11A*	0.023096	0.9968	14	20	5	5	3.176556	0
14	*TCHH*	0.019599	1	34	2	7	9	6.300014	0
15	*SPDYC*	0.033182	0.9687	5	38	3	4	2.166668	0
16	*ZNF100*	0.019278	0.9999	37	7	5	5	4.886202	0
17	*MUC7*	0.025711	0.9737	11	34	3	3	2.389296	0
18	*QRICH2*	0.021925	0.9886	22	24	5	5	2.792701	0
19	*DTL*	0.022612	0.9757	17	32	3	3	2.514485	0
20	*PALM2AKAP2*	0.022917	0.9726	16	35	3	3	2.463767	1

B. Driver ranking among BC-associated genes by cDriver (Top 6 genes)

Rank	Gene	Raw Ranking	Bayes risk	Bayes driver	Risk Rank	Driver Rank	sample no. with mutation	total no. of mutation	CCF sum	Indels count
1	*PTEN*	1	0.055157	1	1	1	11	12	8.82506	7
2	*CBFB*	3	0.040124	0.9998	2	10	5	5	3.488581	2
3	*GATA3*	4	0.032593	0.9999	6	9	4	4	3.842848	3
4	*MAP3K1*	21	0.020886	0.9834	28	29	4	4	2.709942	1
5	*TP53*	26	0.017579	0.9994	60	11	5	5	4.191872	2
6	*PIK3CA*	29	0.017306	0.999	63	12	4	5	4	0

Abbreviations: CCF = cancer cell fraction, no. = numbers

Cancer driver mutation ranking for all genes detected (A), calculated using cDriver. The ranking is based on the combined statistical analyses from two types of Bayes models (bayes.risk and bays.driver per cDriver software). The input data contained all variants with VAF greater than 2%. BC-genes were extracted and ranking is created separately (B). Suggested final cut-off by cDriver was 35.5 in this dataset.

Table 9 cDriver ranking for the TIER-2 BC series

A. Raw ranking by cDriver, involving all variants detected (Top 20 genes)

Rank	Gene	Bayes risk	Bayes driver	Risk Rank	Driver Rank	sample no. with mutation	total no. of mutation	CCF sum	Indels count
1	*PIK3CA*	0.055344	0.9999	6	2	6	8	4.505611	0
2	*TCHH*	0.056475	0.9981	4	4	4	6	3.102168	0
3	*ZNF253*	0.06151	0.9923	2	7	3	3	2.540686	0
4	*AL139011.2*[a]	0.061801	0.99	1	8	3	3	2.453971	0
5	*ZNF676*	0.049351	0.9961	7	5	3	5	3	0
6	*TBX3*	0.047586	0.9953	8	6	3	3	2.968849	2
7	*ZAN*	0.039857	0.9999	16	1	6	9	4.916423	0
8	*RPTN*	0.041298	0.9998	12	3	5	6	4.635565	0
9	*NPIPA5*	0.060709	0.8965	3	14	2	4	1.733615	0
10	*DTD2*	0.056356	0.9009	5	13	2	2	1.787671	0
11	*MAP3K1*	0.046908	0.9563	9	12	3	3	2.18325	1
12	*ZNF714*	0.036348	0.986	30	9	3	3	2.85802	0
13	*PPIP5K1*	0.039689	0.864	17	15	3	4	1.886798	0
14	*STAP2*	0.045249	0.6755	10	24	2	2	1.461538	0
15	*MUC17*	0.032685	0.9716	54	11	4	4	2.686345	0
16	*MAGEC1*	0.032813	0.8568	51	16	2	2	2	0
17	*TTF1*	0.032407	0.7642	57	19	3	3	1.786943	0
18	*SLC39A6*	0.036963	0.4869	26	34	2	2	1.311957	0
19	*CAD*	0.032818	0.5659	50	28	3	3	1.467348	0
20	*ZNF695*	0.036148	0.5097	32	33	2	3	1.349708	0

B. Driver ranking among BC-associated genes by cDriver (Top 3 genes)

Rank	Gene	Raw Ranking	Bayes risk	Bayes driver	Risk Rank	Driver Rank	sample no. with mutation	total no. of mutation	CCF sum	Indels count
1	*PIK3CA*	1	0.055344	0.9999	6	2	6	8	4.505611	0
2	*TBX3*	6	0.047586	0.9953	8	6	3	3	2.968849	2
3	*MAP3K1*	11	0.046908	0.9563	9	12	3	3	2.18325	1

Abbreviations: CCF = cancer cell fraction, no. = numbers

Cancer driver mutation ranking for all genes detected (A), calculated using cDriver. The ranking is based on the combined statistical analyses from two types of Bayes models (bayes.risk and bays.driver per cDriver software). The input data contained all variants with VAF greater than 2%. BC-genes were extracted and ranking is created separately (B). Suggested final cut-off by cDriver was 13.5 in this dataset.

[a] This gene is described as a novel protein and does not have an official HUGO gene name assigned

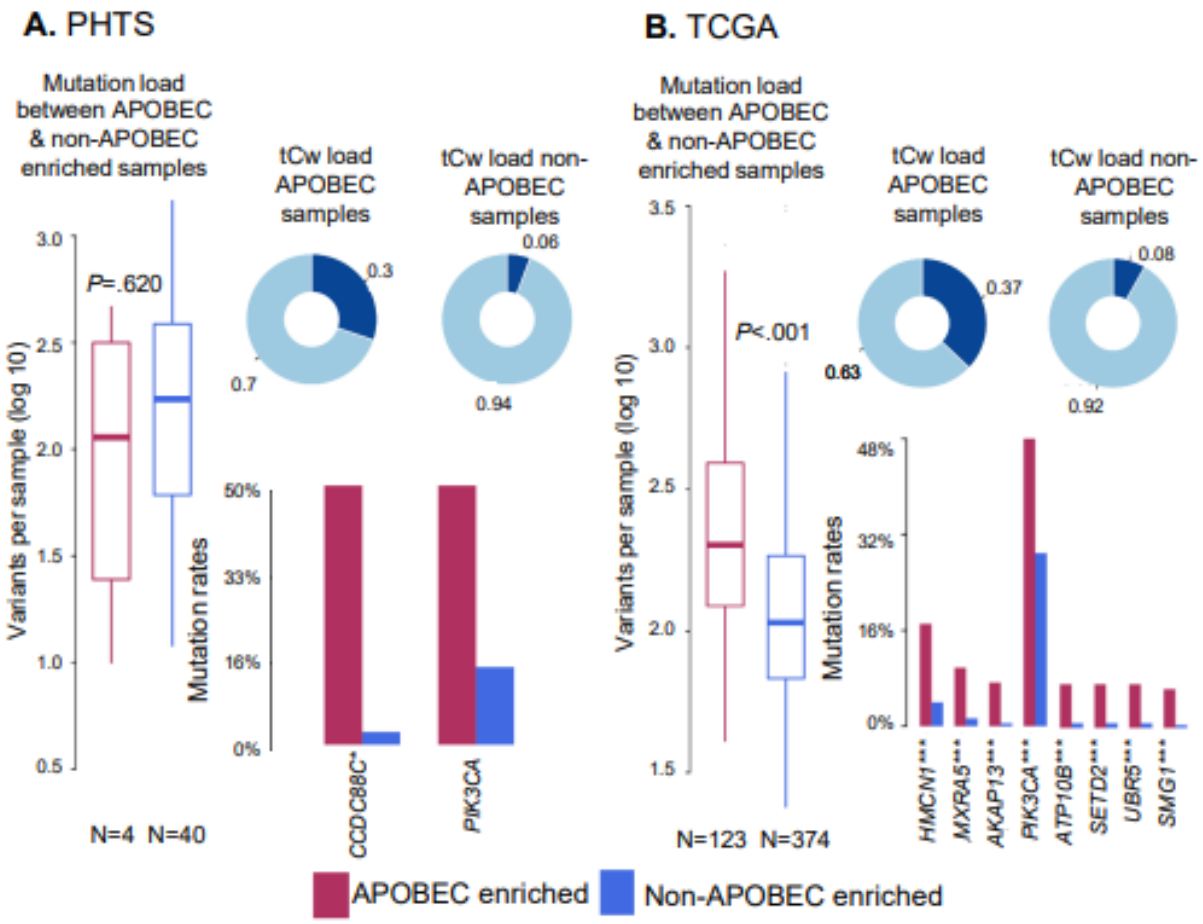
PHTS-derived BC has distinct somatic mutational signatures compared to TCGA.

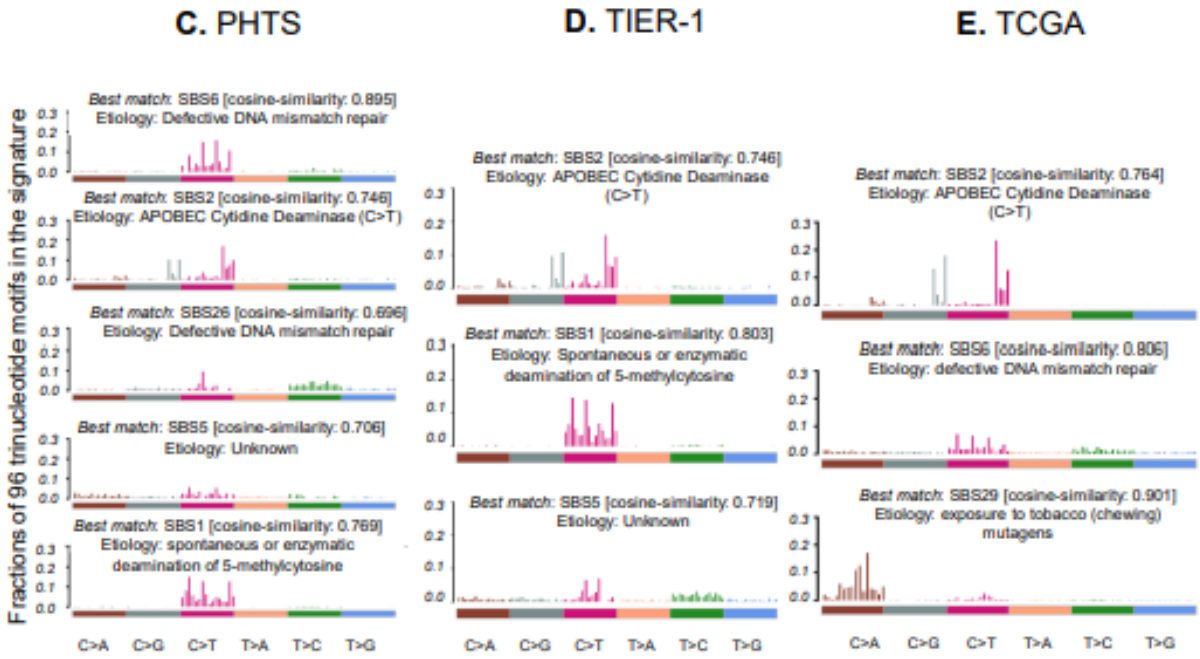
In order to identify genome-wide differences between PHTS and TCGA BCs, we performed analyses including all BC-associated genes and non-BC-associated gene variants detected in our pipeline. Previous studies showed apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like (APOBEC), a deaminase, to be one of the most important endogenous sources of mutagenesis in human cancer, especially in BC[121;126]. APOBEC-related somatic mutations were found enriched in PHTS-all (9.1%), TIER-1 (3.2%), TIER-2 (23.1%), and TCGA (21.0%) BCs. The mutational load between APOBEC and non-APOBEC-enriched samples had no statistically significant difference in PHTS-all BCs ($p=0.62$; **Figure 3A**). In contrast, there was a statistically higher mutational load in APOBEC than in non-APOBEC-enriched samples in TCGA ($p<0.001$; Figure 3B). *PIK3CA* was one of the differentially mutated genes in TCGA (OR = 2.22; 95% CI 1.43 to 3.45; $p<0.001$) but not in PHTS-all, TIER-1 or TIER-2 BCs (**Figure 3A and 3B**). Indeed, no differentially mutated genes were found in TIER-1 or TIER-2 BCs.

Single base substitution (SBS) signature analysis identified five best-match clusters in PHTS-all BCs: SBS1 (spontaneous or enzymatic deamination of 5-methylcytosine), SBS2 (APOBEC Cytidine Deaminase [C>T]), SBS5 (unknown etiology), SBS6 (defective DNA mismatch repair) and SBS26 (defective DNA mismatch repair) (**Figure 3C**). Three out of these signatures (SBS1, SBS2 and SBS5) persisted in TIER=1 BCs **Figure 3D**). In contrast, the

TCGA cohort clustered into three best-matched groups: SBS2, SBS6, and SBS29 (exposure to tobacco [chewing] mutagens) (**Figure 3E**). SBS29 was identified in TCGA but not in PHTS-all or TIER-1 BCs. SBS29 signifies mutagens like tobacco use[127] and may indicate environmental and lifestyle-related causes contributing to carcinogenesis in sporadic BCs but not in PHTS-all or TIER-1 BCs. In PHTS and TIER-1 BCs, we identified the SBS1 signature, which is strongly correlated with older age of cancer onset[127;128]. No significant SBS was found for conversions in TIER-2 BCs.

Figure 3: Genome-wide somatic mutational signature analyses





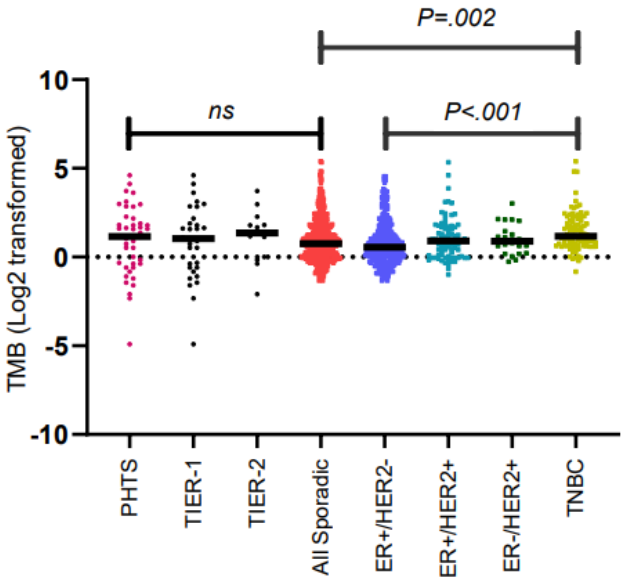
<u>A, B.</u> *Apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like (APOBEC) enrichment estimates are shown for the PHTS-all BCs (Panel A) and TCGA BC cohorts (Panel B). Box plots on the left compare the differences in mutational burden between APOBEC enriched (magenta) and non-APOBEC enriched (blue) samples. The vertical bar represents 95% CI. The pie charts on the right show the proportion of C>T transition events occurring in the TCW motif (tCw load), indicated by dark blue shadowing. APOBEC related mutations were found enriched in 9.1% of samples in the PHTS-all group compared to 31.0 % in TCGA. For PHTS-all BCs, CCDC88C (MIM: 611204) showed a mutational burden which was significantly higher in APOBEC enriched than non-APOBEC enriched samples by Fisher's exact test (odds ratio [OR] = 30.68; 95% interval [CI], 1.19 to 2298.87; p=0.02) but PIK3CA did not (OR = 5.35, 95% CI, 0.33 to 87.36; p=0.15). For TCGA, genes showing statistically higher mutational burdens in APOBEC enriched than non-APOBEC enriched samples by Fisher's exact test are shown in the right lower corner with corresponding barplots indicating differences in mutational loads. Odds ratio, 95% CI and p-values for the top 8 differentiated genes in Panel B are: HMCN1 (MIM: 608548 [OR = 4.85; 95% CI, 2.30 to 10.50; p<0.001]); MXRA5 (MIM: 300938 [OR = 7.94; 95% CI, 2.51 to 29.06; p<0.001]); AKAP13 (MIM: 604686 [OR = 4.42; 95% CI, 2.93 to 139.34; p<0.001]); PIK3CA (OR = 2.22; 95% CI, 1.43 to 3.45; p<0.001); ATP10B (MIM: 619791 [OR = 9.60; 95% CI, 2.35 to 56.03; p<0.001]); SETD2 (MIM: 612778 [OR = 9.60; 95% CI, 2.35 to 56.03; p<0.001]); UBR5 (MIM: 608413 [OR = 16.16; 95% CI 3.37 to 153.71; p<0.001]); and SMG1 (MIM: 607032 [OR = 12.7; 95% CI 2.50 to 124.68, p<0.001]). *** p-value <0.001, ** p-value <0.01, * p-value <0.05.*

<u>C, D, E.</u> *Single base substitution (SBS) signatures in PHTS-all (Panel C), TIER-1 BCs (Panel D) and TCGA (Panel E), respectively. SBS signatures were extracted and compared against known SBS signatures from the Catalogue of Somatic Mutations in Cancer (COSMIC) database. Cosign similarity was calculated to identify best matched signatures and identified signatures are plotted.*

Germline PTEN variants trend to show increased genomic instability, a hallmark of cancer[129], reflecting one of PTEN's non-canonical roles

TIER-2 BCs had the highest TMB with a median of 2.57 (IQR, 3.22; 95% CI, 0.90-4.83), followed by TNBC (median, 2.30; IQR, 2.72; 95% CI, 1.90-3.00), PHTS-all (median, 2.23; IQR 3.53; 95% CI 1.43-3.20), TIER-1 (median, 2.07; IQR, 3.83; 95% CI, 0.77-3.47), ER+/HER2+ (median, 1.88; IQR, 2.19; 95% CI, 1.47-2.23), ER-/HER2+ (median, 1.73; IQR, 1.57; 95% CI, 1.53-2.43), TCGA groups all combined (median, 1.70; IQR, 1.97; 95% CI, 1.60-1.83) and ER+/HER2- (median, 1.47; IQR, 1.63; 95% CI, 1.33-1.67). Although TIER-2 BCs had the highest TMB value, there were no statistically significant differences compared to any of the other groups (**Figure 4**). Among TCGA BC subtypes, TNBC had a statistically higher median TMB than the ER+/HER2- cohort and all sporadic BC combined ($p<0.001$). The failure to demonstrate statistically significant difference in TMB between PHTS groups (especially TIER-2 BCs) and the sporadic BC cohorts may be due to the limited sample size.

Figure 4: Comparative analysis of Tumor Mutational Burden (TMB)



Adjusted *P* values <.05 are considered statistically significant

Median tumor mutational burden (TMB) in Mut/Mb

Groups	Median	IQR	95% CI
PHTS (n=44)	2.23	3.53	1.43 to 3.20
TIER-1 (n=31)	2.07	3.83	0.77 to 3.47
TIER-2 (n-13)	2.57	3.22	0.90 to 4.83
All sporadic (n=497)	1.70	1.97	1.60 to 1.83
ER+/HER2- (n=308)	1.47	1.63	1.33 to 1.67
ER+/HER2+ (n=80)	1.88	2.19	1.47 to 2.23
ER-/HER2+ (n=23)	1.73	1.57	1.53 to 2.43
TNBC (n=86)	2.30	2.72	1.90 to 3.00

Post-hoc analysis

Groups compared	Adjusted *P* value
PHTS vs. All sporadic	.55
TIER-1 vs. All sporadic	>.99
TIER-2 vs. All sporadic	>.99
TIER-1 vs. TIER-2	>.99
PHTS vs. ER+/HER2-	.23
All sporadic vs. TNBC	.002
ER+/HER2- vs. TNBC	<.001

Scatter plot showing log2 transformed tumor mutational burden (TMB) values from each sample derived from the counts of non-synonymous variants with variant allele frequency greater than 5%. In this plot, eight groups were compared with a Kruskal-Wallis test. Each dot in the graph represents one sample. The black horizontal bars near the center of the scatter plots represent log2 transformed median TMB values. The vertical lines running perpendicular to the median bars represent 95% confidence intervals (CI). The raw TMB median values in mutations per megabase (Mut/Mb), interquartile ranges (IQR), and 95% CIs are shown in the right upper panel. Post-hoc sub-analyses were performed with Mann-Whitney test and the results are shown in the right lower panel. Adjusted p-values <0.05 are considered statistically significant. Abbreviations: ns = not statistically significant, IQR = interquartile range, Mut/Mb = mutations per megabase

Discussion

We identified notable differences in the clinical characteristics and somatic variant spectra between PHTS-derived BCs and sporadic counterparts from TCGA. Our data reveal several key findings, providing insights into distinct BC biology in the background of germline *PTEN* alterations and their clinical implications.

First, our findings are consistent with the hypothesis that high penetrance germline variants influence somatic phenotypes[130]. Germline variants of high penetrance in cancer susceptibility genes tend to acquire biallelic inactivation of the gene harboring them and less likely to have other independent gain-of-function driver alterations compared to tumors retaining heterozygosity in the germline variants. Our data support this observation with the high frequency of second (somatic) hits in *PTEN* (presumably resulting in biallelic inactivation), as well as mutual exclusivity of somatic *PTEN* and *PIK3CA* mutations in PHTS-all BCs, but not in TCGA. Data restricted to BCs arising in the background of pathogenic or likely pathogenic *PTEN* germline variants (Tier 1) further support this hypothesis.

Second, our overall data point to somatic variants in *PTEN* as the predominant, somatic oncogenic drivers in PHTS-derived BCs, especially in TIER-1. This is mainly supported by cDriver ranking *PTEN* as the top driver, based on cancer cell fraction, mutational frequency and functional impacts[98]. Interestingly, there were two PHTS breast samples with two somatic hits in *PTEN*. In the case of PHTS-AE, one somatic variant is a pathogenic frameshift

deletion (NM_000314.4: c.855_856del, p.Glu285Aspfs*12) and the other, a likely pathogenic variant at the last nucleotide of exon 8 (NM_000314.4: c.1026G>C, p.Lys342Asn). In the other case (PHTS-L), both somatic variants are classified as pathogenic. One is a frameshift in exon 8 (NM_000314.4: c.741dup, p.Pro248Thrfs*5) and the other one, a missense mutation in exon 6 (NM_000314.4: c.448G>A, p.Glu150Lys). It is therefore challenging to distinguish which may be the driver mutation in either of these cases, and all of them indeed appear damaging. Importantly, missense mutations in *PTEN* can have dominant negative effects[131], potentially making them as damaging or even more damaging than frameshift truncating or nonsense mutations. Also, it is possible we are detecting somatic variants from different cell populations (i.e., different clonal populations) with two distinct somatic hits in *PTEN* for either of these cases. Further investigations, including gene expression analyses, epigenetic studies and examination of the tumor microenvironment, are warranted for mechanism resolution and to elucidate the functional impacts of the second hit somatic variants in *PTEN*.

Third, our results corroborate that germline *PTEN* variants trend to increased genomic instability, a hallmark of cancer[129], reflecting one of the PTEN's non-canonical roles. Accumulating evidence shows that normal PTEN function are important in the maintenance of genome integrity[75]. Despite lack of statistical significance due to limited sample size, the numerically higher median TMB values in PHTS-all, TIER-1, and TIER-2 BCs compared to TCGA also suggest increased genomic instability due to germline *PTEN* variants and in line

with PTEN's role in maintaining genome stability[75;132]. Whether this may lead to multiple global alterations in the genome warrants further investigation. We also showed through mutational signature analyses that the mechanism of mutational burden appears different between PHTS and TCGA BCs.

Fourth, we observed notable differences in mutational signatures between PHTS-all and TIER-1 BCs compared to TCGA. The sporadic breast cancers from TCGA dataset had a signature associated with mutagens such as tobacco-use, indicating environmental and lifestyle etiologies while PHTS-all and TIER-1 groups lacked this signature (**Figure 3C**, **3D**). Interestingly, one of the SBS clusters in PHTS-all and TIER-1 BCs was SBS1, which is strongly correlated with older age of cancer onset. Knowing the PHTS-all and TIER-1 groups had significantly younger median ages of BC diagnosis, detecting this particular signature may indicate that mutational patterns that usually accumulate with time tend to occur much faster in PHTS-derived BCs, especially in the background of pathogenic germline *PTEN* mutations (Tier 1). Of note, no significant SBS signatures were found for conversions in the TIER-2 group, which is attributable to its small sample size.

Fifth, all the *PIK3CA* somatic mutations detected in PHTS BCs have been reported previously and classified as pathogenic or likely pathogenic (**Table 4**). Furthermore, we found that lower CC scores may predict the presence of somatic *PIK3CA* mutations arising in PHTS-all and TIER-1 BCs, illustrating that germline variants in the setting of phenotypic burden (CC score) may inform

somatic mutations in the BC tissue, which can provide clinically useful information and precision management.

Finally, our genomic data reveal TIER-1 and TIER-2 BCs are two genomically different types of tumors. This provides further evidence to support the different pathogenicity predictions between Tier 1 and Tier 2 germline *PTEN* variants, which are expected to influence the biology of carcinogenesis in BC differently. Further in-depth research is warranted to investigate if this proposed hypothesis has clinical significance in determination of pathogenicity in germline variant calling.

This study has several limitations. First, this study may be limited by a small sample size for PHTS, owing to the difficulty in identifying PHTS clinically. Yet, we have a well-annotated, and homogeneous clinical cohort of PHTS individuals, enabling us to provide clinically useful information for this apparently rare but important population. Second, heterogeneity of each BC sample is an inherent limitation of this study design. Third, not being genome sequencing, our pipeline did not identify all genomic variations such as non-coding ones. Fourth, two different exome sequencing platforms were used for the PHTS series. The older platform, namely the Illumina Somatic Exome protocol, had a lower sequencing coverage goal (20x for tumor) and the read distribution along target regions tended to be less uniform, which may have affected sensitivity of variant calling in certain areas. However, the average coverage for tumor samples was 129x (ranges from 42x to 261x) even for the older platform and as high as 261x (range from 182x to 368x) for the newer platform, which appear appropriate.

Fifth, this study focused on small genomic changes, including single nucleotide variants and indels. Thus, our pipeline did not detect large deletions, insertions, and chromosomal rearrangements. Sixth, this study was not designed to examine epigenetic alterations.

Future directions include expanding on the size of our study cohort for PHTS, examining larger changes in the genome beyond single nucleotide variants and indels, studying non-coding and regulatory regions, gene expression analyses and epigenetic studies. Taking fresh frozen breast tumor samples to examine metabolites and microbiomes may also elucidate potential biological mechanism of carcinogenesis and disease progression. More comprehensive analysis with a multiomic approach would further facilitate the understating of underlying biology of PHTS-derived BC. Since breast cancer samples are highly heterogeneous within samples, a microdissection and/or specialized profiling techniques would be useful to obtain information for each histologically distinct region of the breast.

In conclusion, we demonstrate a characteristic genomic landscape in PHTS-derived BCs. Currently, this entity is treated similarly to sporadic BCs, according to the standard of care in the absence of somatic landscape data in PHTS-derived tumors. Although targeting *PTEN*-associated alterations for therapeutic purposes has tremendous challenges[60;88], our findings call for more targeted, personalized strategies to effectively treat PHTS-derived cancers, a population that will only rise in incidence as clinical genetic testing becomes more widely accessible in the clinic.

CHAPTER III: INTEGRATING SOMATIC COPY NUMBER VARIATIONS AND GENE EXPRESSION IN BREAST CANCERS FROM WOMEN WITH PTEN HAMARTOMA TUMOR SYNDROME

Introduction

Phosphatase and tensin homolog (*PTEN*), a tumor suppressor gene[56], is one of the most frequently somatically altered genes in different malignancies including breast cancer (BC)[57]. *PTEN* hamartoma tumor syndrome (PHTS) encompasses individuals harboring a germline *PTEN* mutation, which causes heritable predisposition to specific cancers including breast, thyroid, kidney, endometrial and colon cancers, and melanoma[50]. PHTS-derived BCs have distinct clinical characteristics compared to sporadic counterparts. Women with PHTS have up to 85% lifetime risk of breast cancer (BC), which is notably higher than that in the general population (12.9% lifetime risk)[50]. Furthermore, women with PHTS have a much younger onset of BC diagnosis, as well as a significantly higher incidence of second primary BC[55].

PHTS-derived BCs are distinct not only at the clinical but also at the molecular and genomic levels. Recently, we found that BCs arising in the setting of PHTS had a distinct somatic mutational landscape compared to that of their sporadic counterparts[133]. We demonstrated that PHTS-derived BCs had a high frequency of somatic second hits to the *PTEN* gene (where the underlining germline *PTEN* mutations represent the first hit), which appeared to be driving carcinogenesis. Furthermore, BCs from PHTS patients with germline pathogenic

or likely pathogenic *PTEN* variants (Tier-1 variants), had much fewer somatic mutations in *PIK3CA* compared to those in TCGA and in PHTS-Tier 2 (variant of unknown significance or likely benign variants) BC. Our findings were consistent with the observation that the nature of the underlying germline mutations in cancer tissues influences somatic phenotypes[130].

BC biology and its genomic landscape are complex and need to be understood in the context of large genomic and functional genomic changes such as somatic copy number variation (CNV) and gene expression differences[134;135]. In this study, we further characterized the somatic landscape of PHTS-derived BCs by examining somatic CNVs and the transcriptome.

Methods

Patients and DNA extraction

The PHTS series includes 44 women with a personal history of BC and who had appropriate consents under the Cleveland Clinic institutional review board protocol 8458 (IRB 8458) at the time of sample acquisition. Original formalin-fixed paraffin-embedded (FFPE) samples representing primary breast carcinoma were obtained from healthcare institutions where the pathology specimens were originally collected. DNA was extracted from the FFPE blocks using QIAamp® DNA FFPE Tissue kit (Qiagen, Venlo, Netherlands). Matched blood-derived DNA originating from lymphoblastoid cell lines from the subjects were obtained from the Genomic Medicine Biorepository at the Lerner Research

Institute of the Cleveland Clinic (Cleveland, OH, USA). Baseline patient characteristics including histologic subtypes, BC-specific tumor markers, age of diagnosis, staging, grade, germline *PTEN* variants and their classifications, were extracted from the Cleveland Clinic Genomic Medicine Institute's relational database and as previously described[133].

Exome sequencing

Next generation sequencing (NGS) was performed on the tumor-blood DNA pairs using the Illumina HiSeq platform at the Broad Institute of MIT and Harvard University. The raw data were quality controlled, aligned and sorted by the computational pipeline at the Broad Institute to generate binary alignment map (BAM) files for tumor and blood samples separately. Detailed methods for DNA sample processing and sequencing methodology were previously described[133].

Copy number variation analysis

With the WES data from 44 PHTS-derived BC samples and 558 sporadic BC samples from TCGA as input, the segmentation and raw copy number data were obtained using FACETS (version 0.5.6), an open-source tool to analyze allele-specific copy number variations[120]. The critical value (cval) was specified at 50 to create an input for gistic2 (version 6.15.28)[136], which identifies significantly recurrent copy number alterations in the somatic genome. We used the following setting: a confidence interval of 95%, q-value of 0.01, amplification threshold of 0.3, and deletion threshold of -0.3. For other parameters, we used the default

setting specified by gistic2. We applied the same CNV algorithm to the raw TCGA sporadic BC dataset to make a head-to-head comparison with our PHTS series data.

Patients and RNA extraction

FFPE tissue samples were available from a subset of the PHTS BC series (n=29). RNA was extracted from the available FFPE blocks using AllPrep® DNA/RNA FFPE kit or RNeasy FFPE kit (Qiagen). RNA concentration was measured with the Qubit Fluorometer dsDNA HS (High Sensitivity) Assay kit (Thermo Fisher Scientific, Waltham, MA, USA). While the ideal RNA concentration for sequencing library preparation was 30–40 ng/mL, the range of RNA concentrations of submitted samples was 57.6–540 ng/mL. The range of 280/260 ratio was 1.81 to 2.07. The extracted RNA was sent to the Genomics Core of the Department of Genetics and Genome Sciences at Case Western Reserve University (Cleveland, OH, USA) for library construction. The constructed RNA libraries were then sent to the Genomics Core at the Cleveland Clinic Lerner Research Institute (Cleveland, OH, USA) for RNA sequencing.

RNA-Seq library preparation and sequencing

The SMARTer Stranded Total RNA-Seq Kit v2 Pico Input Mammalian from Takara Bio USA (protocol 050619) was used to prepare RNA-Seq libraries. The total RNA input was adjusted to 100ng in 8ul of nuclease-free water. Since FFPE samples intrinsically have highly degraded RNA, cDNA synthesis was performed

without fragmentation. Subsequent PCR steps utilized the indexes from the SMARTer RNA Unique Dual Index Kit – 24U (634451). Ribosomal cDNA was depleted, and the final amplification included 13 cycles of PCR. Samples were purified with AMPure beads and eluted in 18 µl of 5mM Tris Buffer. Final QC included running samples (diluted 1:1 in water) on the HSD1000 tape on the Agilent TapeStation and obtaining a Qubit reading (Thermo Fisher Scientific).

The constructed libraries were sequenced on an Illumina NovaSeq 6000 using an S2 flow cell, where dual-indexed paired-end 151 bp sequencing was accomplished. Sequencing data were demultiplexed using *bcl2fastq* and FastQC reports were generated to evaluate the sequence quality of each sample.

Differentially expressed gene analysis

Based on the FastQC report of the original FASTQ files generated by RNA sequencing, we noted that the first three base pairs in reverse reads (R2) consistently had low quality scores at the 5' end. Thus, the first three base pairs at the 5' end in R2 FASTQ files were clipped using trimmomatic (version 0.39)[137]. Adaptor and ribosomal sequences were trimmed off using BBmap (version 37.96)[138]. The optimized FASTQ files were then aligned to hg38 using STAR (version 2.7.8)[139].

FastQC reports were again obtained on STAR aligned FASTQ files. Five out of 29 samples had less than 50% uniquely mapped reads, and were excluded from differentially expressed gene analyses. We analyzed the 24 samples which passed quality control using DESeq2 (version 1.34.0) to identify

differentially expressed genes with statistical significance[140], defined as a false discovery rate (FDR) < 0.05 and a log2-fold change > +/- 1 (greater than 2 for overexpression and less than -2 for underexpression).

The hierarchical clustering heatmap was created using pheatmap (version 1.0.12)[141] and the volcano plot was created using EnhancedVolcano (version 1.12.0)[142], using R (version 4.1.2).

CNV and transcriptome correlational analysis

We performed correlational studies to examine which cis-genes are correlationally expressed with the chromosomal peaks detected by gistic2. For each sample, the log2 fold change raw values from DESeq2 were tested for the actual copy change values from gistic2. Pearson correlation analysis[143] was used for genes with normally distributed log2 fold changes, and Spearman correlation[144] for those with non-normal distributions. Normalization test was performed using D'Agostino-Pearson omnibus normality test, Anderson-Darling test, Shapiro-Wilk normality test and Kolmogrov-Smirnov normality test with the default setting with alpha of 0.05 on GraphPad Prism version 9.0 (GraphPad Software, San Diego, CA, USA). Bonferroni correction was performed to identify statistically significant genes associated with the peak regions containing multiple genes.

Breast cancer-associated genes

For targeted analysis, we aggregated lists of genes associated with BC[133]. The selected genes were chosen from the TCGA BC publication[32], NCCN Genetic/Familial High-Risk Assessment: Breast and Ovarian guidelines (version 1.2022-August 11, 2021), 22 previously reported gold standard (GS) genes for BC[98], preliminary BC susceptibility genes and targetable BC-associated genes from the literature[48;98-116]. A total of 84 BC-associated genes were included in the final list (**Table 3**).

Pathway analysis

Output from DESeq2, including HGNC gene ID, log2 fold changes and adjusted p-values, was uploaded into the Ingenuity Pathway Analysis (IPA) software (QIAGEN, Venlo, Netherlands). The data were then subjected to functional annotations and canonical pathway analyses. The IPA's Core Analysis workflow was performed using default parameters. For Benjamini-Hochberg (B-H) correction, the score cut off (A-log or B-H p-value) of greater than 1.3 was used.

CIBERSORT and diversity analyses

The TCGA Breast Invasive Carcinoma (BRCA) RNA-Seq dataset was downloaded using TCGAbiolinks package[145]. Data retrieval was performed by the three main TCGAbiolinks functions: GDCquery, GDCdownload and GDCprepare. The raw feature count matrix was converted to transcripts per million (TPM) and merged with PHTS data. The merged TPM matrix was

processed for differential abundance analysis using the random-forest algorithm, implemented in the DAtest package (https://github.com/Russel88/DAtest/wiki/usage#typical-workflow). Briefly, the performance of differential abundance methods was compared with False Discovery Rate (FDR), Area Under the (Receiver Operator) Curve (AUC), Empirical power (Power), and False Positive Rate (FPR). Based on the DAtest's benchmarking, we selected random forest as the method of choice to perform differential abundance analysis. We assessed the statistical significance ($P < 0.05$) throughout, and whenever necessary, we adjusted p-values for multiple comparisons according to the Benjamini-Hochberg method to control false discovery rate while performing multiple testing on gene abundance according to sample categories. We used CIBERSORT[146] for performing RNA-Seq deconvolution analysis and estimating immune cell fractions in our bulk RNA-Seq data.

Statistical analysis

Statistical analyses were performed with GraphPad Prism version 9.0 (GraphPad Software, San Diego, CA, USA), except for statistical analyses incorporated in maftools (version 2.10.0).[118] P-values < 0.05 were considered statistically significant unless otherwise stated.

Results

Notable copy number variants (CNVs) in PHTS-derived BCs which are distinct from the sporadic BCs from TCGA.

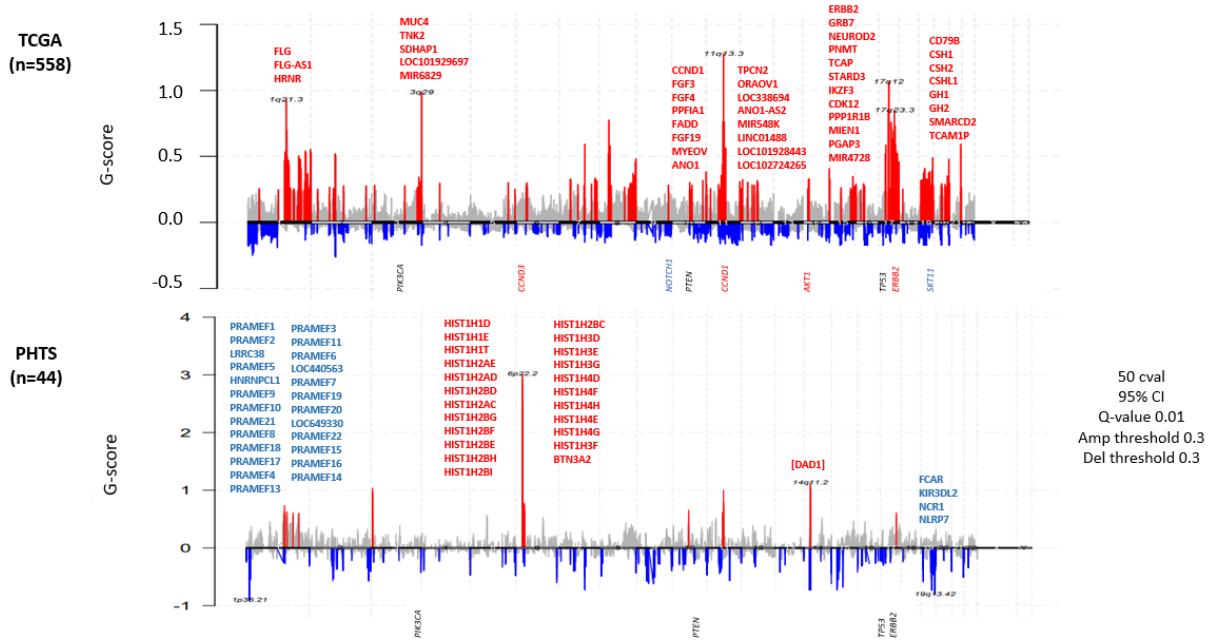
We identified seven significant amplification peaks and 46 significant deletion peaks in PHTS-derived BCs (**Figure 5A, Table 10, 11**). In TCGA BCs, there were 37 amplification peaks and 63 deletion peaks. Four out of seven CNV amplifications (3p26.1, 6p22.2, 10q21.2, 11q13.1) are present in PHTS-derived BC samples but not in sporadic BC samples from TCGA. The most significant peak in this group is at 6p22.2 (**Figure 5B**), which was absent in TCGA. This region contains multiple histone-related genes including *HIST1H1*, *HIST1H2*, *HIST1H3*, and *HIST1H4* families. Nine out of 36 samples (25.0%) with significant amplifications at 6p22.2 had somatic *PTEN* variants which were distinct from their respective germline *PTEN* variants, while only one out of eight samples (12.5%) without a 6p22.2 CNV amplification peak had a somatic *PTEN* hit. This difference trended towards but did not reach significance (Odds ratio [OR] 2.33, 95% CI 0.29 to 28.9, p=0.66).

For deletion peaks, 28 out of 46 regions were present in PHTS BCs but absent in TCGA BCs (**Table 11**). No amplification or deletion peaks in PHTS-derived BCs contained any of the 82 BC associated genes (**Table 10**, **11**) including *ERBB2*, *EGFR*, *PTEN*, and *TP53*. In contrast, amplification peaks containing *CCND3* (6p21.1), *CCND1* (11q13.3), *AKT1* (14q32.33), and *ERBB2* (17q12), and deletion peaks containing *NOTCH1* (9q34.3) and *STK11* (19p13.3) were identified in TCGA BCs.

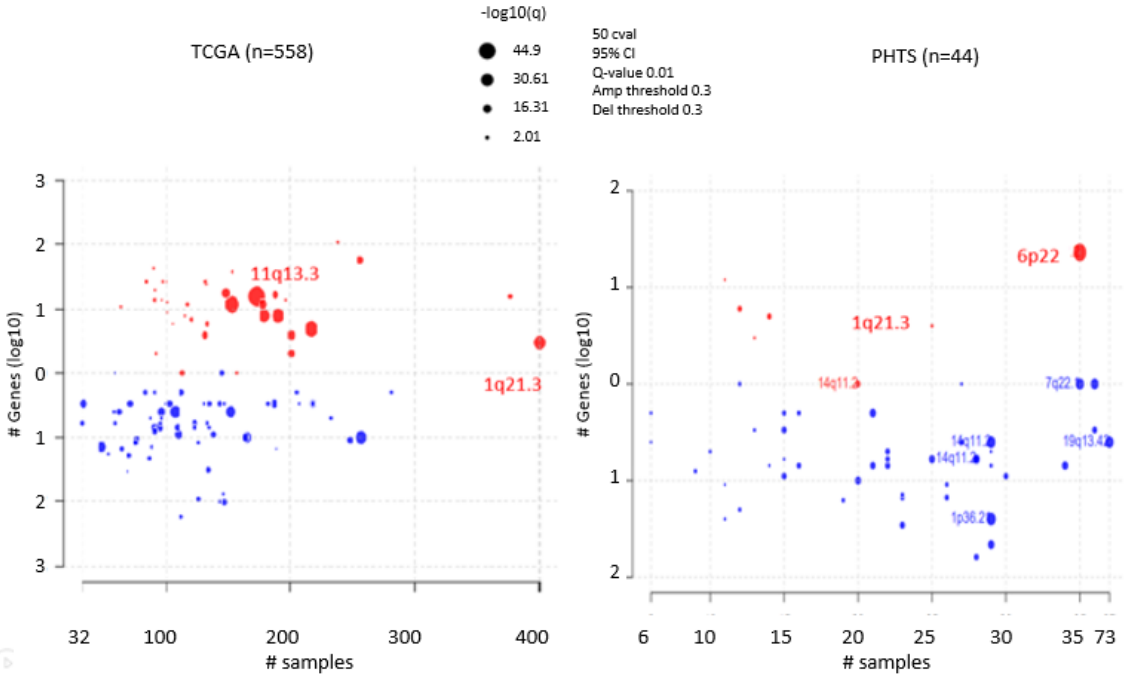
Overall, there were three common amplification peaks between PHTS-derived and TCGA BCs (**Figure 6A**), while there were 14 common deletion peaks shared by the two groups (**Figure 6B**).

Figure 5: Recurrent CNV peaks and altered cytobands in PHTS-derived BCs and sporadic BCs from TCGA

A.



B.

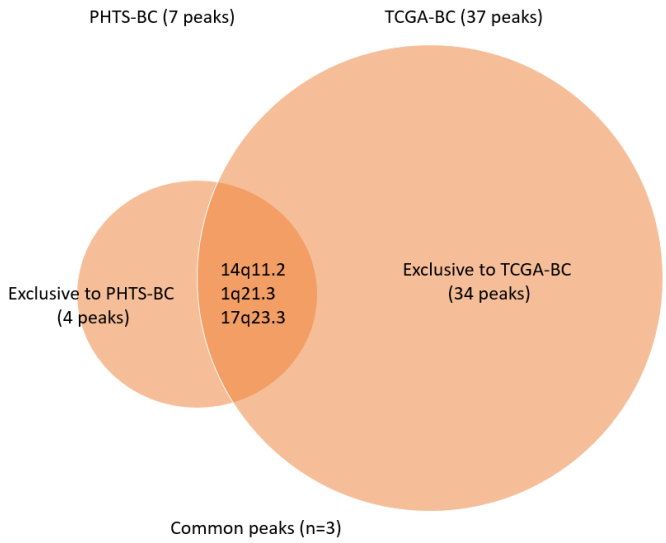


A. Genome plots showing recurrent CNV peaks identified by gistic2 in samples from TCGA (top) and from PHTS-derived BCs (bottom). Red peaks represent recurrent amplification and blue, deletion peaks. Genes contained in some of the significant peak regions are shown. Positions of BC-associated genes are shown at the bottom of each plot, shown in red if an amplification is present and in blue if a deletion is observed.

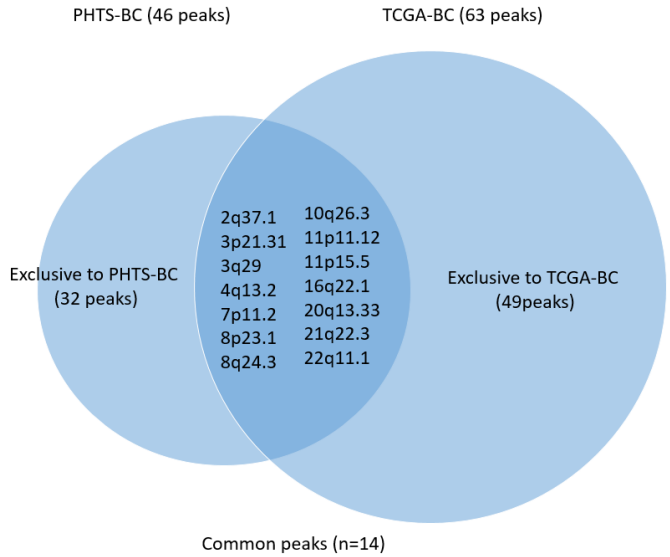
B. Gistic bubble plots showing significantly altered cytobands in BC samples from TCGA (left panel) and in PHTS-derived BCs (right panel). The X-axis represent the number of samples which had CNV alterations and the Y-axis, the number of genes each bubble contains. The bubble size is according to -log10 transformed q values.

Figure 6: CNV peak comparison between PHTS-derived and TCGA BCs showing common and unique peaks

A.



B.



A. Venn diagram showing three common amplification peaks between PHTS-derived BCs and TCGA BCs.

B. Venn diagram showing 14 common deletion peaks between PHTS-derived BCs and TCGA BCs.

Table 10: Amplification peaks detected by gistic2 in PHTS-derived BCs

	Amplification peak ***(no. of genes in the peak region)***	***Genes in the region*** ***(unadjusted p-value)***
1	1q21.3 (n=4)	*ENSA** (p=0.012)
2	3p26.1	-
3	6p22.2 (n=11)	*HIST1H2BI** (p=0.004)
4	10q21.2 (n=3)	*ARID5B* (p=0.024)
5	11q13.1	-
6	14q11.2 (n=1)	*[DAD1]** (p=0.019)
7	17q23.3 (n=24)	-

The number of genes found in each peak region is shown in parenthesis in the first column. Genes in regions whose expression were statistically correlated with the peak intensities are shown in the second column. The p-values shown in column 2 are unadjusted. After Bonferroni correction, based on the number of genes tested within each region, the genes indicated with an asterisk remained statistically significant.

Table 11: Deletion peaks detected by gistic2 in PHTS-derived BCs

	Deletion peak ***(no. of genes in the peak region)***	***Genes in the region*** ***(unadjusted p-value)***
1	1p36.21 (n=19)	*PRAMEF20* (p=0.046)
2	1q44 (n=20)	*OR2M4* (p=0.041), *OR2M3* (p=0.0039), *OR2T6* (p=0.022)
3	2p11.1	-
4	2q37.1	-
5	3p21.31 (n=5)	*MIR-711* (n=0.029)
6	3q29	-
7	4p16.1	-
8	4q13.2	-
9	5p15.33	-
10	5q23.3	-
11	5q35.3 (n=3)	*SCGB3A1* (p=0.034)
12	6p21.32	-
13	6q25.3	-
14	7p13 (n=3)	*POLR2J4** (p=0.013)
15	7p11.2	-
16	7q22.1	-
17	8p23.1	-
18	8q24.3 (n=4)	*LY6E* (n=0.016)
19	9p13.3	-
20	9p11.2 (n=35)	*ANKRD20A1** (p=0.0006), *CNTNAP3B** (n=0.0011), *SPATA31A1* (n=0.0015)
21	9q34.3	-
22	10p12.1 (n=5)	*GAD2* (p=0.018), *FAM238B* (p=0.018)

23	10p11.1	-
24	10q26.3 (n=9)	*KND1* (p=0.0055)
25	11p15.5 (n=3)	*MUC5B** (p=0.0007)
26	11p11.12 (n=59)	*ORSM11* (p=0.018), *OR5J2* (p=0.040)
27	11q14.3 (n=16)	*FOLH1B* (p=0.012)
28	12p13.33	-
29	12q13.2 (n=16)	*OR10P1* (p=0.021)
30	13q12.11 (n=6)	*TPTE2* (p=0.025), *ANKRD26P3* (p=0.032)
31	14q11.2 (peak A)	-
32	14q11.2 (peak B)	-
33	15q11.1 (n=1)	*CHEK2P2** (p=0.035)
34	16p13.11	-
35	16q22.1 (n=5)	ARRS1 (p=0.011)
36	16q22.2	-
37	17p13.1 (n=7)	*MYH1** (p=0.006), *MYH4* (p=0.029)
38	17q21.2	-
39	18p11.23	-
40	19p13.2 (peak A) (n=7)	*OR7G1** (p=0.006), *MBD3L1* (p=0.035)
41	19p13.2 (peak B)	-
42	19q13.33	-
43	19q13.42 (n=4)	*NCR1* (p=0.047)
44	20q13.33	-
45	21q22.3	-
46	22q11.1	-

The number of genes found in each peak region is shown in parenthesis in the first column. Genes in regions whose expression were statistically correlated with the peak intensities are shown in the second column. The p-values shown in column 2 are unadjusted. After Bonferroni correction, based on the number of genes tested within each region, the genes indicated with an asterisk remained statistically significant.

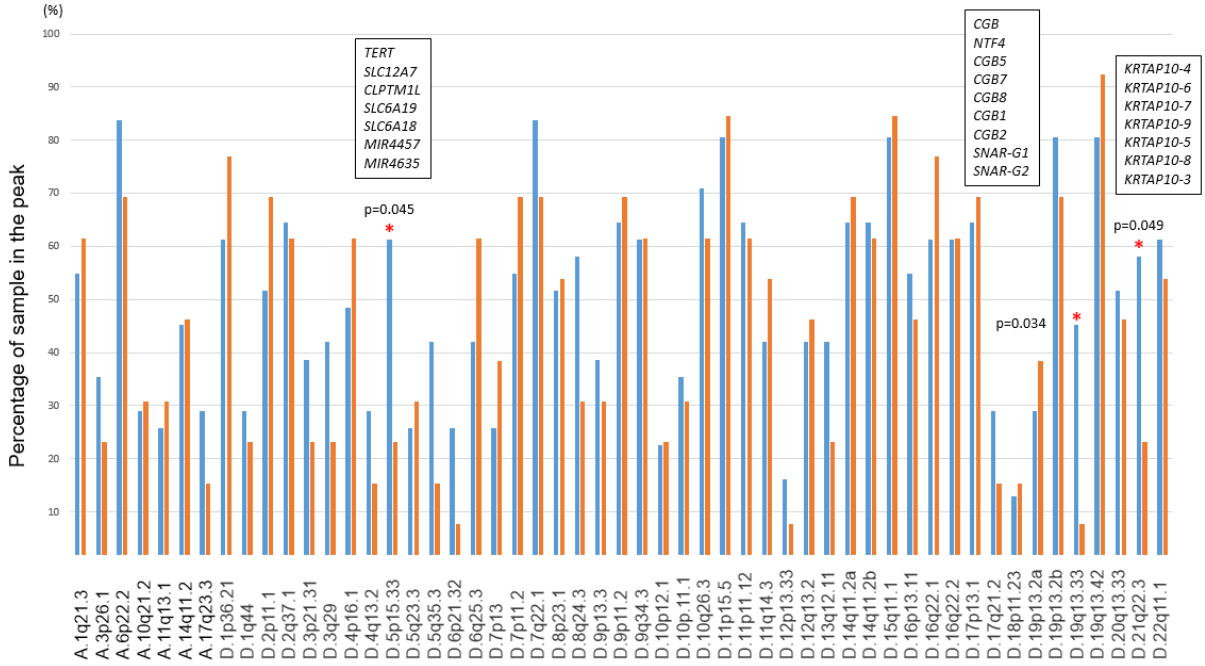
Several genes within amplified or deleted chromosomal regions are correlationally expressed in PHTS-derived BCs

In order to examine if any of the genes present within the identified CNV peaks are overexpressed for amplification peaks (7 peaks) or underexpressed for deletion peaks (46 peaks), we performed correlational studies between the height intensity of each peak and log2-fold expression changes (presence vs absence of the peak) for each gene in the peak regions. For significant peaks in PHTS-derived BCs, we found some genes are correlationally expressed and remained statistically significant after Bonferroni correction. For the amplification peak at 1q21.3, for instance, *ENSA* was correlationally over-expressed. There were no genes identified in the amplification peak at 14q11.2, but the closest gene by distance called *DAD1* was found to be correlationally overexpressed with this peak. For deletion peaks, the following genes were found to be correlationally underexpressed at the corresponding regions: *POLR2J4* at 7p13; *ANKRD20A1* and *CNTNAP3B* at 9p11.2; *MUC5B* at 11p15.5; *CHEK2P2* at 15q11.1; *MYH1* at 17p13.1; and *OR7G1* at 19p13.2 (**Table 11**). Among genes tested in the 6p22.2 region, *HIST1H2BI* was correlationally overexpressed by independent Fisher's exact test (p=0.0043). However, this finding did not remain statistically significant after Bonferroni correction (**Table 11**).

Several CNV peaks are more significant in Tier-1 compared to Tier-2 PHTS-BCs

In our previous work, we studied two groups of tumors classified by the pathogenicity5 of the underlying germline *PTEN* variants. Tier-1 germline *PTEN* variants (n=31) are classified as pathogenic or likely pathogenic, while Tier-2 germline *PTEN* variants (n=13) as variants of unknown significance (VUS, n=8) or likely benign (n=5). To further characterize the differences between these two types of tumors, we examined which CNV peaks are more significantly amplified or deleted using Fisher's exact test. Tier-1 BCs had three more significant deletion peaks compared to Tier-2 BCs, namely at 15p15.33, 19q13.33, and 21q22.3 (**Figure 7**). None of the genes identified in these peaks have correlationally expressed transcripts.

Figure 7: Comparison of CNV peak occurrences showed three chromosomal regions more frequently deleted in Tier-1 over Tier-2 derived BC samples



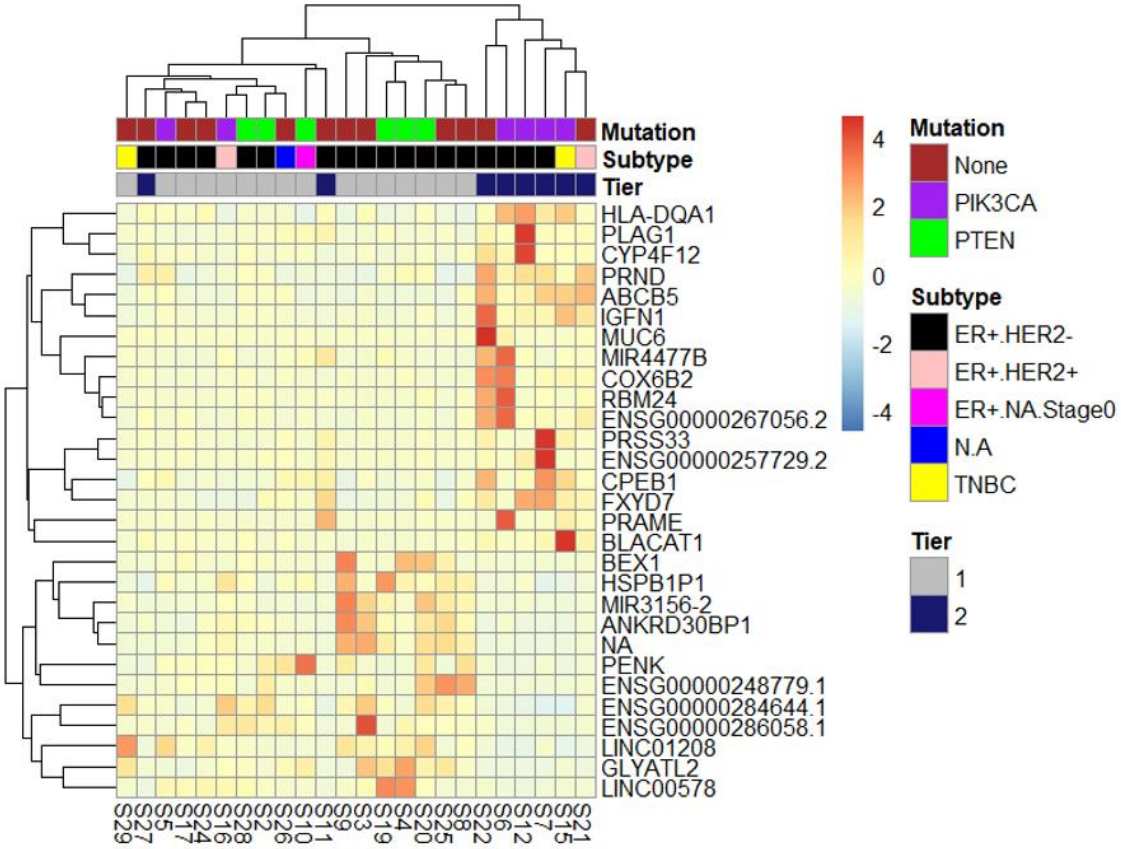
Bar graph showing amplification (A.) and deletion peaks (D.) at each significant CNV peak on the X-axis. The Y-axis represents the percentage of samples with statistically significant peaks present in each region. Blue bars represent Tier-1 derived BC samples, orange, Tier-2. Statistically significant different regions between Tier-1 and Tier-2 BCs are shown with a red asterisk along with the p-value calculated by Fisher's exact test. The genes contained in the regions with statistically significant tests are listed in the white boxes.

Transcriptome analysis reveals several differentially expressed genes in Tier-1 compared to Tier-2 PHTS-BCs

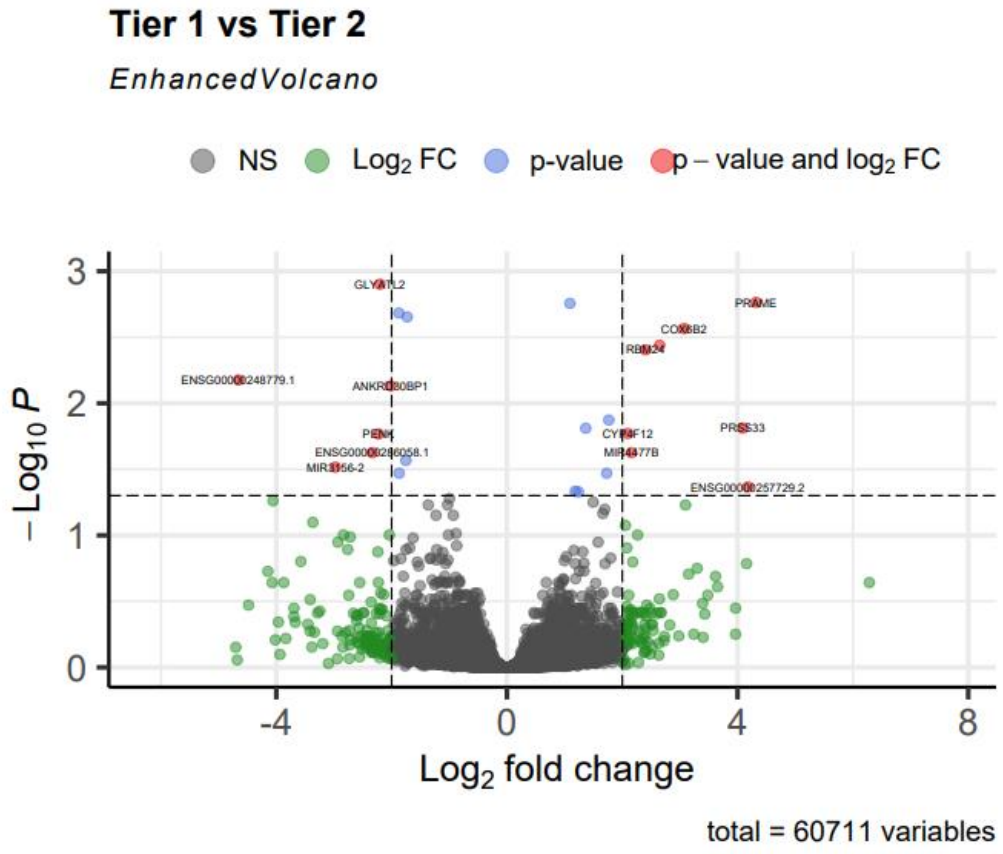
We further examined the differences between Tier-1 and Tier-2 derived BCs at the transcriptome level by performing differential gene expression analysis. A hierarchical clustering heatmap showed Tier-1 and Tier-2 derived BCs clustered into two distinct patterns of differentially expressed genes (**Figure 8A**). We identified a total of 18 differentially expressed genes with 10 overexpressed (*MUC6*, *PRAME*, *RP11_788H181*, *PRSS33*, *COX6B2*, *AC0053364*, *RBM24*, *IGFN1*, *mir-4477*, and *CYP4F12*) and 8 underexpressed (*RP11_53O192*, *BEX1*, *mir-3156*, *ANKRD30B*, *FAR2P1*, *PENK*, *GLYATL2*, and *ANKRD30BP1;* **Figure 8B** and **Table 12).** There were two Tier-2 derived BC samples which clustered among Tier-1 derived BC samples. Overall, we found no clear association between gene expression differences and BC subtypes or the presence of somatic *PTEN* or *PIK3CA* variants (**Figure 8B**).

Figure 8: RNA-seq data showing two distinct groups in PHTS-derived BC samples, identifying alpha-tocopherol degradation as a significant biological pathway in Tier-1 PHTS-BC

A.



B.



<u>A.</u> Heatmap of hierarchical clustering based on 28 differentially expressed (DE) genes, showing the PHTS-derived BC samples cluster into two groups, namely Tier-1 and Tier-2 (log2 fold change +/- 1, p < 0.05). The X-axis lists the sample ID and right Y-axis shows the DE gene IDs included in this analysis.

<u>B.</u> Enhanced volcano plot of RNAseq transcriptome data showing differentially expressed genes in Tier-1 BCs compared to Tier-2 (log2 fold change +/- 1, p < 0.05). The x-axis shows the magnitude of change (2 fold change) and the Y-axis, statistically significance in -log10P.

Table 12: Top differentially expressed genes detected by DESeq2 (Tier-1 compared to Tier-2 BC)

Overexpressed genes	**Fold changes**
MUC6	5.065
PRAME	4.318
RP11_788H181	4.178
PRSS33	4.092
COX6B2	3.071
AC0053364	2.648
RBM24	2.402
IGFN1	2.178
mir-4477	2.161
CYP4F12	2.096

Underexpressed genes	**Fold changes**
RP11_53O192	- 4.657
BEX1	- 4.139
mir-3156	- 2.975
ANKRD30B	- 2.910
FAR2P1	- 2.338
PENK	- 2.227
GLYATL2	- 2.204
ANKRD30BP1	- 2.018

We chose false discovery rate (FDR) threshold of 0.05 and log2 fold change threshold of +/- 1.

Pathway analysis based on the differentially expressed genes reveals ɑ-tocopherol degradation to be a statistically significant canonical pathway in Tier-1 over Tier-2 BCs

We then examined which biological pathways are characteristic of Tier-1 BCs compared to Tier-2 BCs. Using the transcriptomic data from RNA sequencing as input, Ingenuity Pathway Analysis (IPA) revealed three enriched canonical pathways by Fisher's exact test for Tier-1: 1) alpha-Tocopherol Degradation, 2) BEX2 Signaling Pathway, and 3) Oxidative Phosphorylation. After Benjamini-Hochberg correction, the first pathway, alpha-Tocopherol Degradation, remained statistically significant (P=0.037).

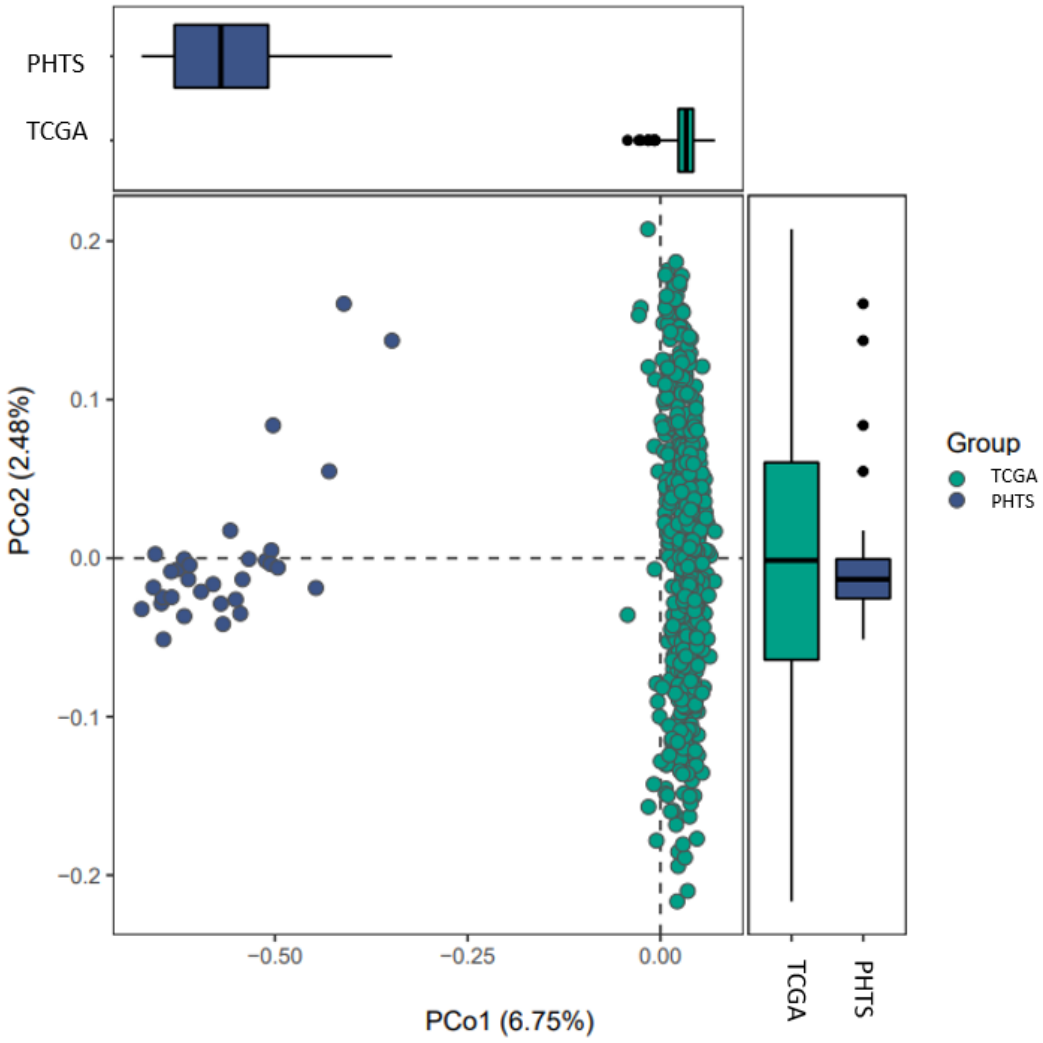
Immune Cell Population Characterization reveals PHTS-derived BCs have immune suppressed and immunotherapy resistant phenotypes compared to TCGA-BCs

In order to characterize patterns of immune cell populations infiltrating or surrounding breast carcinomas, we used CIBERSORT[146] to impute immune cell compositions in PHTS-derived and TCGA-derived BCs. Beta-clustering based on fractions of each cell population showed that the PHTS BC group is distinct from the sporadic TCGA BC counterparts (**Figure 9A**). We identified certain immune cell populations whose proportions are significantly increased in PHTS-derived BCs (t-test $p<0.05$), namely naïve B cells, M0 macrophages, M2 macrophages, resting mast cells, monocytes, activated NK cells, and regulatory T cells (**Figure 10**, **Table 13**). In contrast, the TCGA BCs had significantly greater predicted

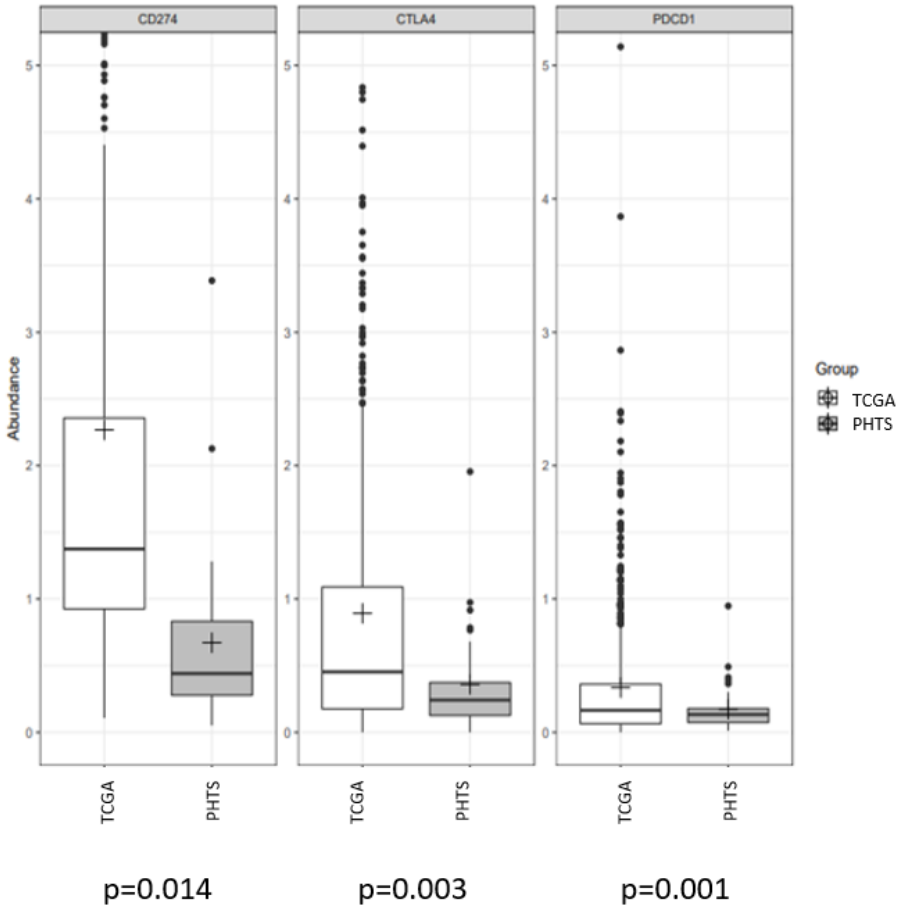
proportions of cell populations including dendritic cells (resting), eosinophils, M1 macrophages, mast cells, CD4 memory activated T cells, CD8 T cells, follicular helper T cells, and gamma delta T cells (**Table 13**). Relatedly, we compared the gene abundance between PHTS-derived BCs and TCGA BCs for PD-L1 (*CD274*), *CTLA4*, and PD-1 (*PDCD1*). The TCGA BCs had significantly increased (Fisher's exact test $p<0.05$) abundance in these genes (*CD274*, p=0.014; *CTLA4*, p=0.003; *PDCD1*, p=0.001; **Figure 9B**).

Figure 9: Immune cell population characterization and immunotherapy target gene abundance in PHTS-derived BCs versus sporadic BCs

A.



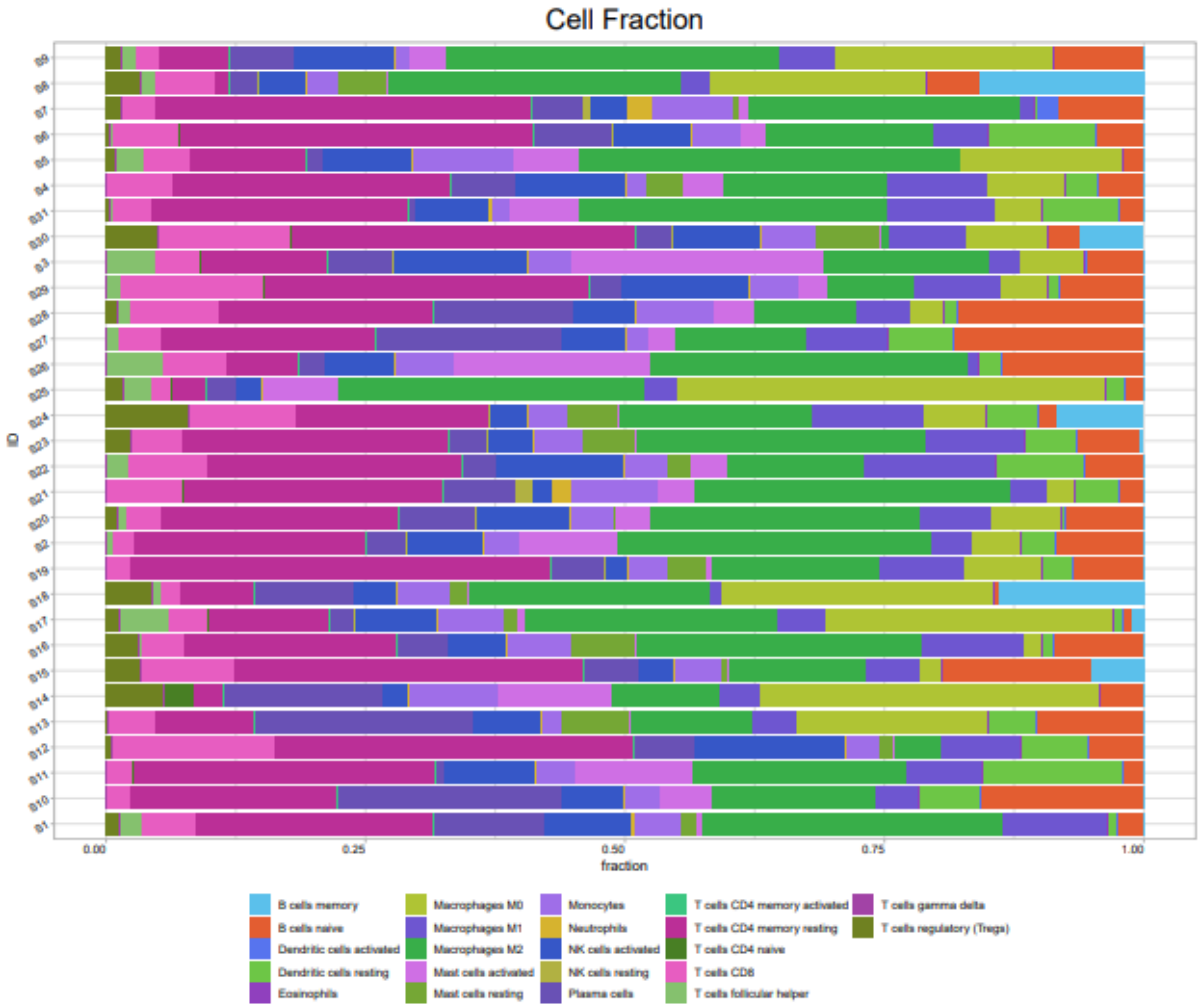
B.



A. Box plot showing beta-clustering based on fractions of each cell population, revealing that the PHTS BC group is distinct from the sporadic TCGA BC counterparts.

B. Box plot showing the gene abundance comparison between PHTS-derived BCs and TCGA BCs by Fisher's exact test for PD-L1 (CD274), CTLA4, and PD-1 (PDCD1). TCGA BCs had significantly increased abundance of these genes (CD274, p=0.014; CTLA4, p=0.003; PDCD1, p=0.001).

Figure 10: CIBERSORT reveals heterogeneity of immune cell populations in PHTS-derived BC samples



CIBERSORT plots showing cell compositions in PHTS-derived BC samples. X-axis shows the estimated percentage of each cell population. Each row represents a sample with the corresponding sample ID.

Table 13: CIBERSORT reveals significant differences in immune cell populations between PHTS-derived and TCGA BCs

Significantly increased cell population	*Increased group*	*P-value*
B cells naïve	*PHTS*	*0.0003*
Dendritic cells resting	*TCGA*	*<0.0001*
Eosinophils	*TCGA*	*<0.0001*
Macrophages M0	*PHTS*	*<0.0001*
Macrophages M1	*TCGA*	*<0.0001*
Macrophages M2	*PHTS*	*<0.0001*
Mast cells	*TCGA*	*0.0004*
Mast cells resting	*PHTS*	*0.0014*
Monocytes	*PHTS*	*<0.0001*
Neutrophils	*TCGA*	*(0.373)*
NK cells activated	*PHTS*	*<0.0001*
T cells CD4 memory activated	*TCGA*	*<0.0001*
T cells CD8	*TCGA*	*0.0011*
T cells follicular helper	*TCGA*	*<0.0001*
T cells gamma delta	*TCGA*	*<0.0001*
T cells regulatory (Tregs)	*PHTS*	*<0.0001*

Comparison between PHTS-derived BCs and sporadic BCs from TCGA in the proportion of each cell population by t-test. For each cell population, the group (PHTS vs TCGA) which had increased proportion compared to the other group is indicated in column 2, along with the P-value in column 3.

Discussion

In this study, we identified distinct somatic CNVs in PHTS-derived BCs compared to sporadic BCs. Overall, we observed notable heterogeneity across PHTS BC samples, which is consistent with the nature of BC biology in general. Our data point to several key findings, which help further characterize PHTS-derived BCs and offer insights into the biology of BCs arising in the setting of germline *PTEN* variants.

Our data revealed somatic CNVs in PHTS-BC which are distinct from those in sporadic BCs from TCGA. The most significant amplification peak was at 6p22.2, which was not observed in TCGA. This peak contains several histone-related genes, including *HIST1H2BI*, which was found to be correlationally expressed with the copy number change. The lack of statistical significance in correlational expression of the other histone genes may be due to the limited sample size. Furthermore, although not statistically significant, there was a higher proportion of somatic *PTEN* variants in the samples which exhibited significant amplification at 6p22.2. The PTEN protein is known to interact with histone H1 to maintain chromatin organization and integrity[76]. Importantly, we previously demonstrated that the tumor mutational burden is increased in PHTS-derived BCs compared to sporadic BCs, which supports genomic instability as an important component of BC biology in PHTS[133]. When PTEN dysfunction negatively affects chromatin stability, this leads to dysregulated gene expression[76]. We therefore speculate that the significant 6p22.2 amplification peak may represent a feedback loop to compensate for the compromised

genome integrity and increased instability. In such a case where PTEN is severely dysfunctional, therapeutic agents targeting DNA damage may be useful, including DNA intercalating agents such as doxorubicin and poly(ADP-ribose) polymerase (PARP) inhibitors[147].

We found that for the amplification peak at 1q21.3, the cis-gene alpha-endosulfine gene (*ENSA*) was found to be correlationally over-expressed in PHTS relative to TCGA BCs. This peak was also present in BC samples from TCGA and is known to be a recurrent amplification in BC. *ENSA* has been found to be highly expressed in triple negative breast cancer (TNBC) and associated with poor survival in this group[148]. Upregulation of *ENSA* has been shown to promote tumor growth by regulating cholesterol biosynthesis[148]. This may be one of the common biological mechanisms for carcinogenesis shared by PHTS-derived and sporadic BCs.

For deletion peaks, it is unknown how altered gene expression due to copy number deletion may contribute to carcinogenesis in BC. RNA Polymerase II Subunit J4, Pseudogene (*POLR2J4*) at 7p13 and Olfactory Receptor Family 7 Subfamily G Member 1 (*OR7G1*) at 19p13.2 are reported to be associated with non-breast cancers[149;150]. Contactin Associated Protein Family Member 3B (*CNTNAP3B*) at 9p11.2 has been reported to be overexpressed in atypical hyperplasia of the breast[151]. Mucin 5B, Oligomeric Mucus/Gel-Forming (*MUC5B*) at 11p15.5 was also found to be correlationally expressed with the peak at 11p15.5. This peak was also identified in the TCGA group. Previous studies have shown *MUC5B* expression was increased in BCs compared to normal breast

epithelium[152], and that *MUC5B* expression is associated with aggressive behavior of BC cell lines[153]. There is currently little evidence in the literature describing checkpoint kinase 2 pseudogene 2 (*CHEK2P2*) at 15q11.1, Myosin Heavy Chain 1 (*MYH1*) at 17p13.1 *or* Ankryn Repeat Domain 20 Family Member A1 (*ANKRD20A1*) at 9q21.11 as significant genes in BC. Whether and how these copy number deletions and gene expression differences play a role in breast carcinogenesis in PHTS warrants further investigation.

There are two biologically distinct groups of PHTS-derived BCs based on the pathogenicity of the underlining germline *PTEN* variants: 1) Tier-1 variants are classified as pathogenic or likely pathogenic; and 2) Tier-2 variants, as variants of unknown significance (VUS) or likely benign[133]. Our previous exome sequencing data revealed that Tier-1 and Tier-2 derived BCs are different at the genomic level. This finding was further supported by transcriptomic analysis data, where Tier-1 and Tier-2 BCs clustered separately. Some genes overexpressed in Tier-1 BCs relative to Tier-2 BCs may contribute to BC tumorigenesis and progression. For instance, some members of the mucin protein family, have been shown to be highly expressed in mucinous BC and associated with negative estrogen receptor (ER) status[154]. Expression of PReferentially expressed Antigen of Melanoma (*PRAME*) was previously shown to correlate with poorer clinical prognosis, including higher rates of distant metastases and decreased overall survival in BC[155]. Additionally, the PRAME protein has been investigated as a potential immunotherapy target[156;157]. Serine Protease 33 (*PRSS33*) and Cytochrome C oxidase subunit 6B2 (*COX6B2*) are not well-characterized in BC

but their expression is associated with other types of cancer[158;159]. Similarly, other overexpressed genes including RNA binding motif protein 24 (*RBM24*), Immunoglobulin-like and fibronectin type III domain containing 1 (*IGFN1*), and Cytochrome P450 family 4 subfamily F member 12 (*CYP4F12*) may have biological contributions to tumorigenesis in Tier-1 PHTS BC but their exact roles are unclear[160-164]. Additional investigation of their association with BC is warranted.

The contribution of Tier-1 underexpressed genes to tumorigenesis and disease progression is even less clear. Little is known about the molecular functions of brain expressed X-linked 1 (*BEX1*) and its exact role in tumorigenesis is still under debate[165;166]. Proenkephalin (*PENK*) is one of the genes encoding for endogenous opioid precursors[167]. Interestingly, downregulation of *PENK* is reported to be associated with defects in cell motility and abnormal adhesion in brain metastasis from BC.[168] Glycine-N-acyltransferase like 2 (*GLYATL2*) is a glycine conjugating enzyme with functions implicated in barrier function and immune response[169]. Very little is known about any association between ankyrin repeat domain 30B (*ANKRD30B*) and any type of cancer.

Although some of the identified differentially expressed genes have been implicated in BC development and progression, we do not think one or just a few genes drive tumorigenesis in PHTS-derived BCs. Thus, we performed pathway analysis, which revealed ɑ-tocopherol degradation to be a significantly impacted canonical pathway in Tier-1 versus Tier-2 BCs. Also known as vitamin E, ɑ-

tocopherol is an antioxidant[170], and an animal and cell-based study has shown that vitamin E may increase PTEN and p53 levels in the rat prostate[171]. Furthermore, in a subtype of Cowden syndrome with no germline *PTEN* mutations but with germline *SDHx* variants, vitamin E appears to protect from oxidative stress and potentially suppresses tumorigenesis[172]. We hypothesize that vitamin E plays an important role in suppressing the development of cancer in cells with dysfunctional PTEN-related pathways. Being a key pathway in Tier-1 BCs, vitamin E degradation may explain the more penetrant nature of pathogenic germline *PTEN* variants including within the Tier-1 BCs due to enhanced elimination of vitamin E, which is supposed to protect cells from carcinogenic oxidative damage. This hypothesis is worth experimentally testing, including in relevant *Pten* animal models.

The immune landscape infiltrating or surrounding breast carcinoma appears distinct between PHTS and TCGA. Overall, beta-clustering revealed these two groups to be significantly different from one another in cell composition, with certain immune cell populations predicted to be significantly increased in proportion either in PHTS or TCGA BCs. More specifically, immune cell populations which are either inactive or suppressive (naïve B cells, M0 macrophages, M2 macrophages, resting mast cells, monocytes, and regulatory T cells) are increased in PHTS BCs. Furthermore, genes encoding immune checkpoint pathways, including PD-L1 (*CD274*), CTLA4, and PD-1 (*PDCD1*), were found less abundant in PHTS-derived BCs, suggesting that the PHTS-derived BCs may be less responsive to immune checkpoint inhibitors.

Consistent with our findings in PHTS BC, previous studies focusing on sporadic BCs also showed that PTEN deficiency in tumors is associated with an immunosuppressive tumor microenvironment (TME) and resistance to immune checkpoint blockade[173;174]. Mechanistically, intrinsic PTEN deficiency in tumor cells stimulates the activation of PI3K signaling and the secretion of VEGF, which lead to the recruitment of immunosuppressive immune cells, abnormal angiogenesis, and resistance to T cell-mediated killing[174]. In contrast, distinct from sporadic cancers, PTEN deficiency in PHTS BCs occurs not only in tumor cells but also in the non-malignant normal cells (germline effect), including immune and stromal cells, which could influence the differentiation, expansion, activation, trafficking, and phenotypes of immune and stromal cells in the TME as well. For example, a previous study has found that genetic depletion of PTEN enhances NK cell cytolytic function against malignant cells, which is consistent with our data that increased proportion of activated NK cells was found in PHTS BCs[175]. Accordingly, to design strategies for immunotherapy in PHTS BCs, the influences of the PTEN pathway in both tumor and non-tumor (especially immune) compartments need to be considered. Notably, mutational signatures were also found to be associated with phenotypes of the TME and responsiveness to immunotherapy. For example, contrary to smoking associated signatures that show better response to immunotherapy, age-related mutational signature was found negatively associated with immune activity, survival outcomes, and the response to immunotherapy in triple-negative BC, melanoma, and/or NSCLC[176;177]. Our finding that PHTS BCs contain enriched age-related

mutational signature provides another potential linkage between PTEN deficiency and defective anti-tumor immune responses in PHTS BCs[133].

This study has several limitations which mirrors those mentioned in chapter II. They include the small sample size for PHTS, heterogeneity of each BC sample, not being whole genome sequencing, two different exome sequencing platforms used for the PHTS series and inability to examine non-coding regions and epigenetic alterations. While chapter II described the study examining small changes such as single nucleotide variant, small insertions and deletions, we attempted to examine larger genomic changes, namely CNV. Still, our pipeline was not designed to detect genomic changes such as chromosomal rearrangements, which may be important in tumorigenesis in PHTS-derived BCs.

Future directions include expanding on the size of our study cohort for PHTS, examining chromosomal rearrangements, non-coding and regulatory regions, and epigenetic studies. As mentioned in chapter II, utilizing fresh frozen breast tumor samples to examine metabolites and microbiomes may also provide insights into the biology of PHTS-derived BCs.

In conclusion, this study revealed key genomic and transcriptomic alterations in PHTS derived BCs which are distinct from those of the sporadic BC group from TCGA. We further revealed a potential key pathway associated with BC biology in PHTS, especially in the setting of pathogenic germline *PTEN* mutations. The alterations we identified enable hypothesis-driven studies to further characterize downstream functional effects contributing to BC carcinogenesis in PHTS. PHTS will only rise in incidence as clinical genetic

testing becomes more widely accessible in the clinic. Currently, there are no PHTS-specific treatment strategies for any type of PHTS component cancers, including BC. More extensive studies both at the clinical, translational and basic science levels are warranted to develop PTEN-targeted and personalized treatments, and perhaps preventatives, to effectively manage PHTS-derived cancers.

CHAPTER IV: CONCLUDING DISCUSSION AND FUTURE DIRECTIONS

General Overview

In this work, we demonstrated that PHTS-derived breast cancer (BC) has a distinct somatic genomic landscape compared to sporadic BC from TCGA. Our next generation sequencing data characterized genomic features of the PHTS-derived BCs, which facilitates the understanding of the biology of PHTS-derived BCs in comparison of sporadic BCs. Based on the data we presented in this body of work, it can be postulated that women with PHTS have a much higher risk of developing BC at younger ages compared to women from the general population for multiple reasons, including: 1) double hit (somatic) variants to the *PTEN* gene; 2) intrinsic tendency to accumulate point mutations much faster than the general population; 3) genomic instability; and 4) predicted impaired immune cell constitution and/or function.

Distinct Genomic Landscape in PHTS-derived Breast Cancer

We showed that somatic variants in *PTEN* (single nucleotide variants, small insertions and deletions) are the predominant, somatic oncogenic events in PHTS-derived BCs. Our data further demonstrated that high penetrance germline variants influence somatic phenotypes. The data restricted to BCs arising in the background of pathogenic or likely pathogenic *PTEN* germline variants further support this hypothesis. The caveat to this interpretation is that it is extremely challenging to determine the presence of true biallelic inactivation using our methodology. To evaluate more accurately, functional studies showing the

absence of PTEN protein expression and/or activity are necessary. Immunohistochemistry (IHC) staining may be feasible to perform to evaluate expression changes, while *in vitro* phosphatase activity assays may be used to interrogate classical PTEN activity.

Copy number variants (CNVs) in PHTS-derived BCs were also found to be distinct from the sporadic BCs from TCGA, especially revealing the significant peak at 6p22.2 containing multiple histone-related genes. This implies that PHTS-derived BC cells are under selective pressure for histone-related molecules. We could hypothesize that higher cell turnover and genomic instability necessitate such changes to endow cancer cells with a survival advantage. Our computational algorithm did not evaluate genomic fusions and rearrangements since these changes are known to be less relevant in BC in general. Further investigation into such large changes, however, may reveal more distinct genomic features in PHTS-BC which do not occur in sporadic BCs, and is hence worth pursuing in future studies.

We further showed that there are significant genome-wide differences between PHTS-derived and TCGA BCs. Interestingly, we detected COSMIC SBS29 signature in sporadic BC but not in PHTS, which signifies mutagens, like tobacco use[127], and may indicate environmental and lifestyle-related causes contributing to carcinogenesis in sporadic BCs but not in PHTS-derived BCs. On the other hand, we identified the SBS1 signature, which is strongly correlated with older age of cancer onset[127;128]. Detecting this signature in the PHTS-derived BCs leads us to speculate that mutational patterns that usually

accumulate with time tend to occur much faster in PHTS-derived BCs, especially in the background of pathogenic germline *PTEN* mutations. This observation may explain the early age of diagnosis of PHTS-BC as well as secondary BC which occurs at much higher frequency in PHTS. This is consistent with Knudson's second hit hypothesis, which further supports our clinical approach of careful breast cancer surveillance and possibility of prophylactic bilateral mastectomy in the clinical setting. Our clinical data, however, lack information on modifiable lifestyle choices such as smoking status, alcohol use, and exercise/activity levels. It would be insightful to incorporate these environmental factors to validate our hypothesis. Future studies should include PHTS-BC cell models to conduct experiments evaluating the DNA damage and repair mechanisms in PHTS-BC. Radiation exposure as well as chemical mutagens can be utilized to artificially induce DNA damage and compare the consequent cell repair efficacy between PHTS and non-PHTS BC models. We hypothesize that PHTS-BC cells are more susceptible to DNA damage, which could lead to carcinogenesis.

We showed that the predicted immune landscape infiltrating or surrounding breast carcinoma were distinct between PHTS-derived and TCGA BCs. Immune cell populations which point to either inactive or suppressive phenotypes were found increased in PHTS-derived BCs. We thus speculate that PTEN deficiency in PHTS-derived BC is associated with an immunosuppressive tumor microenvironment (TME). Furthermore, genes encoding immune checkpoint pathways, including PD-L1 (*CD274*), CTLA4, and PD-1 (*PDCD1*), were found less abundant in PHTS-derived BCs, implying that the PHTS-derived

BCs may be less responsive to immune checkpoint inhibitors. In our study, bulk RNA-seq data were analyzed by a computational pipeline to evaluate the immune cell populations in PHTS-BC tissue, which only allows us to capture the average gene expression of the cells present in the inherently heterogeneous sample. To increase resolution of gene expression of individual genes, single-cell RNA sequencing (scRNA-seq) may be useful, especially to evaluate the immune cell landscape of PHTS-BC[178]. As mentioned previously, PHTS-derived BC had a trend toward increased TMB. Increased TMB can be clinically useful as a therapeutic target or marker for immunotherapy in many types of cancers including BC. However, we need to interpret high TMB values in PHTS-derived BC with caution in the clinical setting because of the predicted immunotherapy resistant phenotypes observed in our study. We hence need more clinical data to investigate how individuals with PHTS who have developed metastatic BC may or may not respond to immunotherapy when it is clinically indicated for sporadic counterparts.

Revealing Clinically Distinct Subtypes of PHTS-derived BCs

We demonstrated that PHTS-derived breast cancers have two biologically distinct groups based on the pathogenicity of the underlying germline *PTEN* variants: 1) Tier-1 BCs, which harbor pathogenic or likely pathogenic germline PTEN variants; and 2) Tier-2 BCs, with variants of unknown significance (VUS) or likely benign variants in *PTEN*.

Pathway analysis based on the differentially expressed genes revealed α-tocopherol degradation to be a statistically significant canonical pathway in Tier-1 over Tier-2. We hypothesize that vitamin E may play an important role in suppressing the development of cancer in cells with dysfunctional PTEN-related pathways. Vitamin E degradation may explain the more penetrant nature of pathogenic germline *PTEN* variants. Future studies are warranted to test this hypothesis with carefully designed experiments using relevant *Pten* animal models as well as using PHTS patient samples. One important limitation of these RNA-seq derived data stems from the fact that we did not have matched pairs from normal breast tissues from our research participants. As a future direction, it would be informative to perform RNA-seq from matched tumor and normal tissues, which will help us identify tumor-specific transcriptomic signatures in PHTS-BC versus sporadic BC, in addition to identifying expression changes that may be pertinent for tumor initiation in the normal tissues harboring the background germline *PTEN* variants.

Limitations and Future Directions for Genomic Investigations in PHTS-BC

PHTS is a rare genetic disorder, which often limits our ability to recruit many (population-level) participants for research. Our genomic study should also be expanded in the future, specifically looking for other genomic changes to reveal important cancer biology in PHTS-BC, including non-coding and regulatory regions, chromosomal rearrangement analyses, and epigenetic studies. If feasible, fresh frozen breast tumor samples should be obtained to

perform microbiome and metabolomics analyses. Ultimately, a multi-omic approach to integrate different types of genomic data mentioned above should be attempted between PHTS-BC and sporadic BC. Furthermore, clinically relevant information, such as tumor markers, tumor grade, clinical staging, histology types, ethnicity, as well as environmental factors such as smoking and alcohol use should be incorporated for a more comprehensive analysis.

One of the biggest limitations of our RNA study was that a head-to-head comparison using RNA-seq between PHTS-BC and sporadic, TCGA BC cohorts was technically impossible, due to a large difference in sequencing technologies and the inability to batch-correct the data as a whole. This prevented us from performing certain analyses such as differentially expressed genes and canonical pathway predictions. Systematically collecting a new series of sporadic BC samples under our IRB protocol will allow us to obtain RNA-seq data using the same sequencing methods to perform a head-to-head a comparison between PHTS-BC and sporadic BC counterparts. This will provide useful information to reveal important canonical pathways in PHTS-BCs compared to sporadic BCs. Additionally, our attempt to integrate CNV data from DNA sequencing and expression data from RNAseq by correlational studies also posed a challenge. Unfortunately, not all FFPE samples taken for DNA analyses were available for RNA extraction. Thus, the CNV data and transcriptome data were not from the same groups of PHTS-BC. Ultimately, obtaining ample tissue samples along with a larger sample size is needed to reproduce and expand our results in the future.

To provide further mechanistic evidence, functional studies using cell lines, animal models, as well as clinical trials on human participants are warranted. Since PHTS-BC tends to be ER+, HER2-, non-high grade, selecting breast cancer cell lines with wildtype *BRCA1* such as MCF7 and MDA-MB-415, which are ER+, HER2-, *BRCA1* wildtype, and luminal A, are reasonable to use. By introducing *PTEN* mutations observed in our study, the transformed cell lines can be used as representative PHTS-BC pre-clinical models for further investigation.[179] For example, the established cell lines can be used to reproduce our results based on FFPE samples, as well as performing sc-RNAseq to further delineate the transcriptome profile. Furthermore, this type of proposed cell lines can be used to validate the key pathway we found in Tier-1 BC, namely the vitamin E deficiency pathway, by various methods including RNA-seq, metabolite measurements, and cell proliferation analyses. Relatedly, induced pluripotent stem cells (iPSC) have been successfully utilized in the Eng lab to develop organoids, specifically of the brain and thyroid. As a future direction, PHTS-BC models using breast organoids should be established, in order to study what early changes these PHTS cells may have throughout the various developmental stages, which may help explain BC susceptibility in the setting of germline *PTEN* mutations.

BCs are heterogeneous not only among affected individuals but even within the same affected breast, which is an inherent challenge when studying BC in general. To study the biology of PHTS-derived BC at a spatial resolution, we are currently conducting transcriptome profiling of PHTS-derived BC cases.

The purpose of this ongoing study is to investigate how the underlying biology may differ between different histologies within the same patient (for instance, invasive carcinoma vs. *in situ* carcinoma, vs. locoregional metastases). This study will help us understand the evolution of the primary BC lesions and the progression of PHTS-derived BCs and provide information to design effective and personalized treatments and preventatives for BCs in the setting of germline *PTEN* mutations.

Finally, the underlying carcinogenic mechanisms of other PHTS component cancers, including uterine, colorectal, kidney cancers and melanoma, have never been investigated in the past and should be explored for their own genomic profiles. For thyroid cancer, the second prevalent component cancer in PHTS next to breast cancer[50], genomic profiling comparing thyroid cancer tissues to PHTS-derived benign thyroid nodules as well as to sporadic thyroid cancer from TCGA has been completed by Dr. Eng's lab (Plitt G. et al., 2023; in press at Cancer Research at the time of drafting this dissertation). Furthermore, PHTS-related cancers serve as an excellent model to study other hereditary cancer syndromes, and even sporadic cancers with somatic alterations in *PTEN*.

Clinical Implications and Future Directions

Based on the biological pathways PTEN is involved in, we hypothesize that the PI3K/AKT/mTOR pathway is undeniably a pertinent treatment target for PHTS-derived malignancies, including BC. Vitamin E degradation pathway is also a candidate target, especially for BCs arising in the background of Tier-1

germline *PTEN* mutations. To identify other targetable biomarker candidates in PHTS derived malignancies, further investigation is warranted as previously described.

We speculate that when PTEN is severely dysfunctional, therapeutic agents targeting DNA damage may be useful, including DNA intercalating agents such as doxorubicin and poly(ADP-ribose) polymerase (PARP) inhibitors[147]. For individuals with PHTS who are at very high risk of breast cancer as well as other malignancies, medical management with an mTOR inhibitor or *PIK3CA* inhibitor, which are both inhibitory to the PI3K pathway, should be investigated in clinical trials as *preventative* measures. If clinically shown to be effective for cancer prevention, some women may be able to delay or avoid prophylactic surgeries such as bilateral mastectomies and hysterectomy which are often done to reduce the risk of breast cancer and uterine cancer, respectively. Furthermore, if RNA-seq data are available from PHTS-BC carcinoma tissue and matching normal tissue, we can potentially design an RNA-based vaccine against PHTS-BC in individuals with germline *PTEN* mutations by identifying common transcripts present in the tumor tissues but absent in matched normal breast tissue.

Our findings demonstrating the presence of two biologically distinct groups of PHTS-BC, based on the pathogenicity of the underlying germline *PTEN* mutations, provides clinically useful information to reclassify germline variants. There are a number of *PTEN* mutations which are initially classified as variants of unknown significance (VUS) due to the rarity of the variants as well as lack of clinical data to predict the biological nature of the variants. At the PTEN

Multidisciplinary Clinic and Center of Excellence at the Cleveland Clinic, we carefully evaluate each variant and classify them based on the nature of the genomic alteration in light of the patients' clinical presentations. Because PHTS is rare, there is a tremendous challenge to make the same level of clinical judgement by healthcare providers who are not familiar with PHTS. Therefore, it is crucial to establish a systematic approach to classify all *PTEN* germline variants in an accurate and standardized manner so that we can tailor screening and treatment options for individuals with PHTS in a more personalized manner no matter where they are diagnosed and treated.

We envision collecting more breast tissues, both malignant and benign, from an expanded PHTS cohort, which includes women with pathogenic and likely pathogenic *PTEN* mutations, variants of unknown significance (VUS), benign and likely benign *PTEN* variants, and those without germline *PTEN* variants but with a diagnosis of clinical Cowden syndrome. If we are to find significant genomic and molecular differences by comparing multi-omic profiles of these different groups, we can further classify them into different BC risk groups based on clinical parameters. This type of model will be extremely useful for breast pathologists, oncologists, and even cancer geneticists upon diagnosing and planning personalized medical and surgical managements of these PHTS patients. Clear classification and national guidelines especially for VUS and clinical Cowden syndrome would be helpful to guide oncologists and general providers to decide when to consider referring their patients to genetic counseling for further testing.

Summary and Significance of the Study

In conclusion, our data revealed several key findings, providing insights into distinct BC biology in the background of germline *PTEN* alterations and their clinical implications. We provide multiple lines of evidence for a distinct somatic genomic landscape of PHTS-derived BCs compared to sporadic counterparts. Despite its distinct somatic genomic landscape, PHTS-derived BC is treated similarly to sporadic BCs, according to the current standard of care. As we learn more about the mechanisms of PHTS-derived BCs and their potential therapeutic targets, we will be better empowered to develop more targeted, personalized strategies to effectively treat and in time, prevent PHTS-derived BCs. This is a population that will only rise in incidence as clinical genetic testing becomes more widely accessible in the clinic.

REFERENCES

1. Arnold, M., Morgan, E., Rumgay, H., Mafra, A., Singh, D., Laversanne, M., Vignat, J., Gralow, J.R., Cardoso, F., Siesling, S., et al. (2022). Current and future burden of breast cancer: Global statistics for 2020 and 2040. Breast 66, 15-23.
2. SEER. (2021). Surveillance, Epidemiology, and End Results (SEER) Program Populations (1969-2019), National Cancer Institute, DCCPS, Surveillance Research Program, released Feburary 2021. In.
3. Polyak, K. (2011). Heterogeneity in breast cancer. J Clin Invest 121, 3786-3788.
4. Sun, Y.S., Zhao, Z., Yang, Z.N., Xu, F., Lu, H.J., Zhu, Z.Y., Shi, W., Jiang, J., Yao, P.P., and Zhu, H.P. (2017). Risk Factors and Preventions of Breast Cancer. Int J Biol Sci 13, 1387-1397.
5. Zhu, H., and Dogan, B.E. (2021). American Joint Committee on Cancer's Staging System for Breast Cancer, Eighth Edition: Summary for Clinicians. Eur J Breast Health 17, 234-238.
6. Giuliano, A.E., Edge, S.B., and Hortobagyi, G.N. (2018). Eighth Edition of the AJCC Cancer Staging Manual: Breast Cancer. Ann Surg Oncol 25, 1783-1785.
7. Rakha, E.A., Reis-Filho, J.S., Baehner, F., Dabbs, D.J., Decker, T., Eusebi, V., Fox, S.B., Ichihara, S., Jacquemier, J., Lakhani, S.R., et al. (2010). Breast cancer prognostic classification in the molecular era: the role of histological grade. Breast Cancer Res 12, 207.

8. Onitilo, A.A., Engel, J.M., Greenlee, R.T., and Mukesh, B.N. (2009). Breast cancer subtypes based on ER/PR and Her2 expression: comparison of clinicopathologic features and survival. Clin Med Res 7, 4-13.

9. Makki, J. (2015). Diversity of Breast Carcinoma: Histological Subtypes and Clinical Relevance. Clin Med Insights Pathol 8, 23-31.

10. Mohammed, A.A. (2021). The clinical behavior of different molecular subtypes of breast cancer. Cancer Treat Res Commun 29, 100469.

11. Dai, X., Li, T., Bai, Z., Yang, Y., Liu, X., Zhan, J., and Shi, B. (2015). Breast cancer intrinsic subtype classification, clinical use and future trends. Am J Cancer Res 5, 2929-2943.

12. Yang, J., Nie, J., Ma, X., Wei, Y., Peng, Y., and Wei, X. (2019). Targeting PI3K in cancer: mechanisms and advances in clinical trials. Mol Cancer 18, 26.

13. Baselga, J., Campone, M., Piccart, M., Burris, H.A., 3rd, Rugo, H.S., Sahmoud, T., Noguchi, S., Gnant, M., Pritchard, K.I., Lebrun, F., et al. (2012). Everolimus in postmenopausal hormone-receptor-positive advanced breast cancer. N Engl J Med 366, 520-529.

14. China Anti-cancer Association Tumor Drug Clinical Research, C., Breast Cancer Expert Committee, N.T.Q.C.C., Tumor Pathology Committee of China Anti-cancer, A., and Boao Institute of Oncology, I. (2022). [Expert consensus on the clinical application of PI3K/AKT/mTOR inhibitors in the treatment of advanced breast cancer]. Zhonghua Zhong Liu Za Zhi 44, 673-692.

15. FDA. (2019). FDA approves alpelisib for metastatic breast cancer. https://www.fda.gov/drugs/resources-information-approved-drugs/fda-approves-alpelisib-metastatic-breast-cancer. Accessed May 9, 2023. .
16. FDA. (2022). FDA approves alpelisib for PIK3CA-related overgrowth spectrum. https://www.fda.gov/drugs/resources-information-approved-drugs/fda-approves-alpelisib-pik3ca-related-overgrowth-spectrum. Accessed May 9, 2023.
17. Goel, S., DeCristo, M.J., McAllister, S.S., and Zhao, J.J. (2018). CDK4/6 Inhibition in Cancer: Beyond Cell Cycle Arrest. Trends Cell Biol 28, 911-925.
18. FDA. (2018). FDA approves first cancer drug through new oncology review pilot that enables greater development efficiency. https://www.fda.gov/news-events/press-announcements/fda-approves-first-cancer-drug-through-new-oncology-review-pilot-enables-greater-development. Accessed May 9, 2023. .
19. FDA. (2017). Palbociclib (IBRANCE). https://www.fda.gov/drugs/resources-information-approved-drugs/palbociclib-ibrance. Accessed May 9, 2023. .
20. FDA. (2017). FDA approves abemaciclib for HR-positive, HER2-negative breast cancer. https://www.fda.gov/drugs/resources-information-approved-drugs/fda-approves-abemaciclib-hr-positive-her2-negative-breast-cancer. Accessed May 9, 2023.
21. FDA. (2021). FDA approves abemaciclib with endocrine therapy for early breast cancer. https://www.fda.gov/drugs/resources-information-approved-

drugs/fda-approves-abemaciclib-endocrine-therapy-early-breast-cancer. Accessed May 9, 2023. In.

22. FDA. (2021). FDA grants regular approval to sacituzumab govitecan for triple-negative breast cancer. https://www.fda.gov/drugs/resources-information-approved-drugs/fda-grants-regular-approval-sacituzumab-govitecan-triple-negative-breast-cancer. Accessed May 9, 2023.
23. FDA. (2023). FDA approves sacituzumab govitecan-hziy for HR-positive breast cancer. https://www.fda.gov/drugs/resources-information-approved-drugs/fda-approves-sacituzumab-govitecan-hziy-hr-positive-breast-cancer. Accessed May 9, 2023.
24. FDA. (2023). FDA approves elacestrant for ER-positive, HER2-negative, ESR1-mutated advanced or metastatic breast cancer. https://www.fda.gov/drugs/resources-information-approved-drugs/fda-approves-elacestrant-er-positive-her2-negative-esr1-mutated-advanced-or-metastatic-breast-cancer. Accessed May 9, 2023.
25. FDA. (221). FDA approves pembrolizumab for high-risk early-stage triple-negative breast cancer. https://www.fda.gov/drugs/resources-information-approved-drugs/fda-approves-pembrolizumab-high-risk-early-stage-triple-negative-breast-cancer. Accessed May 7, 2023.
26. FDA. (2020). FDA grants accelerated approval to pembrolizumab for locally recurrent unresectable or metastatic triple negative breast cancer. https://www.fda.gov/drugs/resources-information-approved-drugs/fda-grants-accelerated-approval-pembrolizumab-locally-recurrent-

unresectable-or-metastatic-triple?utm_medium=email&utm_source=govdelivery. Accessed May 7, 2023.

27. FDA. (2021). FDA grants accelerated approval to dostarlimab-gxly for dMMR advanced solid tumors. https://www.fda.gov/drugs/resources-information-approved-drugs/fda-grants-accelerated-approval-dostarlimab-gxly-dmmr-advanced-solid-tumors. Accessed May 7, 2023.
28. Solinas, C., Gombos, A., Latifyan, S., Piccart-Gebhart, M., Kok, M., and Buisseret, L. (2017). Targeting immune checkpoints in breast cancer: an update of early results. ESMO Open 2, e000255.
29. Hanna, A., and Balko, J.M. (2021). Breast cancer resistance mechanisms: challenges to immunotherapy. Breast Cancer Res Treat 190, 5-17.
30. Martins, F., Sofiya, L., Sykiotis, G.P., Lamine, F., Maillard, M., Fraga, M., Shabafrouz, K., Ribi, C., Cairoli, A., Guex-Crosier, Y., et al. (2019). Adverse effects of immune-checkpoint inhibitors: epidemiology, management and surveillance. Nat Rev Clin Oncol 16, 563-580.
31. McDonald, K.A., Kawaguchi, T., Qi, Q., Peng, X., Asaoka, M., Young, J., Opyrchal, M., Yan, L., Patnaik, S., Otsuji, E., et al. (2019). Tumor Heterogeneity Correlates with Less Immune Response and Worse Survival in Breast Cancer Patients. Ann Surg Oncol 26, 2191-2199.
32. TCGA. (2012). Comprehensive molecular portraits of human breast tumours. Nature 490, 61-70.

33. Andre, F., Ciruelos, E.M., Juric, D., Loibl, S., Campone, M., Mayer, I.A., Rubovszky, G., Yamashita, T., Kaufman, B., Lu, Y.S., et al. (2021). Alpelisib plus fulvestrant for PIK3CA-mutated, hormone receptor-positive, human epidermal growth factor receptor-2-negative advanced breast cancer: final overall survival results from SOLAR-1. Ann Oncol 32, 208-217.

34. Shiovitz, S., and Korde, L.A. (2015). Genetics of breast cancer: a topic in evolution. Ann Oncol 26, 1291-1299.

35. Kuchenbaecker, K.B., Hopper, J.L., Barnes, D.R., Phillips, K.A., Mooij, T.M., Roos-Blom, M.J., Jervis, S., van Leeuwen, F.E., Milne, R.L., Andrieu, N., et al. (2017). Risks of Breast, Ovarian, and Contralateral Breast Cancer for BRCA1 and BRCA2 Mutation Carriers. JAMA 317, 2402-2416.

36. NCCN. (2022). Clinical Practice Guidelines in Oncology (National Comprehensive Cancer Network [NCCN] guidelines) Clinical practice guideline in Oncology: Genetic/Familial High-Risk Assessment: Breast, Ovarian, and Pancreatic. Version 1.2022 – August 11, 2021. In.

37. Robson, M., Im, S.A., Senkus, E., Xu, B., Domchek, S.M., Masuda, N., Delaloge, S., Li, W., Tung, N., Armstrong, A., et al. (2017). Olaparib for Metastatic Breast Cancer in Patients with a Germline BRCA Mutation. N Engl J Med 377, 523-533.

38. FDA. (2022). FDA approves olaparib for adjuvant treatment of high-risk early breast cancer. https://www.fda.gov/drugs/resources-information-approved-

drugs/fda-approves-olaparib-adjuvant-treatment-high-risk-early-breast-cancer. Access May 7, 2023.

39. Lloyd, K.M., 2nd, and Dennis, M. (1963). Cowden's disease. A possible new symptom complex with multiple system involvement. Ann Intern Med 58, 136-142.

40. Starink, T.M., van der Veen, J.P., Arwert, F., de Waal, L.P., de Lange, G.G., Gille, J.J., and Eriksson, A.W. (1986). The Cowden syndrome: a clinical and genetic study in 21 patients. Clin Genet 29, 222-233.

41. Brownstein, M.H., Mehregan, A.H., Bikowski, J.B., Lupulescu, A., and Patterson, J.C. (1979). The dermatopathology of Cowden's syndrome. Br J Dermatol 100, 667-673.

42. Eng, C. (2000). Will the real Cowden syndrome please stand up: revised diagnostic criteria. J Med Genet 37, 828-830.

43. Nelen, M.R., Padberg, G.W., Peeters, E.A., Lin, A.Y., van den Helm, B., Frants, R.R., Coulon, V., Goldstein, A.M., van Reen, M.M., Easton, D.F., et al. (1996). Localization of the gene for Cowden disease to chromosome 10q22-23. Nat Genet 13, 114-116.

44. Liaw, D., Marsh, D.J., Li, J., Dahia, P.L., Wang, S.I., Zheng, Z., Bose, S., Call, K.M., Tsou, H.C., Peacocke, M., et al. (1997). Germline mutations of the PTEN gene in Cowden disease, an inherited breast and thyroid cancer syndrome. Nat Genet 16, 64-67.

45. Marsh, D.J., Kum, J.B., Lunetta, K.L., Bennett, M.J., Gorlin, R.J., Ahmed, S.F., Bodurtha, J., Crowe, C., Curtis, M.A., Dasouki, M., et al. (1999).

PTEN mutation spectrum and genotype-phenotype correlations in Bannayan-Riley-Ruvalcaba syndrome suggest a single entity with Cowden syndrome. Hum Mol Genet 8, 1461-1472.

46. Mester, J., and Eng, C. (2012). Estimate of de novo mutation frequency in probands with PTEN hamartoma tumor syndrome. Genet Med 14, 819-822.

47. Tan, M.H., Mester, J., Peterson, C., Yang, Y., Chen, J.L., Rybicki, L.A., Milas, K., Pederson, H., Remzi, B., Orloff, M.S., et al. (2011). A clinical scoring system for selection of patients for PTEN mutation testing is proposed on the basis of a prospective study of 3042 probands. Am J Hum Genet 88, 42-56.

48. Lee, Y.R., Yehia, L., Kishikawa, T., Ni, Y., Leach, B., Zhang, J., Panch, N., Liu, J., Wei, W., Eng, C., et al. (2020). WWP1 Gain-of-Function Inactivation of PTEN in Cancer Predisposition. N Engl J Med 382, 2103-2116.

49. Ni, Y., He, X., Chen, J., Moline, J., Mester, J., Orloff, M.S., Ringel, M.D., and Eng, C. (2012). Germline SDHx variants modify breast and thyroid cancer risks in Cowden and Cowden-like syndrome via FAD/NAD-dependant destabilization of p53. Hum Mol Genet 21, 300-310.

50. Tan, M.H., Mester, J.L., Ngeow, J., Rybicki, L.A., Orloff, M.S., and Eng, C. (2012). Lifetime cancer risks in individuals with germline PTEN mutations. Clin Cancer Res 18, 400-407.

51. Bubien, V., Bonnet, F., Brouste, V., Hoppe, S., Barouk-Simonet, E., David, A., Edery, P., Bottani, A., Layet, V., Caron, O., et al. (2013). High cumulative risks of cancer in patients with PTEN hamartoma tumour syndrome. J Med Genet 50, 255-263.

52. Nieuwenhuis, M.H., Kets, C.M., Murphy-Ryan, M., Yntema, H.G., Evans, D.G., Colas, C., Moller, P., Hes, F.J., Hodgson, S.V., Olderode-Berends, M.J., et al. (2014). Cancer risk and genotype-phenotype correlations in PTEN hamartoma tumor syndrome. Fam Cancer 13, 57-63.

53. Riegert-Johnson, D.L., Gleeson, F.C., Roberts, M., Tholen, K., Youngborg, L., Bullock, M., and Boardman, L.A. (2010). Cancer and Lhermitte-Duclos disease are common in Cowden syndrome patients. Hered Cancer Clin Pract 8, 6.

54. Yehia, L., Plitt, G., Tushar, A.M., Joo, J., Burke, C.A., Campbell, S.C., Heiden, K., Jin, J., Macaron, C., Michener, C.M., et al. (2023). Longitudinal Analysis of Cancer Risk in Children and Adults With Germline PTEN Variants. JAMA Netw Open 6, e239705.

55. Ngeow, J., Stanuch, K., Mester, J.L., Barnholtz-Sloan, J.S., and Eng, C. (2014). Second malignant neoplasms in patients with Cowden syndrome with underlying germline PTEN mutations. J Clin Oncol 32, 1818-1824.

56. Li, J., Yen, C., Liaw, D., Podsypanina, K., Bose, S., Wang, S.I., Puc, J., Miliaresis, C., Rodgers, L., McCombie, R., et al. (1997). PTEN, a putative protein tyrosine phosphatase gene mutated in human brain, breast, and prostate cancer. Science 275, 1943-1947.

57. Millis, S.Z., Ikeda, S., Reddy, S., Gatalica, Z., and Kurzrock, R. (2016). Landscape of Phosphatidylinositol-3-Kinase Pathway Alterations Across 19784 Diverse Solid Tumors. JAMA Oncol 2, 1565-1573.

58. Ngeow, J., He, X., Mester, J.L., Lei, J., Romigh, T., Orloff, M.S., Milas, M., and Eng, C. (2012). Utility of PTEN protein dosage in predicting for underlying germline PTEN mutations among patients presenting with thyroid cancer and Cowden-like phenotypes. J Clin Endocrinol Metab 97, E2320-2327.

59. Nelen, M.R., Kremer, H., Konings, I.B., Schoute, F., van Essen, A.J., Koch, R., Woods, C.G., Fryns, J.P., Hamel, B., Hoefsloot, L.H., et al. (1999). Novel PTEN mutations in patients with Cowden disease: absence of clear genotype-phenotype correlations. Eur J Hum Genet 7, 267-273.

60. Yehia, L., Ngeow, J., and Eng, C. (2019). PTEN-opathies: from biological insights to evidence-based precision medicine. J Clin Invest 129, 452-464.

61. Ngeow, J., Sesock, K., and Eng, C. (2017). Clinical Implications for Germline PTEN Spectrum Disorders. Endocrinol Metab Clin North Am 46, 503-517.

62. Feng, J., Dang, Y., Zhang, W., Zhao, X., Zhang, C., Hou, Z., Jin, Y., McNutt, M.A., Marks, A.R., and Yin, Y. (2019). PTEN arginine methylation by PRMT6 suppresses PI3K-AKT signaling and modulates pre-mRNA splicing. Proc Natl Acad Sci U S A 116, 6868-6877.

63. Maehama, T., and Dixon, J.E. (1998). The tumor suppressor, PTEN/MMAC1, dephosphorylates the lipid second messenger, phosphatidylinositol 3,4,5-trisphosphate. J Biol Chem 273, 13375-13378.

64. Myers, M.P., Stolarov, J.P., Eng, C., Li, J., Wang, S.I., Wigler, M.H., Parsons, R., and Tonks, N.K. (1997). P-TEN, the tumor suppressor from human chromosome 10q23, is a dual-specificity phosphatase. Proc Natl Acad Sci U S A 94, 9052-9057.

65. Stambolic, V., Suzuki, A., de la Pompa, J.L., Brothers, G.M., Mirtsos, C., Sasaki, T., Ruland, J., Penninger, J.M., Siderovski, D.P., and Mak, T.W. (1998). Negative regulation of PKB/Akt-dependent cell survival by the tumor suppressor PTEN. Cell 95, 29-39.

66. Planchon, S.M., Waite, K.A., and Eng, C. (2008). The nuclear affairs of PTEN. J Cell Sci 121, 249-253.

67. Chang, F., Lee, J.T., Navolanic, P.M., Steelman, L.S., Shelton, J.G., Blalock, W.L., Franklin, R.A., and McCubrey, J.A. (2003). Involvement of PI3K/Akt pathway in cell cycle progression, apoptosis, and neoplastic transformation: a target for cancer chemotherapy. Leukemia 17, 590-603.

68. Whitman, M., Kaplan, D.R., Schaffhausen, B., Cantley, L., and Roberts, T.M. (1985). Association of phosphatidylinositol kinase activity with polyoma middle-T competent for transformation. Nature 315, 239-242.

69. Paplomata, E., and O'Regan, R. (2014). The PI3K/AKT/mTOR pathway in breast cancer: targets, trials and biomarkers. Ther Adv Med Oncol 6, 154-166.

70. Whitman, M., Downes, C.P., Keeler, M., Keller, T., and Cantley, L. (1988). Type I phosphatidylinositol kinase makes a novel inositol phospholipid, phosphatidylinositol-3-phosphate. Nature 332, 644-646.

71. Carpenter, C.L., Duckworth, B.C., Auger, K.R., Cohen, B., Schaffhausen, B.S., and Cantley, L.C. (1990). Purification and characterization of phosphoinositide 3-kinase from rat liver. J Biol Chem 265, 19704-19711.

72. Carracedo, A., and Pandolfi, P.P. (2008). The PTEN-PI3K pathway: of feedbacks and cross-talks. Oncogene 27, 5527-5541.

73. Amorim, I.S., Lach, G., and Gkogkas, C.G. (2018). The Role of the Eukaryotic Translation Initiation Factor 4E (eIF4E) in Neuropsychiatric Disorders. Front Genet 9, 561.

74. Yehia, L., and Eng, C. (2021). PTEN Hamartoma Tumor Syndrome. In GeneReviews(R), M.P. Adam, H.H. Ardinger, R.A. Pagon, S.E. Wallace, L.J.H. Bean, G. Mirzaa, and A. Amemiya, eds. (

75. Hou, S.Q., Ouyang, M., Brandmaier, A., Hao, H., and Shen, W.H. (2017). PTEN in the maintenance of genome integrity: From DNA replication to chromosome segregation. Bioessays 39.

76. Chen, Z.H., Zhu, M., Yang, J., Liang, H., He, J., He, S., Wang, P., Kang, X., McNutt, M.A., Yin, Y., et al. (2014). PTEN interacts with histone H1 and controls chromatin condensation. Cell Rep 8, 2003-2014.

77. Shen, W.H., Balajee, A.S., Wang, J., Wu, H., Eng, C., Pandolfi, P.P., and Yin, Y. (2007). Essential role for nuclear PTEN in maintaining chromosomal integrity. Cell 128, 157-170.

78. Liu, F., Wagner, S., Campbell, R.B., Nickerson, J.A., Schiffer, C.A., and Ross, A.H. (2005). PTEN enters the nucleus by diffusion. J Cell Biochem 96, 221-234.

79. Hou, B., Xu, S., Xu, Y., Gao, Q., Zhang, C., Liu, L., Yang, H., Jiang, X., and Che, Y. (2019). Grb2 binds to PTEN and regulates its nuclear translocation to maintain the genomic stability in DNA damage response. Cell Death Dis 10, 546.

80. Misra, S., Ghosh, G., Chowdhury, S.G., and Karmakar, P. (2021). Non-canonical function of nuclear PTEN and its implication on tumorigenesis. DNA Repair (Amst) 107, 103197.

81. He, J., Zhang, Z., Ouyang, M., Yang, F., Hao, H., Lamb, K.L., Yang, J., Yin, Y., and Shen, W.H. (2016). PTEN regulates EG5 to control spindle architecture and chromosome congression during mitosis. Nat Commun 7, 12355.

82. Freeman, D.J., Li, A.G., Wei, G., Li, H.H., Kertesz, N., Lesche, R., Whale, A.D., Martinez-Diaz, H., Rozengurt, N., Cardiff, R.D., et al. (2003). PTEN tumor suppressor regulates p53 protein levels and activity through phosphatase-dependent and -independent mechanisms. Cancer Cell 3, 117-130.

83. Liu, Y., Du, X., Zhang, S., Liu, Y., Zhang, Q., Yin, Q., McNutt, M.A., and Yin, Y. (2017). PTEN regulates spindle assembly checkpoint timing through MAD1 in interphase. Oncotarget 8, 98040-98050.

84. Choi, B.H., Xie, S., and Dai, W. (2017). PTEN is a negative regulator of mitotic checkpoint complex during the cell cycle. Exp Hematol Oncol 6, 19.

85. Feng, J., Liang, J., Li, J., Li, Y., Liang, H., Zhao, X., McNutt, M.A., and Yin, Y. (2015). PTEN Controls the DNA Replication Process through MCM2 in Response to Replicative Stress. Cell Rep 13, 1295-1303.

86. Gil, A., Andres-Pons, A., Fernandez, E., Valiente, M., Torres, J., Cervera, J., and Pulido, R. (2006). Nuclear localization of PTEN by a Ran-dependent mechanism enhances apoptosis: Involvement of an N-terminal nuclear localization domain and multiple nuclear exclusion motifs. Mol Biol Cell 17, 4002-4013.

87. Morgan, F.C., Yehia, L., McDonald, C., Martinez-Agosto, J.A., Hardan, A.Y., Tamburro, J., Sahin, M., Bayart, C., Eng, C., and Developmental Synaptopathies, C. (2023). Characterizing dermatologic findings among patients with PTEN hamartoma tumor syndrome: Results of a multicenter cohort study. J Am Acad Dermatol 89, 90-98.

88. McLoughlin, N.M., Mueller, C., and Grossmann, T.N. (2018). The Therapeutic Potential of PTEN Modulation: Targeting Strategies from Gene to Protein. Cell Chem Biol 25, 19-29.

89. Mester, J.L., Ghosh, R., Pesaran, T., Huether, R., Karam, R., Hruska, K.S., Costa, H.A., Lachlan, K., Ngeow, J., Barnholtz-Sloan, J., et al. (2018). Gene-specific criteria for PTEN variant curation: Recommendations from the ClinGen PTEN Expert Panel. Hum Mutat 39, 1581-1592.

90. Teresi, R.E., Zbuk, K.M., Pezzolesi, M.G., Waite, K.A., and Eng, C. (2007). Cowden syndrome-affected patients with PTEN promoter mutations demonstrate abnormal protein translation. Am J Hum Genet 81, 756-767.

91. Wang, Y., Romigh, T., He, X., Tan, M.H., Orloff, M.S., Silverman, R.H., Heston, W.D., and Eng, C. (2011). Differential regulation of PTEN expression by androgen receptor in prostate and breast cancers. Oncogene 30, 4327-4338.

92. Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25, 1754-1760.

93. McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., et al. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res 20, 1297-1303.

94. Huang, K.L., Mashl, R.J., Wu, Y., Ritter, D.I., Wang, J., Oh, C., Paczkowska, M., Reynolds, S., Wyczalkowski, M.A., Oak, N., et al. (2018). Pathogenic Germline Variants in 10,389 Adult Cancers. Cell 173, 355-370 e314.

95. Saunders, C.T., Wong, W.S., Swamy, S., Becq, J., Murray, L.J., and Cheetham, R.K. (2012). Strelka: accurate somatic small-variant calling from sequenced tumor-normal sample pairs. Bioinformatics 28, 1811-1817.

96. Lai, Z., Markovets, A., Ahdesmaki, M., Chapman, B., Hofmann, O., McEwen, R., Johnson, J., Dougherty, B., Barrett, J.C., and Dry, J.R. (2016). VarDict: a novel and versatile variant caller for next-generation sequencing in cancer research. Nucleic Acids Res 44, e108.

97. Wang, M., Luo, W., Jones, K., Bian, X., Williams, R., Higson, H., Wu, D., Hicks, B., Yeager, M., and Zhu, B. (2020). SomaticCombiner: improving

the performance of somatic variant calling based on evaluation tests and a consensus approach. Sci Rep 10, 12898.

98. Zapata, L., Susak, H., Drechsel, O., Friedlander, M.R., Estivill, X., and Ossowski, S. (2017). Signatures of positive selection reveal a universal role of chromatin modifiers as cancer driver genes. Sci Rep 7, 13124.

99. Kaur, P., Porras, T.B., Ring, A., Carpten, J.D., and Lang, J.E. (2019). Comparison of TCGA and GENIE genomic datasets for the detection of clinically actionable alterations in breast cancer. Sci Rep 9, 1482.

100. Hamdi, Y., Soucy, P., Adoue, V., Michailidou, K., Canisius, S., Lemacon, A., Droit, A., Andrulis, I.L., Anton-Culver, H., Arndt, V., et al. (2016). Association of breast cancer risk with genetic variants showing differential allelic expression: Identification of a novel breast cancer susceptibility locus at 4q21. Oncotarget 7, 80140-80163.

101. Woditschka, S., Evans, L., Duchnowska, R., Reed, L.T., Palmieri, D., Qian, Y., Badve, S., Sledge, G., Jr., Gril, B., Aladjem, M.I., et al. (2014). DNA double-strand break repair genes and oxidative damage in brain metastasis of breast cancer. J Natl Cancer Inst 106.

102. Del Valle, J., Rofes, P., Moreno-Cabrera, J.M., Lopez-Doriga, A., Belhadj, S., Vargas-Parra, G., Teule, A., Cuesta, R., Munoz, X., Campos, O., et al. (2020). Exploring the Role of Mutations in Fanconi Anemia Genes in Hereditary Cancer Patients. Cancers (Basel) 12.

103. Neidhardt, G., Hauke, J., Ramser, J., Gross, E., Gehrig, A., Muller, C.R., Kahlert, A.K., Hackmann, K., Honisch, E., Niederacher, D., et al. (2017).

Association Between Loss-of-Function Mutations Within the FANCM Gene and Early-Onset Familial Breast Cancer. JAMA Oncol 3, 1245-1248.

104. Fang, C.B., Wu, H.T., Zhang, M.L., Liu, J., and Zhang, G.J. (2020). Fanconi Anemia Pathway: Mechanisms of Breast Cancer Predisposition Development and Potential Therapeutic Targets. Front Cell Dev Biol 8, 160.

105. Nizialek, E.A., Peterson, C., Mester, J.L., Downes-Kelly, E., and Eng, C. (2013). Germline and somatic KLLN alterations in breast cancer dysregulate G2 arrest. Hum Mol Genet 22, 2451-2461.

106. Bartkova, J., Tommiska, J., Oplustilova, L., Aaltonen, K., Tamminen, A., Heikkinen, T., Mistrik, M., Aittomaki, K., Blomqvist, C., Heikkila, P., et al. (2008). Aberrations of the MRE11-RAD50-NBS1 DNA damage sensor complex in human breast cancer: MRE11 as a candidate familial cancer-predisposing gene. Mol Oncol 2, 296-316.

107. Zhang, B., Beeghly-Fadiel, A., Long, J., and Zheng, W. (2011). Genetic variants associated with breast-cancer risk: comprehensive research synopsis, meta-analysis, and epidemiological evidence. Lancet Oncol 12, 477-488.

108. Kuiper, R.P., Nielsen, M., De Voer, R.M., and Hoogerbrugge, N. (1993). NTHL1 Tumor Syndrome, GeneReviews(R) M.P. Adam, H.H. Ardinger, R.A. Pagon, S.E. Wallace, L.J.H. Bean, G. Mirzaa, and A. Amemiya. PMID: 32239880

109. Heikkinen, K., Rapakko, K., Karppinen, S.M., Erkko, H., Knuutila, S., Lundan, T., Mannermaa, A., Borresen-Dale, A.L., Borg, A., Barkardottir, R.B., et al. (2006). RAD50 and NBS1 are breast cancer susceptibility genes associated with genomic instability. Carcinogenesis 27, 1593-1599.

110. Yang, X., Song, H., Leslie, G., Engel, C., Hahnen, E., Auber, B., Horvath, J., Kast, K., Niederacher, D., Turnbull, C., et al. (2020). Ovarian and Breast Cancer Risks Associated With Pathogenic Variants in RAD51C and RAD51D. J Natl Cancer Inst 112, 1242-1250.

111. Baker, J.L., Schwab, R.B., Wallace, A.M., and Madlensky, L. (2015). Breast cancer in a RAD51D mutation carrier: case report and review of the literature. Clin Breast Cancer 15, e71-75.

112. Yehia, L., Niazi, F., Ni, Y., Ngeow, J., Sankunny, M., Liu, Z., Wei, W., Mester, J.L., Keri, R.A., Zhang, B., et al. (2015). Germline Heterozygous Variants in SEC23B Are Associated with Cowden Syndrome and Enriched in Apparently Sporadic Thyroid Cancer. Am J Hum Genet 97, 661-676.

113. Sun, J., Wang, Y., Xia, Y., Xu, Y., Ouyang, T., Li, J., Wang, T., Fan, Z., Fan, T., Lin, B., et al. (2015). Mutations in RECQL Gene Are Associated with Predisposition to Breast Cancer. PLoS Genet 11, e1005228.

114. Park, D.J., Tao, K., Le Calvez-Kelm, F., Nguyen-Dumont, T., Robinot, N., Hammet, F., Odefrey, F., Tsimiklis, H., Teo, Z.L., Thingholm, L.B., et al. (2014). Rare mutations in RINT1 predispose carriers to breast and Lynch syndrome-spectrum cancers. Cancer Discov 4, 804-815.

115. Kluzniak, W., Wokolorczyk, D., Rusak, B., Huzarski, T., Gronwald, J., Stempa, K., Rudnicka, H., Kashyap, A., Debniak, T., Jakubowska, A., et al. (2019). Inherited variants in XRCC2 and the risk of breast cancer. Breast Cancer Res Treat 178, 657-663.

116. Yehia, L., and Eng, C. (2020). PTEN hamartoma tumour syndrome: what happens when there is no PTEN germline mutation? Hum Mol Genet 29, R150-R157.

117. Barnell, E.K., Ronning, P., Campbell, K.M., Krysiak, K., Ainscough, B.J., Sheta, L.M., Pema, S.P., Schmidt, A.D., Richters, M., Cotto, K.C., et al. (2019). Standard operating procedure for somatic variant refinement of sequencing data with paired tumor and normal samples. Genet Med 21, 972-981.

118. Mayakonda, A., Lin, D.C., Assenov, Y., Plass, C., and Koeffler, H.P. (2018). Maftools: efficient and comprehensive analysis of somatic variants in cancer. Genome Res 28, 1747-1756.

119. Viechtbauer, W. (2010). Conducting Meta-Analyses in R with the metafor Package. Journal of Statistical Software 36, 1-48.

120. Shen, R., and Seshan, V.E. (2016). FACETS: allele-specific copy number and clonal heterogeneity analysis tool for high-throughput DNA sequencing. Nucleic Acids Res 44, e131.

121. Roberts, S.A., Lawrence, M.S., Klimczak, L.J., Grimm, S.A., Fargo, D., Stojanov, P., Kiezun, A., Kryukov, G.V., Carter, S.L., Saksena, G., et al.

(2013). An APOBEC cytidine deaminase mutagenesis pattern is widespread in human cancers. Nat Genet 45, 970-976.

122. Gaujoux, R., and Seoighe, C. (2010). A flexible R package for nonnegative matrix factorization. BMC Bioinformatics 11, 367.

123. Melendez, B., Van Campenhout, C., Rorive, S., Remmelink, M., Salmon, I., and D'Haene, N. (2018). Methods of measurement for tumor mutational burden in tumor tissue. Transl Lung Cancer Res 7, 661-667.

124. Cohen, J. (1988). Statistical power analysis for the behavioral sciences, Hillsdale, NJ: L. Erlbaum Associates.

125. Knudson, A.G., Jr. (1971). Mutation and cancer: statistical study of retinoblastoma. Proc Natl Acad Sci U S A 68, 820-823.

126. Swanton, C., McGranahan, N., Starrett, G.J., and Harris, R.S. (2015). APOBEC Enzymes: Mutagenic Fuel for Cancer Evolution and Heterogeneity. Cancer Discov 5, 704-712.

127. Tate, J.G., Bamford, S., Jubb, H.C., Sondka, Z., Beare, D.M., Bindal, N., Boutselakis, H., Cole, C.G., Creatore, C., Dawson, E., et al. (2019). COSMIC: the Catalogue Of Somatic Mutations In Cancer. Nucleic Acids Res 47, D941-D947.

128. Alexandrov, L.B., Kim, J., Haradhvala, N.J., Huang, M.N., Tian Ng, A.W., Wu, Y., Boot, A., Covington, K.R., Gordenin, D.A., Bergstrom, E.N., et al. (2020). The repertoire of mutational signatures in human cancer. Nature 578, 94-101.

129. Hanahan, D., and Weinberg, R.A. (2011). Hallmarks of cancer: the next generation. Cell 144, 646-674.

130. Srinivasan, P., Bandlamudi, C., Jonsson, P., Kemel, Y., Chavan, S.S., Richards, A.L., Penson, A.V., Bielski, C.M., Fong, C., Syed, A., et al. (2021). The context-specific role of germline pathogenicity in tumorigenesis. Nat Genet 53, 1577-1585.

131. Papa, A., Wan, L., Bonora, M., Salmena, L., Song, M.S., Hobbs, R.M., Lunardi, A., Webster, K., Ng, C., Newton, R.H., et al. (2014). Cancer-associated PTEN mutants act in a dominant-negative manner to suppress PTEN protein function. Cell 157, 595-610.

132. Fan, X., Kraynak, J., Knisely, J.P.S., Formenti, S.C., and Shen, W.H. (2020). PTEN as a Guardian of the Genome: Pathways and Targets. Cold Spring Harb Perspect Med 10.

133. Brewer, T., Yehia, L., Bazeley, P., and Eng, C. (2022). Exome sequencing reveals a distinct somatic genomic landscape in breast cancer from women with germline PTEN variants. Am J Hum Genet 109, 1520-1533.

134. Cancer Genome Atlas Research, N., Weinstein, J.N., Collisson, E.A., Mills, G.B., Shaw, K.R., Ozenberger, B.A., Ellrott, K., Shmulevich, I., Sander, C., and Stuart, J.M. (2013). The Cancer Genome Atlas Pan-Cancer analysis project. Nat Genet 45, 1113-1120.

135. Zack, T.I., Schumacher, S.E., Carter, S.L., Cherniack, A.D., Saksena, G., Tabak, B., Lawrence, M.S., Zhsng, C.Z., Wala, J., Mermel, C.H., et al.

(2013). Pan-cancer patterns of somatic copy number alteration. Nat Genet 45, 1134-1140.

136. Mermel, C.H., Schumacher, S.E., Hill, B., Meyerson, M.L., Beroukhim, R., and Getz, G. (2011). GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. Genome Biol 12, R41.

137. Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30, 2114-2120.

138. Bushness, B. (2014). BBMap: A Fast, Accurate, Splice-Aware Aligner. In, United States.

139. Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29, 15-21.

140. Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 15, 550.

141. Kolde, R. (2019). pheatmap: Pretty Heatmaps. In.

142. Blighe, K.R., Sharmila; Lewis, Myles (2021). EnhancedVolcano: Publication-ready volcano plots with enhanced colouring and labeling. In.

143. Benesty, J.C., Jingdong; Huang, Yiteng; Cohen, Israel (2009). Pearson correlation coefficient. In Noise reduction in speech processing. (Springer), pp 37--40.

144. Schober, P., Boer, C., and Schwarte, L.A. (2018). Correlation Coefficients: Appropriate Use and Interpretation. Anesth Analg 126, 1763-1768.

145. Colaprico, A., Silva, T.C., Olsen, C., Garofano, L., Cava, C., Garolini, D., Sabedot, T.S., Malta, T.M., Pagnotta, S.M., Castiglioni, I., et al. (2016). TCGAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data. Nucleic Acids Res 44, e71.

146. Newman, A.M., Liu, C.L., Green, M.R., Gentles, A.J., Feng, W., Xu, Y., Hoang, C.D., Diehn, M., and Alizadeh, A.A. (2015). Robust enumeration of cell subsets from tissue expression profiles. Nat Methods 12, 453-457.

147. Gong, L., Govan, J.M., Evans, E.B., Dai, H., Wang, E., Lee, S.W., Lin, H.K., Lazar, A.J., Mills, G.B., and Lin, S.Y. (2015). Nuclear PTEN tumor-suppressor functions through maintaining heterochromatin structure. Cell Cycle 14, 2323-2332.

148. Chen, Y.Y., Ge, J.Y., Zhu, S.Y., Shao, Z.M., and Yu, K.D. (2022). Copy number amplification of ENSA promotes the progression of triple-negative breast cancer via cholesterol biosynthesis. Nat Commun 13, 791.

149. Lu, X., Yu, Y., Liao, F., and Tan, S. (2019). Homo Sapiens Circular RNA 0079993 (hsa_circ_0079993) of the POLR2J4 Gene Acts as an Oncogene in Colorectal Cancer Through the microRNA-203a-3p.1 and CREB1 Axis. Med Sci Monit 25, 6872-6883.

150. Permuth, J.B., Pirie, A., Ann Chen, Y., Lin, H.Y., Reid, B.M., Chen, Z., Monteiro, A., Dennis, J., Mendoza-Fandino, G., Group, A.S., et al. (2016). Exome genotyping arrays to identify rare and low frequency variants

associated with epithelial ovarian cancer risk. Hum Mol Genet 25, 3600-3612.

151. Gregory, K.J., Roberts, A.L., Conlon, E.M., Mayfield, J.A., Hagen, M.J., Crisi, G.M., Bentley, B.A., Kane, J.J., Makari-Judson, G., Mason, H.S., et al. (2019). Gene expression signature of atypical breast hyperplasia and regulation by SFRP1. Breast Cancer Res 21, 76.

152. Sonora, C., Mazal, D., Berois, N., Buisine, M.P., Ubillos, L., Varangot, M., Barrios, E., Carzoglio, J., Aubert, J.P., and Osinaga, E. (2006). Immunohistochemical analysis of MUC5B apomucin expression in breast cancer and non-malignant breast tissues. J Histochem Cytochem 54, 289-299.

153. Valque, H., Gouyer, V., Gottrand, F., and Desseyn, J.L. (2012). MUC5B leads to aggressive behavior of breast cancer MCF7 cells. PLoS One 7, e46699.

154. Rakha, E.A., Boyce, R.W., Abd El-Rehim, D., Kurien, T., Green, A.R., Paish, E.C., Robertson, J.F., and Ellis, I.O. (2005). Expression of mucins (MUC1, MUC2, MUC3, MUC4, MUC5AC and MUC6) and their prognostic significance in human breast cancer. Mod Pathol 18, 1295-1304.

155. Epping, M.T., Hart, A.A., Glas, A.M., Krijgsman, O., and Bernards, R. (2008). PRAME expression and clinical outcome of breast cancer. Br J Cancer 99, 398-403.

156. Xu, Y., Zou, R., Wang, J., Wang, Z.W., and Zhu, X. (2020). The role of the cancer testis antigen PRAME in tumorigenesis and immunotherapy in human cancer. Cell Prolif 53, e12770.

157. Naik, A., Thomas, R., Al-Khadairi, G., Bacha, R., Hendrickx, W., and Decock, J. (2021). Cancer testis antigen PRAME: An anti-cancer target with immunomodulatory potential. J Cell Mol Med 25, 10376-10388.

158. Cheng, C.C., Wooten, J., Gibbs, Z.A., McGlynn, K., Mishra, P., and Whitehurst, A.W. (2020). Sperm-specific COX6B2 enhances oxidative phosphorylation, proliferation, and survival in human lung adenocarcinoma. Elife 9.

159. Nie, K., Li, J., He, X., Wang, Y., Zhao, Q., Du, M., Sun, H., Wang, J., Lyu, J., Fang, H., et al. (2020). COX6B2 drives metabolic reprogramming toward oxidative phosphorylation to promote metastasis in pancreatic ductal cancer cells. Oncogenesis 9, 51.

160. Zhang, M., Han, Y., Liu, J., Liu, L., Zheng, L., Chen, Y., Xia, R., Yao, D., Cai, X., and Xu, X. (2020). Rbm24 modulates adult skeletal muscle regeneration via regulation of alternative splicing. Theranostics 10, 11159-11177.

161. Liu, Y., Teng, L., Fu, S., Wang, G., Li, Z., Ding, C., Wang, H., and Bi, L. (2021). Highly heterogeneous-related genes of triple-negative breast cancer: potential diagnostic and prognostic biomarkers. BMC Cancer 21, 644.

162. Baker, J., Riley, G., Romero, M.R., Haynes, A.R., Hilton, H., Simon, M., Hancock, J., Tateossian, H., Ripoll, V.M., and Blanco, G. (2010). Identification of a Z-band associated protein complex involving KY, FLNC and IGFN1. Exp Cell Res 316, 1856-1870.

163. Lefebvre, C., Bachelot, T., Filleron, T., Pedrero, M., Campone, M., Soria, J.C., Massard, C., Levy, C., Arnedos, M., Lacroix-Triki, M., et al. (2016). Mutational Profile of Metastatic Breast Cancers: A Retrospective Analysis. PLoS Med 13, e1002201.

164. Bylund, J., Bylund, M., and Oliw, E.H. (2001). cDna cloning and expression of CYP4F12, a novel human cytochrome P450. Biochem Biophys Res Commun 280, 892-897.

165. Doi, T., Ogawa, H., Tanaka, Y., Hayashi, Y., and Maniwa, Y. (2020). Bex1 significantly contributes to the proliferation and invasiveness of malignant tumor cells. Oncol Lett 20, 362.

166. Geng, H.T., Cheng, Z.W., Cao, R.J., Wang, Z.B., Xing, S.Z., Guo, C., Wang, F., Liu, C.M., Chen, S.S., and Cheng, Y.F. (2018). Low expression of BEX1 predicts poor prognosis in patients with esophageal squamous cell cancer. Oncol Rep 40, 2778-2787.

167. Fricker, L.D., Margolis, E.B., Gomes, I., and Devi, L.A. (2020). Five Decades of Research on Opioid Peptides: Current Knowledge and Unanswered Questions. Mol Pharmacol 98, 96-108.

168. Salhia, B., Kiefer, J., Ross, J.T., Metapally, R., Martinez, R.A., Johnson, K.N., DiPerna, D.M., Paquette, K.M., Jung, S., Nasser, S., et al. (2014).

Integrated genomic and epigenomic analysis of breast cancer brain metastasis. PLoS One 9, e85448.

169. Waluk, D.P., Schultz, N., and Hunt, M.C. (2010). Identification of glycine N-acyltransferase-like 2 (GLYATL2) as a transferase that produces N-acyl glycines in humans. FASEB J 24, 2795-2803.

170. Traber, M.G., and Atkinson, J. (2007). Vitamin E, antioxidant and nothing more. Free Radic Biol Med 43, 4-15.

171. Dashtiyan, A.A., Sepehrimanesh, M., Tanideh, N., and Afzalpour, M.E. (2017). The effect of endurance training with and without vitamin E on expression of p53 and PTEN tumor suppressing genes in prostate glands of male rats. Biochim Open 4, 112-118.

172. Ni, Y., and Eng, C. (2012). Vitamin E protects against lipid peroxidation and rescues tumorigenic phenotypes in cowden/cowden-like patient-derived lymphoblast cells with germline SDHx variants. Clin Cancer Res 18, 4954-4961.

173. Cheng, F., and Eng, C. (2019). PTEN Mutations Trigger Resistance to Immunotherapy. Trends Mol Med 25, 461-463.

174. Vidotto, T., Melo, C.M., Castelli, E., Koti, M., Dos Reis, R.B., and Squire, J.A. (2020). Emerging role of PTEN loss in evasion of the immune response to tumours. Br J Cancer 122, 1732-1743.

175. Briercheck, E.L., Trotta, R., Chen, L., Hartlage, A.S., Cole, J.P., Cole, T.D., Mao, C., Banerjee, P.P., Hsu, H.T., Mace, E.M., et al. (2015). PTEN is a negative regulator of NK cell cytolytic function. J Immunol 194, 1832-1840.

176. Chen, H., Chong, W., Yang, X., Zhang, Y., Sang, S., Li, X., and Lu, M. (2020). Age-related mutational signature negatively associated with immune activity and survival outcome in triple-negative breast cancer. Oncoimmunology 9, 1788252.

177. Chong, W., Wang, Z., Shang, L., Jia, S., Liu, J., Fang, Z., Du, F., Wu, H., Liu, Y., Chen, Y., et al. (2021). Association of clock-like mutational signature with immune checkpoint inhibitor outcome in patients with melanoma and NSCLC. Mol Ther Nucleic Acids 23, 89-100.

178. Yu, X., Abbas-Aghababazadeh, F., Chen, Y.A., and Fridley, B.L. (2021). Statistical and Bioinformatics Analysis of Data from Bulk and Single-Cell RNA Sequencing Experiments. Methods Mol Biol 2194, 143-175.

179. Dai, X., Cheng, H., Bai, Z., and Li, J. (2017). Breast Cancer Cell Line Classification and Its Relevance with Breast Tumor Subtyping. J Cancer 8, 3131-3141.