HUMAN BASAL-LIKE BREAST CANCER IS REPRESENTED BY ONE OF THE TWO MAMMARY TUMOR SUBTYPES IN DOGS

by

JOSHUA LEE WATSON

(Under the Direction of Shaying Zhao)

ABSTRACT

About 20% of breast cancers in humans are basal-like, a subtype that is often triple negative and difficult to treat. An effective animal model for basal-like breast cancer (BLBC) is currently lacking and urgently needed. To determine if spontaneous mammary tumors in pet dogs could meet this need, we subtyped canine mammary tumors and evaluated the dog-human molecular homology at the subtype level. We applied various subtyping strategies to the RNA-seq data of 236 canine mammary tumors sequenced by others and us, and consistently identified two molecularly distinct subtypes. One subtype is mostly basal-like and clusters with human BLBC in PAM50 classification, while the other subtype does not cluster with any human breast cancer subtype. Furthermore, the canine basal-like subtype recaptures key molecular features (e.g., cell cycle gene upregulation) and gene expression patterns that characterize human BLBC. However, about 33% of canine basal-like tumors are estrogen receptor negative (ER-) and progesterone receptor positive (PR+), which is rare in human breast cancer. Further analysis reveals that these ER-PR+ canine tumors harbor additional basal-like features, including upregulation of genes of interferon-γ response and of the

Wnt-pluripotency pathway. Interestingly, we observed an association of PGR expression with gene silencing in all canine tumors, and with the expression of T cell exhaustion markers (e.g., PDCD1) in ER-PR+ canine tumors. In summary, we identify a canine mammary tumor subtype that molecularly resembles human BLBC overall, and thus could serve as a vital spontaneous animal model of this devastating breast cancer subtype.

INDEX WORDS: Dog-human comparison, canine mammary tumor subtyping, RNA-seq, basal-like breast cancer, estrogen receptor, progesterone receptor, Wnt signaling, interferon-γ response, PAM50 subtyping

HUMAN BASAL-LIKE BREAST CANCER IS REPRESENTED BY ONE OF THE TWO MAMMARY TUMOR SUBTYPES IN DOGS

by

JOSHUA LEE WATSON

Bachelor of Science, Appalachian State University, 2015

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2023

HUMAN BASAL-LIKE BREAST CANCER IS REPRESENTED BY ONE OF THE TWO MAMMARY TUMOR SUBTYPES IN DOGS

by

JOSHUA LEE WATSON

Major Professor:	Shaying Zhao
Committee:	Kevin Dobbin
	Ying Xu
	Kosuke Funato

Electronic Version Approved:

Ron Walcott
Vice Provost for Graduate Education and Dean of the Graduate School
The University of Georgia
May 2023

DEDICATION

I find it difficult to name a single person to dedicate this dissertation to, as I find myself so grateful to so many for so much.

ACKNOWLEDGEMENTS

I would first like to acknowledge my principal investigator, Dr. Shaying Zhao, for her guidance over the course of my PhD, in both scientific rigor and leadership. I am grateful for all I was able to learn under her direction, as it made me a better scientist. I would also like to thank my committee members, Drs. Kevin Dobbin, Ying Xu, and Kosuke Funato for their assistance throughout this project.

I also would like to acknowledge my lab mates former and current, as they were able to help me in many ways. I thank Dr. Tianfang Wang for her key role in the initial work for this, and in particular, I thank Dr. Yuan Feng, who helped throughout, and assisted greatly with our outside endeavors.

Last but most certainly not least, I acknowledge my friends and family. Thank you all, so much, for everything you have done for me over these years. I thank my parents, Dr. Ricky Watson, and Mrs. Barbara Watson, for their kindness and support. I also thank my sister, Rebecca Watson, Esq., for her friendship. My friends are too myriad to list here, but their understanding and solidarity have helped me dearly. I also wish to thank my girlfriend Molly Stevens for her love throughout the writing.

I wish words were sufficient to display my gratitude.

TABLE OF CONTENTS

LIST OF TABLES

Page

LIST OF FIGURES

Page

CHAPTER 1

INTRODUCTION AND LITERATURE REVIEW

STRUCTURE

The first chapter of this dissertation will begin with a full review of the basal-like breast cancer subtype, beginning with epidemiology, then an examination of the subtype's defining features- histology, metastasis, and genomic landscape (epigenetic, mutation, copy number alterations, and gene expression), followed by defining the current problems in its treatment. Next, it will give a review of canine cancers with respect to their use as a comparative oncology model for breast cancer research. It will explain the logistic and genomic advantages of using canine mammary tumors for comparative oncology research and highlight the current information missing from the field that hampers the use of canine mammary tumors for this purpose and define our goals in performing this research.

The second chapter of this dissertation, titled "Human basal-like breast cancer is represented by one of the two mammary tumor subtypes in dogs", will describe our research performed to address the problem explained in the first chapter, in the form of a manuscript currently under review in the journal *Breast Cancer Research*. Briefly, we identified a spontaneous canine mammary tumor subtype that molecularly resembles human basal-like breast cancer overall, and thus could serve as a vital subtype-specific spontaneous animal model of this devastating breast cancer subtype. Our study also

sheds light on the dog-human difference in the mammary tumor histology and the hormonal cycle.

The final chapter of this dissertation will sum up our conclusions and what questions in the field this study answers, as well as highlight the new questions our results bring up which remain viable future directions for study.

BASAL-LIKE BREAST CANCER

Breast cancer affects roughly 300,590 and kills 43,700 people annually in the United States, making it the fourth deadliest cancer in the US, even though only women compose over 99% of diagnosed cases [1]. Breast cancer is a heterogenous disease, consisting of five well-established molecularly distinct subtypes, as defined by the Prediction Analysis of Microarray 50 (PAM50) gene panel: Luminal A, Luminal B, HER2-enriched, Normal-like, and Basal-like [2]. Between 70-80% of Basal-like breast cancers (BLBC) are “triple negative breast cancer” (TNBC), a clinical diagnosis meaning they lack the receptors ER and PR and have lower ERBB2 [3, 4]. Although the two terms are often used interchangeably, this is not accurate, as non-basal-like TNBC and BLBC with expression of one of the three previously mentioned receptors (typically ESR1 in humans) exist [3, 4]. BLBC is generally considered to have the worst prognosis for long term survival, with a five-year survival rate of 77% compared to an overall average above 90% [5]. The five-year survival rate of BLBC is bimodal, with a large proportion of BLBC patients having a very high risk of recurrence and mortality within five years, and a small subpopulation that have very good health outcomes [6]. Additionally, BLBC

is the subtype most likely to affect younger women, and is significantly more common in African American women than in any other race [7]. Further compounding the problem of BLBC is the lack of effective treatments. While there are drugs available to target ER and ERBB2, BLBC lacks these, and thus the only treatment options are surgery, radiation, and chemotherapy [8]. BLBC possesses a variety of distinct features that distinguish it from other breast cancers and hamper its treatment, highlighting the need for a more effective model for its research.

Features

Histology and metastasis

Histologically, BLBCs are more likely to have high grade (less cellular differentiation), with one study having 85.5% of BLBCs pathologically graded at level 3 [9]. They also have a higher mitotic index than other subtypes. Invasive ductal carcinoma is the most common histological type in BLBC (and in breast cancers in general), but it can also be found in invasive lobular carcinoma, mixed carcinoma, and a multitude of others [10]. Lymphocytic infiltration is also common in BLBC tumors, even though it is associated with better disease-free survival [6, 9]. Although BLBC is not significantly more likely to have distant metastases than other subtypes, it has a significantly higher fatality rate when metastases do occur [11]. Distant metastases in BLBC tend to happen in the brain, liver, and lung, but are more rare in the bone [11].

Genomic landscape

BLBC is highly distinct on the molecular level when compared to other breast cancers, but it is also a heterogenous subtype [12]. Nevertheless, it has several

defining traits, aside from the previously mentioned lack of ER and PR and low ERBB2. The expression of the PAM50 genes in BLBC separates them from any other breast cancer subtype [2]. A wide variety of genomic pathways are enriched in BLBC, supporting functions such as the maintenance of pluripotency, proliferation, cell cycle, and invasion [4, 6, 13, 14]. These include Wnt (both canonical and non-canonical), IL6/JAK/STAT3, and epithelial-mesenchymal transition, among others [15-19]. To the contrary, there are many pathways with tumor suppressing function that are often negatively altered in BLBC, such as those involved in DNA damage response and cell cycle regulation [20].

As lymphocytic infiltration is common in BLBC, several immune signatures are commonly detected, such as T cell receptor and Natural Killer (NK) cell pathways, among others [3]. Importantly, BLBCs very often have increased expression of PD-L1, an immune checkpoint protein that binds to PD-1 on T cells (which is also common on tumor infiltrating lymphocytes detected in BLBC) and prevents the immune system from killing the cancer cell [3]. High expression of PD-1 and PD-L1 is one hallmark of T cell exhaustion [21]. In this state, T cells become less able to kill the cells they are activated to target (e.g., a BLBC cell) [21]. BLBC almost always has high expression of the keratin genes 5, 6, and 17 (KRT5/6/17) [13]. These keratins, known as “basal cytokeratins”, are what the subtype name “basal-like” derives from. High expression of epidermal growth factor receptor (EGFR) is also a hallmark of BLBC. EGFR drives multiple oncogenic processes, such as DNA synthesis, cell proliferation, and metastasis

[22]. While EGFR is a specifically targetable gene (by the antibody drug cetuximab), recent clinical trials targeting it in BLBC have shown limited results [8].

The mutation landscape of BLBC varies greatly. While BLBC generally carries a higher tumor mutation burden (TMB; defined as the number of somatic non-synonymous mutations and small indels per megabases of coding sequence) than other subtypes, only a few mutations are commonly detected in BLBC across studies [13, 20]. Notably, BLBC is consistently highly mutated in TP53 (over 80% of cases), often leading to lost function of TP53 [13, 20]. PIK3CA is the second most commonly mutated gene in BLBC, at 7% of cases, but even then, this percentage is much lower than PIK3CA mutations in other subtypes [13]. Luminal A, Luminal B, and HER2-enriched cancers have PIK3CA mutations in 49%, 32%, and 42% of samples respectively [13]. RB1 is also a significantly mutated gene in BLBC and is often co-mutated with TP53 [13]. Other significantly mutated genes in BLBC are TRIM6-TRIM34, ATP2B2, and PNPLA3 [13]. Almost all hereditary breast cancer, which occurs due to germline mutations in BRCA1/2, develops into BLBC as well, although not all BLBC are hereditary cancers [3, 23]. Germline mutations in BRCA1/2 have been detected in up to 19.5% of BLBC cancers [3].

One of the most common traits of BLBC is high genomic instability with regard to copy number alteration, with a major study showing the median altered genome fraction in these cancers at roughly 50% [20]. The gene MYC (located on 8q) is focally amplified in between 40-65% of BLBC cases, and interestingly, ERBB2 is also frequently amplified in BLBC, even though its mRNA expression is typically low [13, 24].

KRAS and MDM2 are also frequently amplified [13, 24]. As for deletions, the genes INPP4B and PTEN are deleted in 30% and 35% of BLBC samples, respectively [13]. Many of the previously-mentioned recurrently mutated genes (TP53, BRCA1/2, and RB1) are also recurrently deleted in BLBC at rates higher than 30% [24]. More broad copy number alterations in BLBC include significant gains on 1q and 10p, but significant deletions in 8p and 5q [13].

Treatments

As mentioned previously, BLBC has limited treatment options due to the lack of targetable hormone receptors and low HER2 expression. While other breast cancer types enjoy a wide range of treatment options, radiation and especially chemotherapy remain the principal treatments for BLBC [8]. Some subsets of BLBC can be targeted by certain, non-BLBC specific drugs. For example, cancers with a BRCA1/2 mutation can be targeted with PARP (poly-(ADP-ribose) polymerase protein) inhibitors, which prevents the protein PARP from recruiting mutated (and thus non-functional) BRCA1/2 to attempt DNA damage repair [8]. Unfortunately, cancer cells can become resistant to PARP inhibitors by a variety of methods [8]. Another treatment option for BLBC is immunotherapy, the efficacy of which can be predicted by TMB [25]. Recently, the immune checkpoint (PD-1 and PD-L1) inhibitors atezolizumab and pembrolizumab have been approved for triple-negative breast cancer treatment [8]. These drugs have been shown to be effective in extending survival for a matter of months, but patients still do not receive a significant long-term benefit [8]. The dearth of effective treatments for BLBC is indicative of a gap in the current translational research models.

Translational research models

Mouse models of breast cancer are logistically popular, due to a relatively lower cost to keep and short generation times [26]. However, they have some differences with humans that limit translational research. First, mice can tolerate a much higher dose of most drugs relative to body weight than humans can, leading to therapeutic doses that would be toxic for a human patient [26, 27]. Lab mice tend to be kept in totally sterile environments and have stunted or non-existent immune systems, which does not reflect the immune microenvironment in which human cancers develop [27]. The lack of immune response in mice also hampers the understanding of how immunotherapy will work in a human patient. A missing immune system is also a principal limitation of patient derived xenograft models, as they do not have natural tumor-host interactions [28]. Ultimately, the combination of these problems means that over 90% of cancer drugs fail to achieve regulatory approval in human clinical trials [26, 27, 29]. This gap highlights a critical need for a spontaneous preclinical model of cancer to better reflect the essence of BLBC development.

CANINE MAMMARY TUMORS

Canine mammary tumors offer a valuable solution to the problem of translational models for BLBC research and have several critical logistical advantages. First, there are over 80 million pet dogs in the United States alone, and these pets live in the same area and home as their human masters, being exposed to the same environmental carcinogenic factors [30]. Second, mammary tumors are common in intact (unspayed) bitches, at an estimated annual incidence rate at 198 per 100,000 [31], which is comparable to the rate of 125 per 100,000 for breast cancer in women in the United

States [32]. Up to 26% of unspayed bitches may develop a malignant mammary tumor over the course of their lives [31]. Third, canine mammary tumors spontaneously occur in animals with an intact immune system, overcoming many limitations of traditional cancer models such as cell lines, patient derived xenografts, and genetically modified rodent models [33, 34]. This pattern of tumor development is more in line with the essence of human cancer development. Dog hepatic (liver) enzymes are also more homologous to those in humans than mouse liver enzymes are, meaning dogs are more likely to process drugs in a similar way to humans [35]. Furthermore, dogs as a research model benefit greatly from a well-sequenced reference genome [36]. A great deal of research has recently been done using canine cancers of many types as translational research models [37-40]. Further benefitting the use of canine cancer as a model for humans is the National Cancer Institute's (NCI) recently funded project, the Integrated Canine Data Commons (ICDC), a database for canine cancer data styled after The Cancer Genome Atlas' (TCGA) data portals. Canine mammary tumors in particular have several molecular features that suggest they could be effective preclinical models of BLBC to accelerate bench-to-bedside translation.

Molecular features

Histologically, canine mammary tumors are more diverse than human breast cancers. Adenomas and carcinomas with only a single proliferating cell lineage (most commonly epithelial, but possibly also myoepithelial cells) are known as "simple" [29, 41]. The histology of these tumors (which are slightly over half of canine mammary tumors) is more like that of humans. A significant proportion (around one third) of canine mammary tumors have "complex" histology, in which multiple cell lineages, such

as both epithelial and myoepithelial cells, proliferate [29, 41]. Canine mammary tumors may also be mixed carcinomas (~10%), which are complex carcinomas that also have mesenchymal tissue cells (such as cartilage) proliferating as well [41].

Canine mammary tumors (among several canine cancers) have been shown to have a lower TMB than their human counterparts, but as in human breast cancer, TP53 mutations are strongly correlated with higher TMB [42]. PIK3CA is the most commonly mutated gene in canine mammary tumors, especially at the hotspot H1047, but interestingly lacks mutations at the hotspot E542/545 [42]. Overall, the PI3K-Akt pathway is altered (by either mutation or copy number alteration) in over 55% of canine mammary tumors [43]. Other significantly commonly mutated genes in canine mammary tumors were KAT6B, KRAS, AKT1, and NRIP1 [42]. Among these mutations, AKT1 is the only one found exclusively in complex carcinomas, suggesting a possible tissue-specific role for this mutation [43].

Somatic copy number alteration is also common in canine mammary tumors across studies and platforms, with three studies on different platforms showing malignant mammary tumors having a significantly higher proportion of the genome altered than benign tumors [33, 43, 44]. Despite these studies showing similar proportions of the genome being altered, the actual genes affected differ based on the sequencing technology used. A large-scale study using whole exome sequencing data found amplifications in genes such as EGFR and HRAS and deletions in PTEN, but unexpectedly did not find a significant amplification of MYC [43]. Other studies (using canine-genome specific array hybridization for copy number alterations) showed that

malignant and simple carcinomas are likely to harbor amplifications in MYC and deletions in PTEN [33, 44].

Missing link

For all the promise and similarities canine mammary tumors have, there are three key roadblocks that must be addressed for its effective use as a model. First, about half of canine cancers have complex or mixed histology, in which multiple cell lineages (e.g., epithelial and myoepithelial cells) proliferate [33, 41, 45]. Human breast cancers, to the contrary, almost always derive from epithelial cells, like canine simple carcinomas. The few exceptions to this (such as adenomyoepithelioma) occur in <1% of breast cancers [33, 46, 47]. For example, epithelial-myoepithelial carcinoma, which is the closest human equivalent of canine complex carcinoma, has less than 50 recorded cases [48]. To accurately use the canine mammary tumor as a model for human breast cancer, this difference in cell type composition must be accounted for.

Second, the pattern of reproductive hormone production (estrogen and progesterone specifically) and estrus cycle differs greatly between humans and dogs. In humans, the whole cycle is roughly 28 days, with an approximately two week long luteal phase, in which the corpus luteum forms on the ovary and releases progesterone (and some estrogen) [49]. Without fertilization, it degrades after roughly two weeks and stops producing progesterone, triggering menstruation [49]. Canines, on the other hand, typically are diestrous, meaning they go through their reproductive cycle only twice per year, and some components of the cycles last longer. Specifically, the luteal (or diestrus) phase in dogs lasts approximately two months, nearly the same length as pregnancy [50]. This leads to the canine corpus luteum producing high levels of

progesterone for far longer than a human would be exposed to [50, 51]. This is important in breast cancer, as the presence of progesterone has been shown to induce expression of PGR [50, 52].

Previous studies have been done using the PAM50 gene expression pattern to compare canine mammary tumors to human breast cancers, with differing results. A study from Liu et al. 2014 showed that, in 82 out of 100 randomly sampled hierarchical clusters, canine simple carcinomas and human BLBC cluster together away from any other cancer [33]. This study, however, has a small sample size. To the contrary, another study from Bergholtz et al. 2022 used a larger set of canine RNA-seq to perform the PAM50 subtype assignment, and found that canine tumors assigned to the Luminal A subtype were far more similar to human Luminal A cancers than the canine Basal-like tumor were to human BLBC [53, 54]. The PAM50 gene expression pattern is useful but imperfect for the dog, especially considering that some genes (such as PGR) might be expressed differently than in humans, and some PAM50 genes (such as NAT1) do not even exist in the canine genome [55]. The discrepancy between these varied studies shows that independently developed, genome-wide expression subtyping is necessary for the dog as a cancer model.

Larger sets of high-quality RNA-seq data of canine mammary tumors have been recently released, allowing for larger scale mRNA expression subtyping studies [43]. The goal of this work was to develop independent subtypes based on mRNA expression patterns in canine mammary tumors, and with those subtypes, directly compare their expression patterns to human breast cancer subtypes.

CHAPTER 2

HUMAN BASAL-LIKE BREAST CANCER IS REPRESENTED BY ONE OF THE TWO MAMMARY TUMOR SUBTYPES IN DOGS[1]

---

1. Watson J, Wang T, Ho K-L, Feng Y, Dobbin KK, Zhao S: *Human basal-like breast cancer is represented by one of the two mammary tumor subtypes in dogs*. bioRxiv 2023:2023.2003.2002.530622.

Currently (as of 3/27/2023) under editorial review in *Breast Cancer Research*.

ABSTRACT

Background

About 20% of breast cancers in humans are basal-like, a subtype that is often triple negative and difficult to treat. An effective translational model for basal-like breast cancer (BLBC) is currently lacking and urgently needed. To determine if spontaneous mammary tumors in pet dogs could meet this need, we subtyped canine mammary tumors and evaluated the dog-human molecular homology at the subtype level.

Methods

We subtyped 236 canine mammary tumors from 3 studies by applying various subtyping strategies on their RNA-seq data. We then performed PAM50 classification with canine tumors alone, as well as with canine tumors combined with human breast tumors. We investigated differential gene expression, signature gene set enrichment, expression association, mutational landscape, and other features for dog-human subtype comparison.

Results

Our independent genome-wide subtyping consistently identified two molecularly distinct subtypes among the canine tumors. One subtype is mostly basal-like and clusters with human BLBC in cross-species PAM50 classification, while the other subtype does not cluster with any human breast cancer subtype. Furthermore, the canine basal-like subtype recaptures key molecular features (e.g., cell cycle gene upregulation, TP53 mutation) and gene expression patterns that characterize human BLBC. It is enriched in histological subtypes that match human breast cancer, unlike the other canine

subtype. However, about 33% of canine basal-like tumors are estrogen receptor negative (ER-) and progesterone receptor positive (PR+), which is rare in human breast cancer. Further analysis reveals that these ER-PR+ canine tumors harbor additional basal-like features, including upregulation of genes of interferon-γ response and of the Wnt-pluripotency pathway. Interestingly, we observed an association of PGR expression with gene silencing in all canine tumors, and with the expression of T cell exhaustion markers (e.g., PDCD1) in ER-PR+ canine tumors.

Conclusions

We identify a canine mammary tumor subtype that molecularly resembles human BLBC overall, and thus could serve as a vital spontaneous animal model of this devastating breast cancer subtype. Our study also sheds light on the dog-human difference in the mammary tumor histology and the hormonal cycle.

Keywords

Dog-human comparison, canine mammary tumor subtyping, RNA-seq, basal-like breast cancer, estrogen receptor, progesterone receptor, Wnt signaling, interferon-γ response, PAM50 subtyping

BACKGROUND

Human breast cancer is heterogeneous, consisting of well-established molecularly distinct subtypes [13, 20, 56-63]. One of these subtypes is basal-like breast cancer (BLBC; human BLBC will be referred to as hBLBC hereafter), which makes up roughly 15-20% of human breast cancers and has the worst prognosis of all subtypes [13, 20, 56-63]. About 70% of hBLBCs are triple negative, expressing neither estrogen receptor (ER) nor progesterone receptor (PR) and without HER2 amplification or overexpression [13, 20, 56-63]. These cancers also tend to have increased rates of cell proliferation and metastasis [13, 20, 56-63]. All these highlight the need for an effective translational model for hBLBC, which is critically missing at present [26, 27, 34].

Mammary cancers in pet dogs naturally occur in animals with an intact immune system [33, 34], overcoming many limitations of traditional cancer models such as cell lines and genetically modified rodent models. These canine cancers more accurately emulate human breast cancers in etiology, complexity, heterogeneity, behavior, treatment, and outcome [29, 33, 34, 64]. They are also common in bitches, with an annual incidence rate estimated at 198 per 100,000 [31], which is comparable to the rate of 125 per 100,000 for breast cancer in women in the United States [32]. Mammary cancer is especially common in bitches that are not spayed or are spayed after the second estrus, with the risk for malignant tumor development expected at 26% [31]. Thus, canine mammary tumors have the potential to serve as a much-needed translational model of

hBLBC, effectively bridging a current gap between preclinical models and human clinical trials to accelerate bench-to-bedside translation.

The effective use of the canine model is, however, complicated by issues including the dog-human hormonal cycle difference, e.g., the luteal phase lasts ~14 days for humans but ~2 months for dogs. Another difference is histology. About 50% of canine mammary tumors are complex or mixed, with multiple cell lineages (e.g., epithelial and myoepithelial cells) proliferating [33, 41, 45]. These histologies (e.g., adenomyoepithelioma) are, however, are very rare (<1%) in human breast cancers [33, 46, 47]. It remains unknown how these differences shape the molecular homology and difference between canine and human mammary tumors.

The same as human breast cancers, spontaneous canine mammary cancers are heterogeneous and consist of distinct subtypes [41, 45]. Thus, subtype level dog-human comparison is needed to evaluate the dog-human homology. Canine mammary cancers have been histopathologically and clinically subtyped, as well as molecularly subtyped with immunohistochemical markers established for human breast cancer (anti-ER, PR, HER2, -CK 5/6 and -CK14) [41, 45, 65, 66]. However, to our knowledge, canine mammary tumors have not been independently subtyped by genome-wide molecular studies.

For dog-human comparison, our previous study indicates that one histological subtype, simple carcinoma, molecularly resembles hBLBC in cross-species PAM50 classification with human and canine tumors [33]. However, this study is limited by its small sample size. Another group has performed PAM50 classification on the RNA-seq data recently published for 154 canine mammary tumors [54], and reported a higher homology between canine and human luminal A tumors than between canine and human basal-like tumors [53]. This study, however, did not perform cross-species PAM50 to directly compare human and canine tumors. Moreover, many of the canine luminal A tumors are complex and mixed tumors, histologically differing from the vast majority of human luminal A tumors [33, 46, 47].

To address these discrepancies and deficiencies, we set out to independently subtype canine mammary tumors using RNA-seq data of 236 tumors [33, 54, 67], and then perform dog-human comparison at the subtype level, as described below.

MATERIALS AND METHODS

Data collection

Canine RNA-seq data was downloaded from the Sequence Read Archive (SRA) database, including data of 154 mammary tumors from PRJNA489087 (excluding 4 metastatic osteosarcoma and fibrosarcomas) [54] and of 63 mammary tumors from PRJNA561580 [67]. RNA-seq data of 25 mammary tumor samples sequenced in house [33] were also included (PRJNA203086 and PRJNA912710). Human breast cancer

RNA-seq data were downloaded from the National Cancer Institute (NCI) Genomic Data Commons (GDC) database, and the PAM50 classification of these cancers was obtained from the cBioportal database [68]. Gene expression microarray data of two canine mammary tumor studies (GSE20718 and GSE22516) and one human breast cancer (GSE20685) were downloaded from the Gene Expression Omnibus (GEO) database [69-71], and processed with affyPackage [72]. Other information was obtained from relevant publications of these studies. Canine genome canFam3.1 and gene annotation canFam3 1.99 GTF were downloaded from the Ensembl database. Canine mutation data and tumor mutation burden (TMB) values from whole exome sequencing analysis were obtained from a previous publication [42].

Canine sample collection and RNA-seq

Fresh-frozen (FF) canine tissues and spontaneous tumors were obtained from the canine tissue archive bank at Ohio State University, as previously described [73]. Samples were collected from client-owned dogs that developed the disease spontaneously, under the guidelines of the Institutional Animal Care and Use Committee for use of residual diagnostic specimens and with owner informed consent. The case information was provided by the tissue bank. The research received the ethical approval from the Institutional Animal Care and Use Committee.

Cryosectioning of FF tissues, H&E staining, and cryomicrodissection were performed as described [73] to enrich tumor cells for tumor samples. Genomic DNA and RNA were

extracted from the dissected tissues using the AllPrep DNA/RNA Mini Kit (cat. no. 80204) from QIAGEN (Germantown, MD, USA). Only samples with a 260/280 ratio of ~2.0 (RNA) and showing no degradation and other contaminations were subjected to further quality control with qRT-PCR analysis with a panel of genes as previously described [73, 74]. RNA-seq libraries were constructed using KAPA Stranded mRNA-Seq Kit. The samples were subjected to 75 or 125 bp paired-end sequencing using Illumina HiSeq 2500 or NextSeq 500 at Georgia Genomics Facility.

Canine RNA-seq data quality control (QC) and processing

Canine RNA-seq data were processed as described [73-75]. Briefly, RNA-seq read pairs were mapped to the canine reference genome canFam3 using HISAT2 (version 2.21) [76]. Concordantly (for paired-end RNA-seq data only) and uniquely mapped pairs were identified and were used to calculate the mapping rate of each sample. Such pairs with at least one read with ≥1 bp overlapping a coding sequence (CDS) region of the canFam3 1.99 GTF annotation were used to calculate the CDS-targeting rate.

Quality control of canine RNA-seq data was performed as described [75]. First, MultiQC [77] (version 1.5) was used to examine GC content and duplicate level. Base quality distribution before and after Trimmomatic trimming was also examined. Second, the distributions of per sample read-pair total amount, mapping quality, and CDS targeting rate were examined to identify and exclude samples that fail to meet the

cutoffs. A total of 6 canine RNA-seq samples from PRJNA561580 failed the QC and were excluded from further analysis (Figure 3.1A-E).

For each sample that passed QC measures, Subread (version 2.0.0) [78] was used to identify read pairs that are uniquely and, for paired-end RNA-seq, concordantly mapped to the exonic regions of the canFam3 1.99 GTF annotation, the sum of which yields raw RNA-seq counts. Cufflinks version 2.2.0 [79] was used to calculate the FPKM (fragments per kilobase of exon per million mapped) value of each gene in each sample, which was then converted to TPM (transcript per million). For studies that combine RNA-seq data from multiple sources, comBat [80] was applied to correct batch effect for TPM values, and ComBat-seq [81] was used to correct batch effect for mapped RNA-seq read count values.

ER, PR, and HER2 status

Samples with an ESR1 or PGR expression level of $\mathrm{FPKM} \leq 1$ and $\mathrm{FPKM} > 1$ were classified as ER or PR negative and positive respectively. Samples with an ERBB2 expression level of $\mathrm{FPKM} \leq 35$ and $\mathrm{FPKM} > 35$ were classified as HER2 not enriched and enriched, respectively.

Canine mammary tumor subtyping

A gene was selected if it has an official gene name associated with its Ensembl gene ID in the canFam3 1.99 GTF annotation and is expressed (with $\mathrm{FPKM} \geq 1$ in at least one

sample across a cohort or study). This yields 13,416 genes in the discovery set (paired-end RNA-seq data) [33, 54] and 13,608 genes in the validation set (single-end RNA-seq data) [67] (see Results). The NMF R package [82] was then applied on all of these selected genes, as well as on the top 5000, 2000, 1000, and 500 most variable genes among them, with 30 runs for the rank determination. These analyses consistently divided 143 out of 179 samples of the discovery set into two subtypes. For validation, K-means clustering, consensus clustering, and hierarchical clustering via R packages stats, ConsensusClusterPlus, and pvclust respectively [83-85] were used to subtype the 143 samples using the top 10% most variable genes. The same process was repeated to subtype the samples in the validation set, using only the top 2000 most variable genes for NMF, as this gene set gives the most separation using basis components from NMF.

PAM50 classification of canine and human tumors

A total of 43 canine homologues of the 50 PAM50 genes were identified in the canFam3 1.99 annotation file. These 43 genes were then used to perform PAM50 classification with canine samples alone, as well as with canine and human combined samples, for RNA-seq studies [33, 54, 67]. For microarray studies [70, 71], 40 of the PAM50 genes were identified based on the probes and used to classify canine samples alone and canine-human combined samples. The PAM50 subtypes of human breast cancer samples were downloaded from the cBioportal database or relevant publications. For canine and human combined sample PAM50 clustering for RNA-seq studies, 60 human

tumors were randomly sampled from the cancer genome atlas (TCGA) breast cancer study for each of the luminal A, luminal B, hBLBC, and HER2-enriched subtypes. These samples, along with all 27 human normal-like tumors, were merged with all 143 subtyped canine tumors for PAM50 classification analysis. The clustering dendrogram was then cut at the minimum number of clusters that maximally separate hBLBC tumors from hLumA tumors using the R package dendextend [86]. The number of tumors of each canine or human subtype in each cluster was counted. If a cluster contains the majority of tumors of a human subtype as well as the majority of a canine subtype, the human and canine subtypes were considered matched. This process was repeated 100 times to ensure each human tumor was sampled at least once.

Multidimensional scaling on the Euclidean distance matrix was performed for each of the 100 random samplings, from which the mahalanobis distance between the centers of any two subtypes was calculated using the R package 'GenAlgo' v2.2.0 [87].

The above process was repeated for microarray studies [69-71], except that only 40 canine homologues of the 50 PAM50 genes were identified, and 30 tumors per subtype were randomly sampled from the human dataset [69].

Differentially expressed (DE) genes and gene set enrichment analysis (GSEA)

DESeq2 [88] was used to identify DE genes between subtypes or subgroups. Genes with ≥ 2 fold change in read count and the Benjamini-Hochberg (BH) adjusted $p \leq 0.05$

were considered differentially expressed. Enriched functions of DE genes were investigated with the GSEA [89] and DAVID [90] web tools. Pathway and signature gene sets were acquired from previous publications [73, 74, 91]. Single sample GSEA (ssGSEA v. 10.1.0) was performed using Genepattern [92].

Correlation and other statistical analyses

Genes with $\mathrm{FPKM} > 1$ in at least 10% of the samples of a subtype were chosen for ESR1 or PGR correlation analysis. PGR was excluded from hBLBC as <10% of samples have an $\mathrm{FPKM} > 1$. Positively or negatively correlated genes were defined as those with BH adjusted $\mathrm{p} \leq 0.05$ and correlation coefficient $|\mathrm{R}| \geq 0.3$ for both Pearson and Spearman correlation analysis. The software enrichR [93] was used to identify transcription factors targeting each group of significantly correlated genes. Wilcoxon rank sum tests were used for statistical comparison between subtypes or subgroups.

RESULTS

Canine mammary tumors consist of two distinct subtypes

We first performed non-negative matrix factorization (NMF) [82], a widely used subtyping strategy, to subtype 179 canine mammary tumors with paired-end RNA-seq data (the discovery set), after combining 154 tumors sequenced by Kim et al. [54] and 25 tumors sequenced by us [33] followed by batch correction. NMF subtyping was repeated with all 13,416 genes that are expressed in at least one tumor, as well as with the top 5000, 2000, 1000, and 500 most variable genes among the 13,416 expressed

genes. The analysis consistently clustered 143 of 179 tumors into two subtypes (Figure 2.1). To validate this, we also subtyped these tumors using other popular strategies, including K-means, consensus clustering, and hierarchical clustering via multiscale bootstrap resampling (Figure 3.2A-C) [83, 84]. These strategies consistently identified the same two subtypes as the NMF approach (Figure 3.2D).

We then performed the same subtyping analyses on the other large study of canine mammary tumors (n=57) (the validation set), whose RNA-seq data are single-end [67], differing from the discovery set. These tumors also clustered into two subtypes, consistent with the discovery set (Figure 3.2E-F).

The two subtypes differ in tumor histology and invasiveness. One subtype is significantly ($\mathrm{p} = 0.007$) enriched in simple adenomas/carcinomas (where only one cell lineage proliferates prominently), while the other subtype is enriched in complex or mixed adenomas/carcinomas ($\mathrm{p} < 0.001$) (where more than one cell lineages are proliferating prominently) (Figure 2.1) [33, 41, 54]. Moreover, the simple adenomas/carcinomas-enriched subtype contains significantly ($\mathrm{p} < 10^{-6}$) more cases with lymph node invasion (Figure 2.1). Interestingly, the other subtype contains significantly ($\mathrm{p} = 0.0014$) more Maltese dogs (Figure 2.1).

The two subtypes display distinct molecular features. Among the top most mutated genes in this cohort [42], one subtype is enriched in PIK3CA hotspot mutation

H1047R/L ($p = 0.0034$) and KAT6B mutation ($p = 0.036$), while the other subtype (simple adenomas/carcinomas-enriched and with more lymph node invasion) is enriched in TP53 mutation ($p = 0.043$) (Figure 2.1). However, we did not observe any significant difference in KRAS mutation and TMB between the two subtypes (Figure 2.1). For pathways, the two subtypes differ in mutations and copy number alterations of genes in PI3K signaling ($p = 0.0039$) (Figure 2.1), the most altered pathway in canine mammary tumors [42].

Other molecular differences between the two subtypes include PAM50 classification, ER and PR expression status, and others, and will be described in more details below.

Canine and human basal-like tumors cluster together in PAM50 classification

We performed PAM50 classification [2, 94] on the 179 canine tumors from the discovery set. About 74% of the tumors were classified as either the basal-like (n=62, 35%) or luminal A (n=71, 39%) subtype (Figure 2.2A; Table 2.1). Importantly, 90% of the basal-like tumors belong to one of the two subtypes shown in Figure 2.1, while 90% of the luminal A tumors belong to the other subtype. For this reason and reasons described below, the two subtypes shown in Figure 2.1 are named canine basal-like mammary tumor (cBLMT) and canine non-basal-like mammary tumor (cNBLMT) respectively.

To quantitatively assess the canine-human homology at the subtype level, we performed cross-species PAM50 classification as described [33]. Briefly, we randomly

sampled 60 human tumors from each of the luminal A, luminal B, HER2-enriched, and basal-like subtypes from the TCGA RNA-seq study [13, 57]. These, along with all 27 normal-like tumors in TCGA, amount to 267 human tumors covering all five intrinsic subtypes. We then performed PAM50 clustering on these human tumors together with all 143 subtyped canine tumors shown in Figure 2.1. This analysis was repeated 100 times, ensuring that each TCGA tumor was sampled at least once.

In 66 of 100 random sampling analyses, cBLMTs and hBLBCs clustered together and away from tumors of any other canine or human subtype (Figure 2.2B-C). To the contrary, in 87 of 100 random sampling analyses, cNBLMTs, the other canine subtype, did not cluster with tumors of any human subtype (Figure 2.2B-C). These observations are supported by multidimensional scaling of each cross-species PAM50 classification, as shown by the example provided in Figure 2.2D. We then calculated the mahalanobis distance [87] between the centers of canine and/or human subtypes for each of the 100 random sampling analyses. The distributions clearly indicate that the mahalanobis distances between hBLBC and cBLMT are significantly shorter than those between hBLBC and human luminal A (hLumA) or cNBLMT, as well as those between hLumA and cBLMT or cNBLMT (Figure 2.2E). These results support that cBLMT molecularly resembles hBLBC, but cNBLMT molecularly differs from hLumA.

To validate this finding, we attempted to conduct the same analyses on the 57 tumors from the validation set [67]. PAM50 analysis of these canine tumors alone indeed

classified a majority of the tumors as basal-like (n=21) or luminal A (n=17), consistent with the discovery set (Figure 3.3A). Moreover, many of the basal-like canine tumors have a PAM50 gene expression pattern that closely matches hBLBC (Figure 3.3B). However, likely due to having single-end RNA-seq data, these canine tumors could not co-cluster with human breast tumors (whose RNA-seq data are paired-end) in cross-species PAM50 classification even after batch correction.

We next performed the same analyses on the gene expression microarray data of two canine studies (n=27; n=13) and of one human study (n=327) [69-71]. Consistent with the RNA-seq analysis described above, canine-only PAM50 clustering classified most of these canine tumors as either basal-like or luminal A (Figure 3.3C). Moreover, cross-species PAM50 classification clustered cBLMTs and hBLBCs together in a majority of random sampling analyses (Figure 3.3C-D), further supporting the molecular homology between cBLMT and hBLBC.

The cBLMT subtype captures key molecular features of hBLBC

We identified differentially expressed (DE) genes between cBLMT and cNBLMT. The 1,123 genes upregulated in cBLMT are significantly enriched in cell cycle (e.g., DREAM targets, G2M checkpoint) and other functions that characterize hBLBC [13], as well as in genes that are known to be upregulated in hBLBC [91] (Figure 2.3A; Table 2.1). Conversely, the 497 genes downregulated in cBLMT are significantly enriched in genes that are known to be downregulated in hBLBC [91], as well as in functions including ion

transport and ESR1 targets (Figure 2.3A; Table 2.1). We conducted the same analysis with the validation set, and observed similar findings (Figure S4A).

We then performed single sample gene set enrichment analysis (ssGSEA) with the signature gene sets activated in hBLBC and gene sets known to be up- or downregulated in each human breast cancer subtype [13, 91, 92]. This analysis indicates cell cycle, cell proliferation, MYC targets, β-catenin-TCF targets, and epithelial mesenchymal transition (EMT) genes are all significantly more upregulated in cBLMTs, compared to cNBLMTs (Figure 2.3B). Moreover, the gene set known to be highly expressed in hBLBC is significantly upregulated in cBLMTs, while the gene set known to be lowly expressed in hBLBC is significantly downregulated in cBLMTs [91] (Figure 2.3C). This pattern is, however, not observed in cNBLMTs, as the gene set known to be highly expressed in hLumA tumors [91] does not show significant upregulation in cNBLMTs (Figure 2.3C). These results are largely supported by findings with the validation set (Figures S4B-C). The analyses indicate that cBLMT captures key molecular features of hBLBC examined, while cNBLMT fails to do so with those of hLumA.

The cBLMT subtype contains ER-PR-, ER-PR+, and ER+PR+ tumors

hBLBCs consist of approximately 70% ER-PR- (expressing neither ER nor PR) and 30% ER+PR- tumors, with ER-PR+ and ER+PR+ tumors extremely rare, based on the ESR1 and PGR transcript abundance levels (Figures 2.4A-B). However, among

cBLMTs, ER-PR+ and ER+PR+ tumors are significantly more frequent, accounting for 33% and 29% respectively, while ER-PR- and ER+PR- tumors only make up 29% and 9% respectively (Figures 2.4C-D; Table 2.1). Notably, about 78% of ER-PR+ cBLMTs were classified as basal-like in PAM50 analysis, similar to the percentage for ER-PR- cBLMT (85%) (Table 2.1). PAM50 classification also categorized 30% of ER+PR+ cBLMTs as basal-like, a proportion significantly higher than in cNBLMTs (7%) (96% of cNBLMTs are ER+PR+) (Table 2.1). The observation of higher PGR expression in cBLMTs and canine tumors than in human breast tumors (Figures 2.4A-D) is supported by other canine and human studies (Figures 3.5A-D) [67, 70, 71, 95]. We noted no significant differences in the dog's spayed status or the PIK3CA mutation status among the ER+/- PR+/- cBLMT subgroups (Figures 2.4E-F), indicating that the higher PGR expression is unlikely to be associated with either factor.

All cBLMT subgroups capture key molecular features of hBLBC

We compared each of ER-PR-, ER-PR+, and ER+PR+ cBLMT subgroups to cNBLMT (the ER+PR- subgroup contains only 6 samples and was thus excluded from the analysis; see Table 2.1). We found that the DE genes are enriched in functions that characterize hBLBC [13, 60, 96] (Figures 2.5A-C; Table 2.1). Briefly, cell cycle and Wnt signaling are enriched among upregulated genes in each cBLMT subgroup, and the enrichment is especially significant in ER-PR- and ER-PR+ cBLMTs (Figures 2.5A-C). Signatures for impaired BRCA2 function and p53 signaling, both known features of

hBLBC [13], are also enriched among upregulated genes in ER-PR- and ER-PR+ cBLMTs (Figures 2.5A-C; Table 2.1).

ER-PR+ cBLMTs harbor upregulated INF-γ response genes but downregulated IFNG

We performed DE analysis between cBLMT subgroups and found additional hBLBC characteristics [13, 57, 60, 97, 98] specific to each subgroup (Figures 2.5D-F; Table 2.1). Compared to ER+PR+ and ER-PR+ cBLMTs, upregulated genes in ER-PR- cBLMTs are significantly enriched in functions such as EMT (Figures 2.5D-E). Meanwhile, upregulated genes in ER+PR+ and ER-PR+ cBLMTs are enriched in functions including interferon-gamma (INF-γ) response (Figures 2.5D-E). Other notable findings include that Wnt-signaling-initiated pluripotency genes and IL6/JAK/STAT3 signaling genes are significantly upregulated in ER-PR+ cBLMTs, compared to ER-PR- cBLMTs (Figures 2.5D-E). Hedgehog signaling genes are significantly upregulated in both ER-PR- and ER+PR+ cBLMTs, compared to ER-PR+ cBLMTs (Figures 2.5D-E).

Interestingly, while the INF-γ response genes are upregulated in both ER+PR+ and ER-PR+ cBLMTs (Figures 2.5D-E), the INF-γ gene (IFNG) itself is significantly downregulated in ER-PR+ cBLMTs than in ER+PR+ cBLMTs (Figure 2.6A). This indicates that T cell exhaustion may occur in ER-PR+ cBLMTs [97]. To investigate this possibility, we examined the expression of 8 known T cell exhaustion markers [99], but did not find a significant difference among the three subgroups (Figure 2.6A).

As 6 out of 8 T cell exhaustion markers, including PDCD1 (encoding PD-1), express higher in ER-PR+ and/or ER+PR+ cBLMTs (Figure 2.6A), we examined the association of each marker with PGR or ESR1 in expression. We found that four markers, including PDCD1, HAVCR2, CTLA4, and TIGIT, have a significant positive association with PGR in ER-PR+ cBLMTs (Figure 2.6B and Figure 3.6). No such associations were found for PGR in other cBLMT subgroups or the cNBLMT subtype, or for ESR1 in any cBLMT subgroup or cNBLMT (Figure 2.6C and Figure 3.6).

PGR expression is associated with gene silencing

To further understand PR in canine tumors, we investigated genes that correlate with PGR or ESR1 in transcript abundance in both human and canine tumors. The same as in hLumA tumors, over 1000 genes were found to be positively correlated with ESR1 in cNBLMTs (Figure 2.7A). Importantly, both sets of genes are enriched in the same functions, including cell cycle (Figure 2.7A). For PGR, about 2.7 times as many (776 versus 289) positively correlated genes were identified for cNBLMT than hLumA (Figure 2.7A). Moreover, while the PGR-correlated 289 genes in hLumA tumors are enriched in largely the same functions as those of the ESR1-correlated genes, the 776 PGR-correlated genes in cNBLMTs are enriched in interferon-γ response (Figure 2.7A).

Fewer positively correlated genes with ESR1 or PGR were identified in basal-like tumors in both species. For ESR1, we found about 400 genes in cBLMTs, which are

enriched in functions including fatty acid metabolisms, oxidative phosphorylation, and SUZ12 targets, and 82 genes in hBLBCs, which are not enriched in any specific functions (Figure 2.7A). For PGR, while no genes were found in hBLBC, about 300 genes in cBLMT were identified and are enriched in interferon-γ response and GATA1 targets (Figure 2.7A).

Negatively correlated genes show a larger dog-human difference, especially for the basal-like subtype and PGR. For ESR1 in non-basal-like subtypes, 394 genes in hLumA tumors, enriched in functions including EMT, KRAS signaling, and SUZ12 targets, and 38 genes in cNBLMTs, enriched in functions such as oxidative phosphorylation, were identified (Figure 2.7B). However, in basal-like subtypes, only 2 genes were negatively correlated with ESR1 in hBLBCs, compared to 58 in cBLMTs that are enriched in cell cycle-related functions such as mitotic spindle and G2M checkpoints (Figure 2.7B). For PGR, while no genes were identified in either hLumA or hBLBC tumors, 142 genes in cNBLMT, enriched in functions such as hypoxia and glycolysis, and 71 genes in cBLMT, not enriched in any specific functions, were found (Figure 2.7B).

DISCUSSION

Taking advantage of the recently published RNA-seq data for hundreds of canine mammary tumor cases, we performed, to our knowledge, the first genome-wide and independent (not using any known biomarkers) subtyping of this cancer common in

bitches that are intact or spayed late. The study identifies two subtypes, and further shows that one subtype molecularly resembles hBLBC, while the other subtype appears not to match any human breast cancer subtypes. This conclusion is consistent with our previous study [33], but differs from a recent publication reporting that canine and human luminal A tumors have more molecular homology than canine and human basal-like tumors [53]. While our canine-only PAM50 analysis also classifies most tumors of one canine subtype as luminal A, the same as the study by Bergholtz et al. [53], our cross-species PAM50 analysis clearly separates canine luminal A tumors from hLumA tumors, unlike basal-like tumors (as such, we named the two canine subtypes as basal-like, cBLMT, and non-basal-like, cNBLMT) (note that Bergholtz et al. [53] did not perform cross-species PAM50 classification). We further show that cBLMTs capture key molecular features and expression patterns of hBLBCs, whereas cNBLMTs fail to do the same with hLumA tumors. Our results are supported by the histology of the canine subtypes. cBLMTs are enriched in simple carcinomas and simple adenomas, where only one cell lineage prominently proliferates. cNBLMTs, however, are enriched in complex or mixed carcinomas and adenomas, where multiple cell lineages (e.g., epithelial cells and myoepithelial cells) proliferate [33, 41, 45]. Complex or mixed tumors (e.g., adenomyoepithelioma) are very rare in human breast cancer [33, 46, 47]; thus, cNBLMTs do not match hLumA tumors histologically.

Although cBLMT co-clusters with hBLBC in PAM50 classification and captures the key molecular features and gene expression patterns of hBLBC, cBLMT contains ER-PR+

and ER+PR+ tumors, which are rare in hBLBC. Moreover, while ER+PR- tumors are common in hBLBC and other breast cancer subtypes, ER-PR+ tumors are nearly nonexistent in any human breast cancer subtype. This is because that in human breast cells, PR is induced by ER [100, 101] and thus, without ER, PR will not be expressed. One notable difference between the estrous cycle in bitches and the menstrual cycle in women is the luteal phase, which lasts 14 days for humans but 2 months for dogs. As the canine mammary glands are constantly exposed to a high level of progesterone during the luteal phase [102, 103], it is possible that the PGR gene is still actively transcribed after the ESR1 gene is silenced in dogs, resulting in the ER-PR+ tumors. More studies are needed to investigate this possibility.

Despite the difference in the PR expression status, ER+PR+ and ER-PR+ cBLMTs capture key molecular features of hBLBCs, the same as ER-PR- cBLMTs. These include upregulation of cell cycle genes and Wnt signaling. Upregulation of cell cycle genes could lead to high cell proliferation, a molecular characteristic of hBLBC [13]. Activated Wnt signaling is also a well-known feature of hBLBC [60, 96]. For example, WNT5B, one of the major Wnt signaling molecules, is known to drive the hBLBC phenotype, both in vitro and in vivo, by activating both canonical and non-canonical Wnt signaling [19]. WNT5B is upregulated in both ER-PR+ and ER+PR+ cBLMTs. Wnt signaling-driving pluripotency genes are upregulated in ER-PR+ cBLMTs, which likely further drives the basal-like features of these tumors [96].

One notable finding from our study is that in ER-PR+ cBLMTs, interferon-γ response genes are upregulated, but the interferon-γ gene (IFNG) itself is downregulated. Moreover, ER-PR+ cBLMTs appear to express more of the immune checkpoint genes PDCD1 (encoding PD-1) and CTLA4, and only in these tumors, PGR is positively correlated with PDCD1 and CTLA4. These results are consistent with T cell exhaustion, where T cells are hypofunctional [104]. T cell exhaustion occurs in many human cancers, including hBLBCs [97], and presents challenges and opportunities in cancer immunotherapy [104]. Due to the small sample size, we cannot conclude definitively that T cell exhaustion indeed occurs in cBLMTs. Once this possibility is validated with further studies, cBLMTs could be a valuable model to investigate the relationship among progesterone, PR, and T cell exhaustion. Importantly, cBLMTs may be good models to test novel immunotherapies targeting T cell exhaustion [104].

Our study reveals that unlike ESR1, PGR is associated with gene silencing in canine mammary tumors. However, the silenced genes appear to be random and not enriched in any particular functions, especially in ER-PR+ cBLMTs. Interestingly, we find that many of the ESR1 or PGR-correlated genes are targets of SUZ12, a component of polycomb repressive complex 2 (PRC2) that primarily methylates lysine 27 of histone H3 (e.g., H3K27me3, a marker of transcriptionally silent chromatin). Further studies are needed to determine if PRC2 is responsible for PGR-associated gene silencing.

CONCLUSIONS

We identify two molecular subtypes in spontaneous canine mammary tumors. One subtype, cBLMT, molecularly and histologically resembles hBLBC, a breast cancer subtype that lacks an effective treatment and has the worst clinical outcomes. The other subtype, cNBLMT, appears not to match any human breast cancer subtype molecularly and histologically. While cBLMTs also consist of ER-PR+ and ER+PR+ tumors (which may be related to the long luteal phase of the estrous cycle in dogs), a difference from hBLBC, we note that these tumors capture the key molecular features of hBLBCs, the same as ER-PR- cBLMTs. Thus, cBLMTs could serve as a much-needed spontaneous animal model for hBLBC, filling a critical gap in breast cancer research. Moreover, while much more studies are needed, ER-PR+ cBLMTs may provide a valuable system to study T cell exhaustion, as well as estrogen/ER-independent roles of progesterone and PR in gene silencing.

List of abbreviations

NMF: non-negative matrix factorization

GEO: Gene Expression Omnibus

NCI: National Cancer Institute

GDC: Genomic Data Commons

TCGA: The Cancer Genome Atlas

cBLMT: canine basal-like mammary tumors

cNBLMT: canine nonbasal-like mammary tumors

ER: Estrogen receptor

PR: Progesterone receptor

TPM: Transcripts per million

FPKM: Fragments per kilobase of exon per million mapped

QC: Quality control

CDS: coding sequence

TMB: tumor mutation burden

hBLBC: human basal-like breast cancer

hLumA: human luminal A breast cancer

EMT: epithelial mesenchymal transition

DE: Differentially expressed

PRC2: Polycomb repressive complex 2

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

The RNA-seq data generated from this study are submitted to the SRA database at https://www.ncbi.nlm.nih.gov/bioproject/PRJNA912710. All other RNA-seq and microarray data sets analyzed in this study are obtained from the SRA and GEO databases at: https://www.ncbi.nlm.nih.gov/bioproject/PRJNA489087/, https://www.ncbi.nlm.nih.gov/bioproject/PRJNA203086/, https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA561580, https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE20718, https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE22516, and https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=gse20685.

Competing interests

The authors declare that they have no competing interests.

Funding

This research was funded by the National Cancer Institute (NCI), National Institute of Health (NIH) grants R01 CA252713 and R01 CA182093. The Ohio State University College of Veterinary Medicine Biospecimen Repository, one of the sample sources, is supported by NCATS UL1TR001070 and NCI P30CA016058.

Authors' contributions

SZ, TF, KD, and JW conceived the experiment design. JW and SZ wrote the manuscript. KH performed all QC analysis. YF performed DE analysis and plotted Figure 3A. KD provided advice on statistical analyses. TF performed initial NMF and PAM50 classification. JW conducted all other analyses.

Acknowledgements

We thank Ms. Jin Qian for her contribution to the study; Dr. Holly Borghese and Dr. William C. Kisseberth for their help in sample collection; the Georgia Advance Computing Resource Center (GACRC) for providing the computing power for this work; Dr. Michael Skaro for advice on figure design; and the authors who published canine and human data used in this work.

Table 2.1. Features of canine mammary tumor subtypes and subgroups

Subtype	cBLMT[1]				cNBLMT[2]
Subgroup	ER-PR-	ER-PR+	ER+PR+	ER+PR-	96% ER+PR+
Basal-like	17	18	6	3	5
Luminal A	0	0	4	2	55
Luminal B	2	1	6	1	8
HER2-enriched	1	3	2	0	2
Normal-like	0	1	2	0	4
Enriched functions of upregulated genes in subgroup	Hedgehog signaling; EMT; Wnt/β-catenin signaling	Interferon-γ response; IL6/JAK/STAT 3 signaling; Wnt & pluripotency	Interferon-γ response; Hedgehog signaling	ND[3]	
Histological & clinical features	Simple adenoma/carcinoma enriched; tumor with lymph node invasion enriched				Complex or mixed tumors enriched.
Enriched functions of upregulated genes in subtype	Cell cycle; proliferation; MYC target; Wnt signaling; EMT; basal-like upregulation				Ion transport; ESR1 targets; basal-like downregulation
Mutation enrichment	TP53 mutation				PIK3CA H1047R/L; KAT6B mutation; PI3K-AKT-mTOR pathway alteration
Cross-species PAM50	Clustered with hBLBC				Not clustered with any human breast cancer subtype

[1]cBLMT: canine basal-like mammary tumor.

[2]cNBLMT: canine non-basal-like mammary tumor.

[3]cBLMT ER+PR- subgroup has a small sample size and is not investigated.

FIGURES

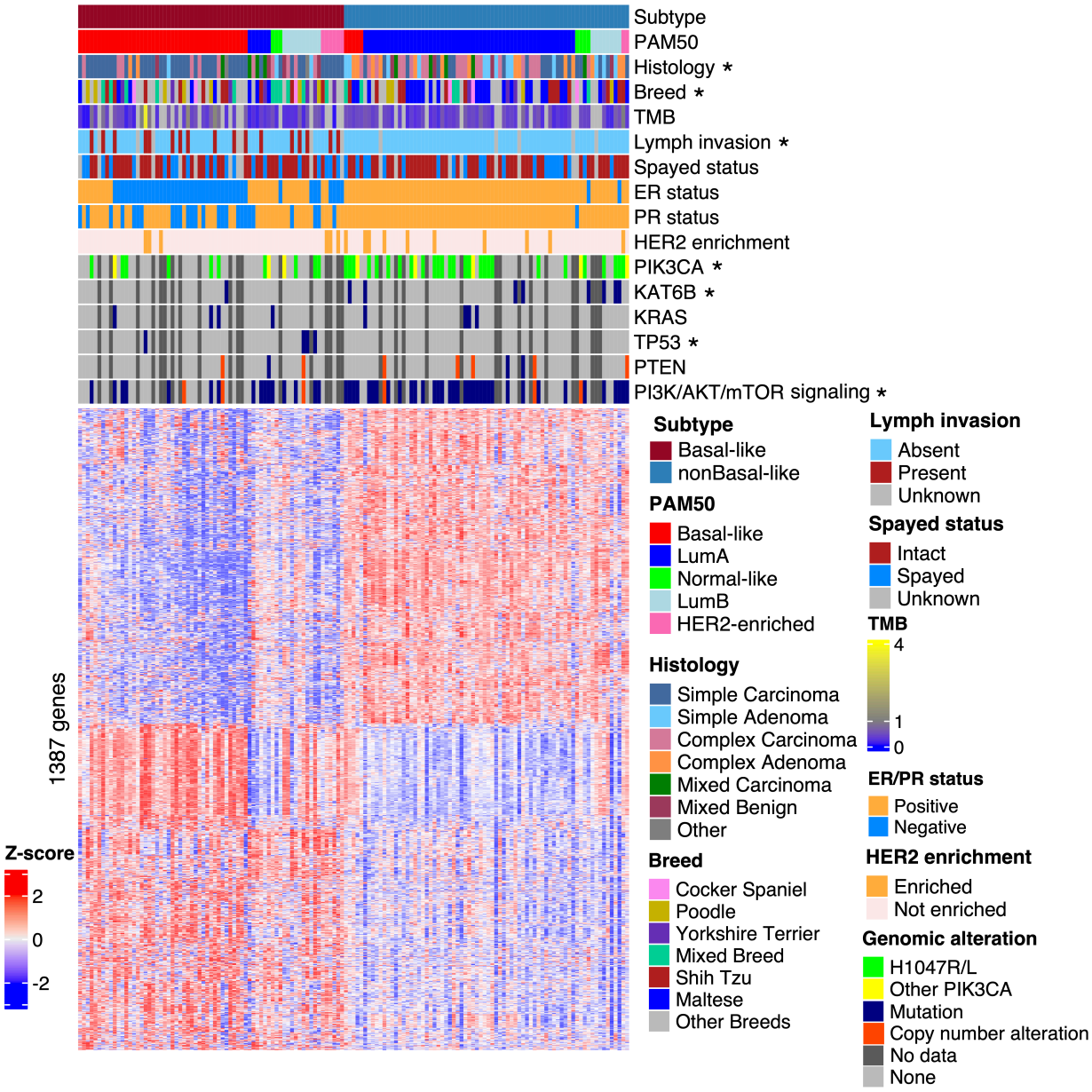


Figure 2.1: Two molecularly distinct subtypes of canine mammary tumors were identified.

Heatmap of the row scaled $\log_2$(TPM) values of 1,387 metagenes from 143 canine mammary tumors (columns), ordered from left to right by subtypes, PAM50 classification, and the ESR1 expression level from high to low. The metagenes (rows),

identified from the NMF analysis (see Methods), are ordered by hierarchical clustering. Lymph invasion is defined as having tumor cells in peritumoral lymphatic vessels and/or regional lymph nodes. Tumor mutation burden (TMB), defined as the number of somatic base substitutions and small indels per megabase (Mb) of callable coding sequence, is obtained from a previous publication [42]. For ER/PR status, a tumor is considered “negative” or “positive” if its ESR1/PGR has a FPKM value of ≤ 1 or > 1, respectively. For HER2 enrichment, a sample is considered “not enriched” or “enriched” if its ERBB2 has a FPKM value of ≤ 35 or >35, respectively. “Other PIK3CA” represents all non-H1047 coding mutations in PIK3CA. “No data” represents samples with no mutation data. Annotation row titles marked with a “*” indicate a significant ($p \leq 0.05$) difference in enrichment between the subtypes.

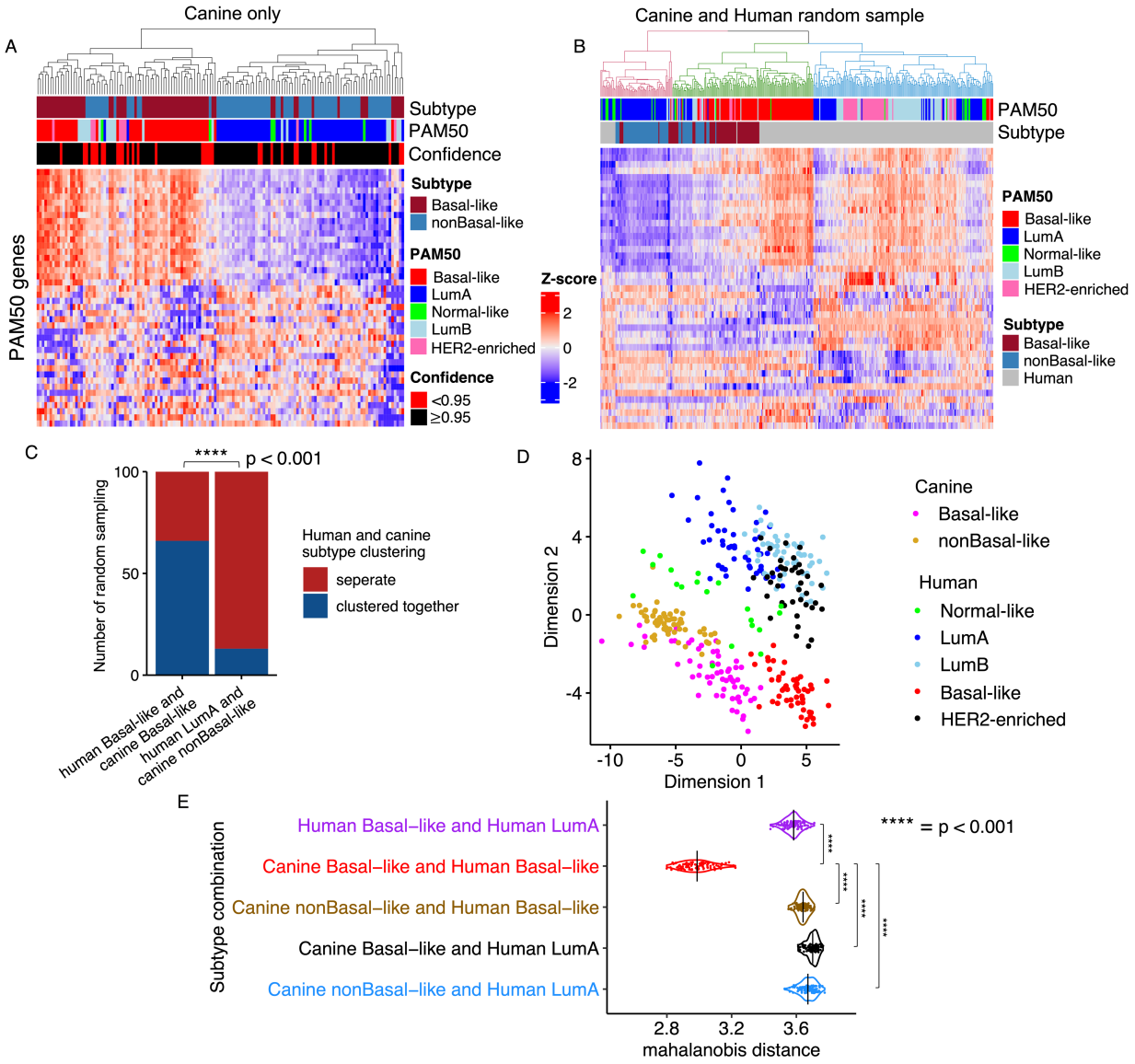


Figure 2.2: PAM50 classification groups canine and human basal-like tumors, but separates canine non-basal-like and human tumors.

A. PAM50 classification of 143 subtyped canine mammary tumors. The heatmap shows hierarchical clustering of the canine tumors using the row scaled $\log_2$(TPM) values of 43 out of the 50 PAM50 genes. The bars indicate the canine subtype from Figure 2.1, as well as the PAM50 subtype and confidence score of each tumor.

B. An example of cross-species PAM50 classification. All 143 subtyped canine tumors, together with 267 human tumors (60 tumors per subtype of luminal A (LumA), luminal B (LumB), basal-like or HER2-enriched randomly sampled from TCGA database, along with all 27 normal-like tumors from TCGA) were subjected to PAM50 classification. The dendrogram is colored to indicate the minimum clusters that maximally separate the hLumA tumors from hBLBC tumors. This cross-species PAM50 classification were repeated 100 times.

C. Bar plot showing the number of random samplings in which human and canine basal-like tumors clustered together or separately, compared to those of human luminal A and canine non-basal-like tumors. The p-value is based on Fisher's exact test.

D. Multidimensional scaling plot of the cross-species PAM50 classification shown in B. Each dot represents a tumor from a subtype specified by the color as indicated in the legend.

E. Violin plot indicating the distribution of the mahalanobis distances between the centers of two subtypes on the multidimensional scaled plot (D) of each of the 100 cross-species PAM50 classifications (B) achieved via random samplings (see Methods). The p-values were obtained from Wilcoxon tests.

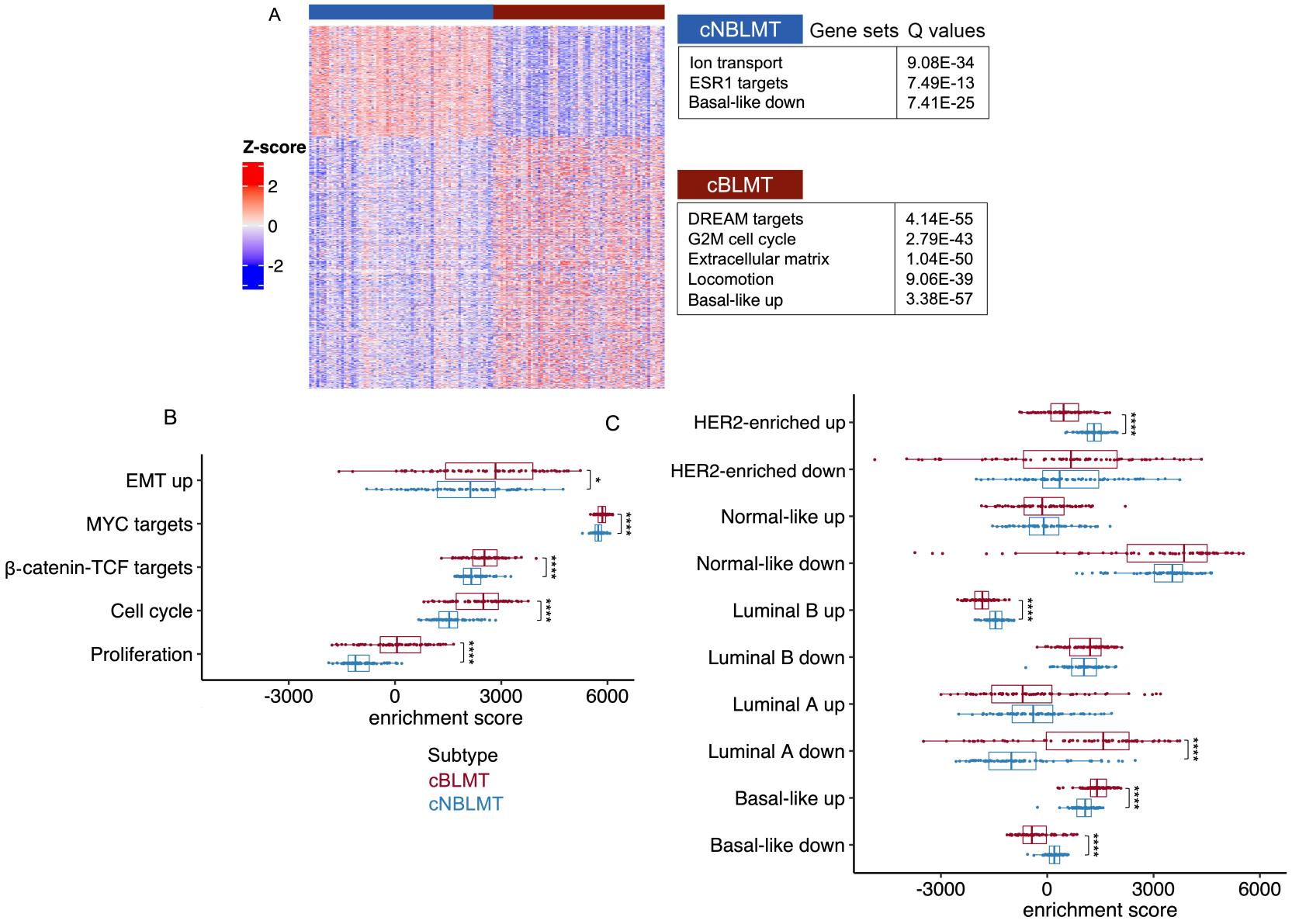

A
Z-score
2
0
-2
cNBLMT
Gene sets Q values
Ion transport 9.08E-34
ESR1 targets 7.49E-13
Basal-like down 7.41E-25
cBLMT
DREAM targets 4.14E-55
G2M cell cycle 2.79E-43
Extracellular matrix 1.04E-50
Locomotion 9.06E-39
Basal-like up 3.38E-57
B
EMT up
MYC targets
β-catenin-TCF targets
Cell cycle
Proliferation
-3000
0
3000
6000
enrichment score
Subtype
cBLMT
cNBLMT
C
HER2-enriched up
HER2-enriched down
Normal-like up
Normal-like down
Luminal B up
Luminal B down
Luminal A up
Luminal A down
Basal-like up
Basal-like down
-3000
0
3000
6000
enrichment score

Figure 2.3: Differentially expressed (DE) gene analysis indicates the enrichment of hBLBC signatures in cBLMT.

A. Heatmap of the row scaled log_2(TPM) values of 1,620 DE genes between cBLMT and cNBLMT samples, identified with an expression fold change of >2 and a BH adjusted p-value of < 0.01 for each DE gene (see Methods). The enriched functions among each DE gene group are indicated.

B & C. Distribution of single sample gene set enrichment analysis (ssGSEA) scores of canine tumors with hBLBC signature gene sets (B), as well as with gene sets with expression patterns characterizing each human breast cancer subtype [91] shown (C). P-values are from Wilcoxon tests. *: $p < 0.05$; ****: $p < 0.0001$.

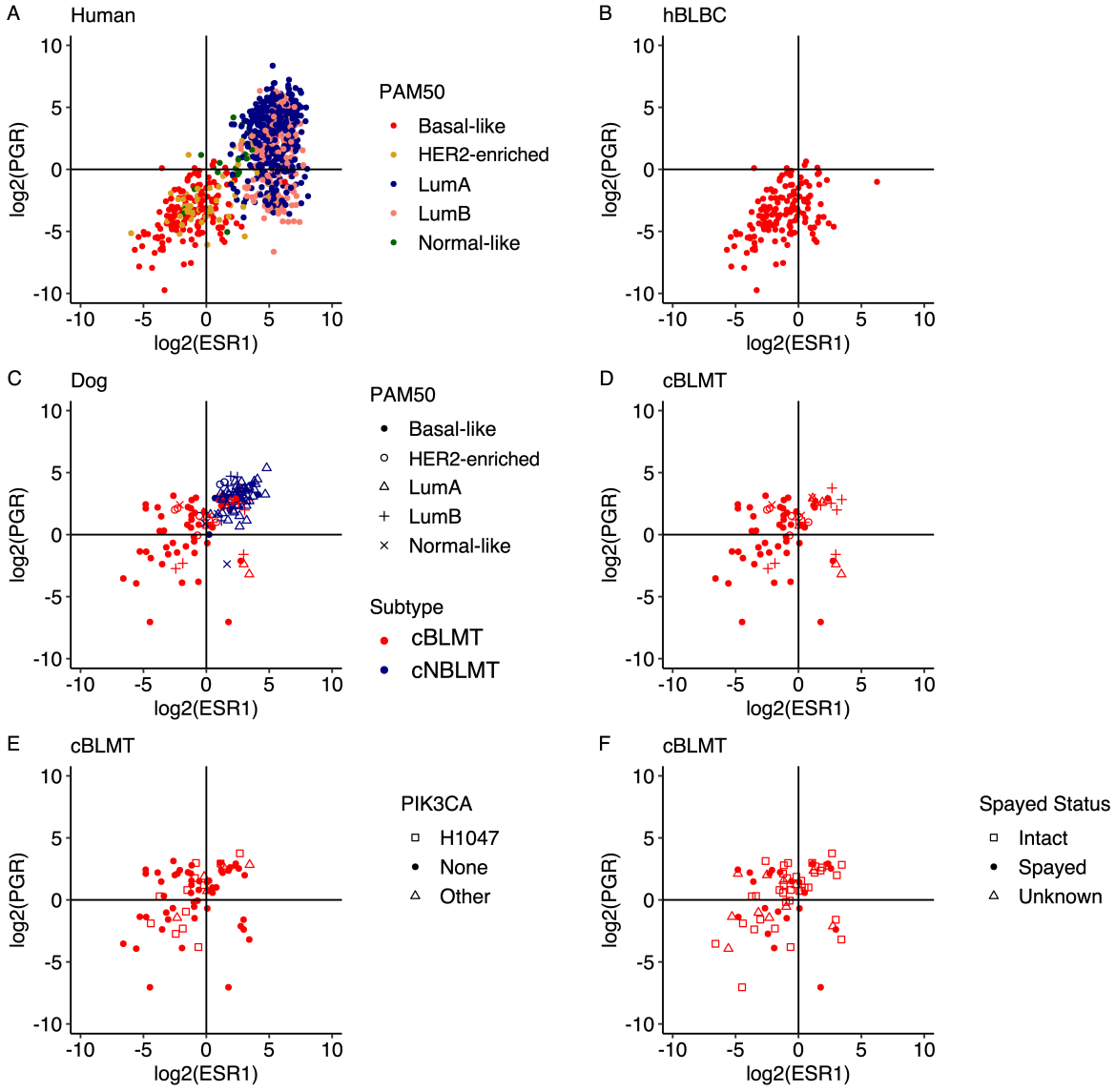


Figure 2.4: cBLMT contains more ER-PR+ tumors than hBLBC.

Scatter plots show the $\log_2$(FPKM) values of ESR1 and PGR for all TCGA human breast cancers with a PAM50 subtype (n=827) (A), hBLBCs (n=140) (B), both cBLMTs and cNBLMTs (n=143) (C), as well as cBLMTs (n=69) with the PAM50 subtype (D), PIK3CA mutation status (E), dog's spayed status (F) indicated.

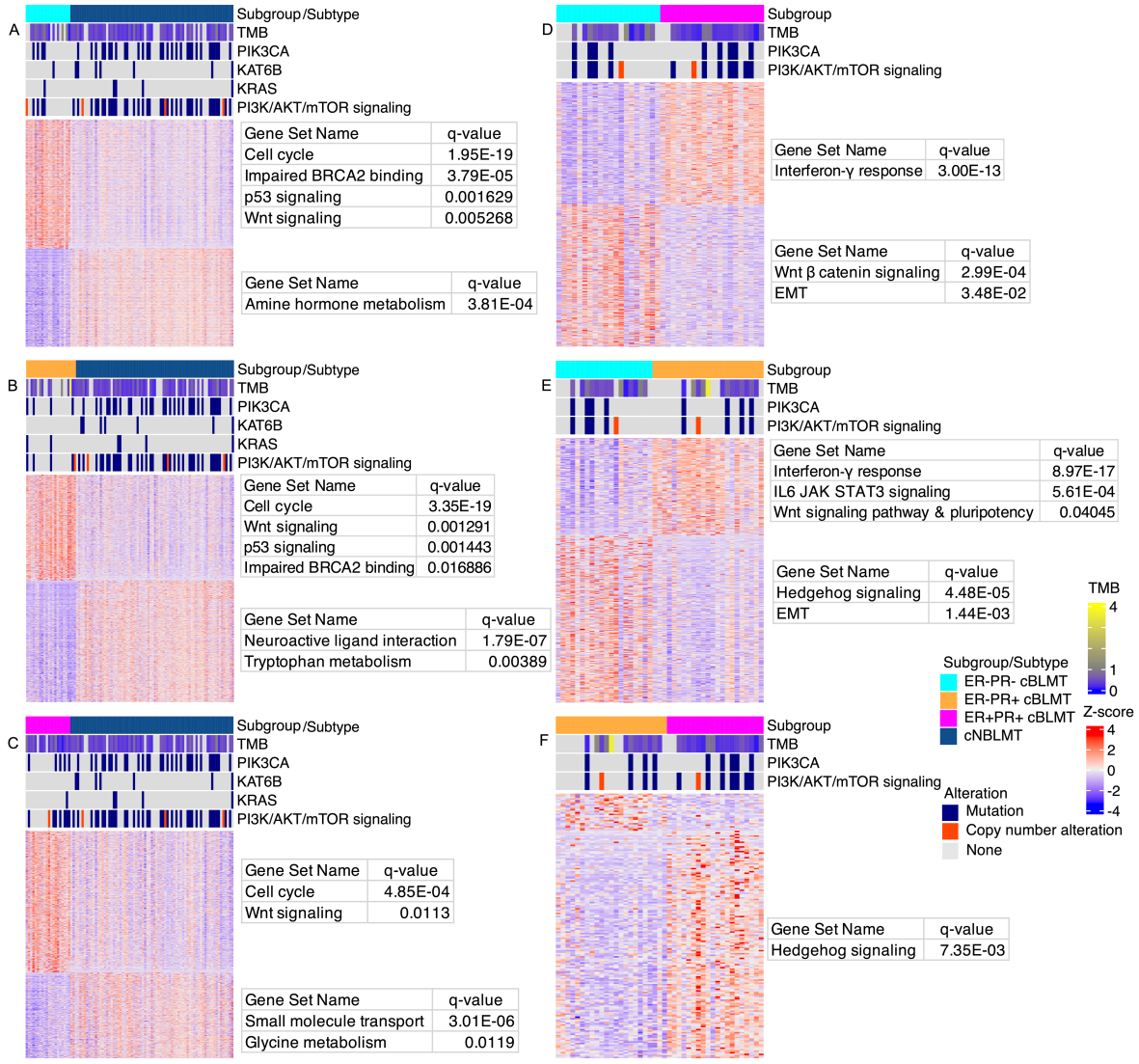


Figure 2.5: Basal-like features are maintained in all cBLMT subgroups, especially in ER-PR+ and ER-PR- cBLMTs.

A-C. Heatmaps showing the DE genes identified between cNBLMT and each of the ER-PR- (A), ER-PR+ (B), and ER+PR+ (C) cBLMT subgroups, with a BH adjusted p-value of < 0.05 and an expression-fold change of > 2. The heatmaps are presented as described in Figure 2.3A, along with TMB and the most mutated genes indicated as

described in Figure 2.1. A tumor was classified ER+ or PR+ if its ESR1 or PGR gene has a FPKM value of > 1, respectively; otherwise, the tumor was classified ER- or PR-. D-F. Heatmaps of DE genes identified between ER+/- PR+/- cBLMT subgroups, presented as described in A-C.

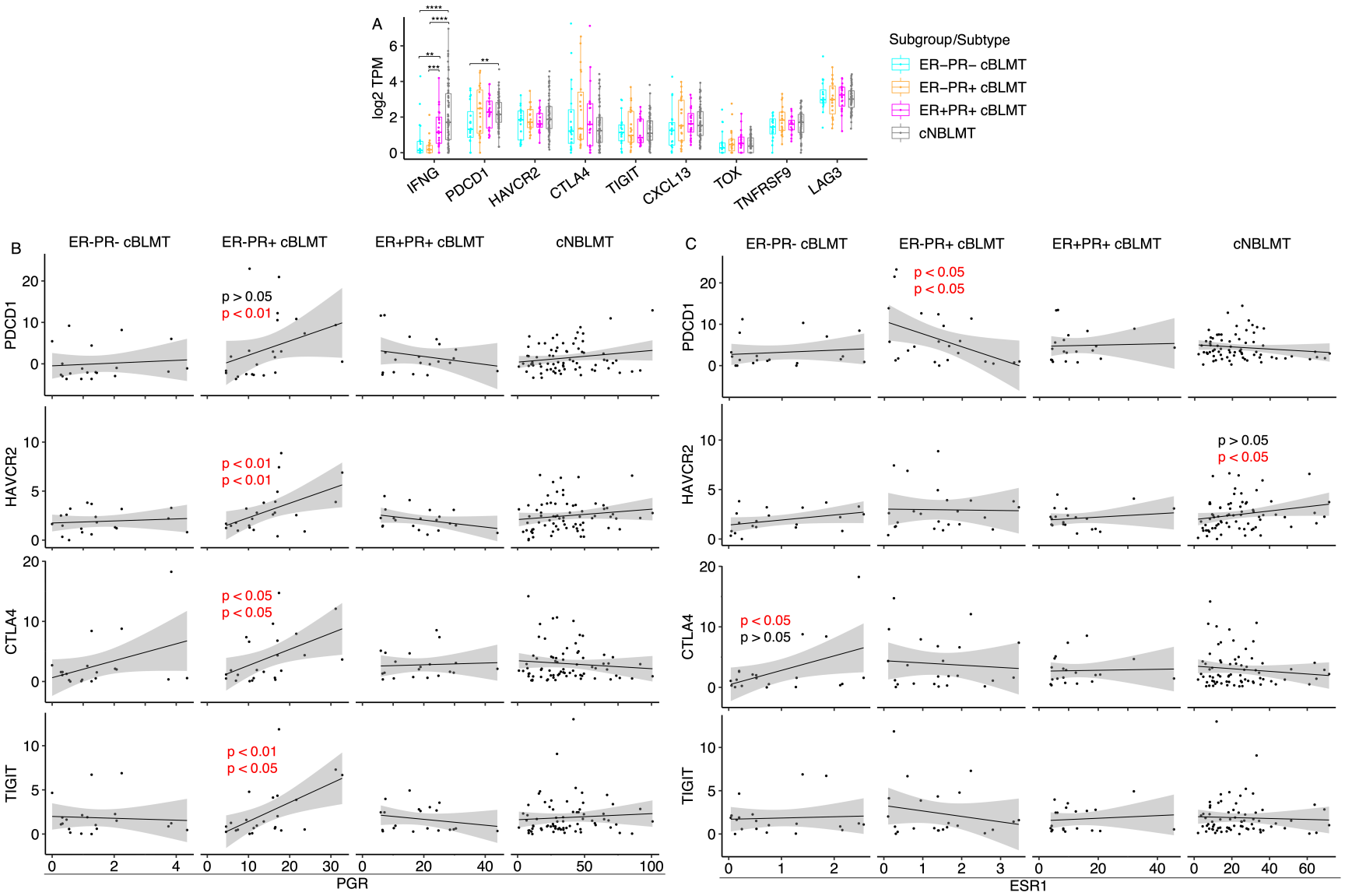

A
log2 TPM
IFNG
PDCD1
HAVCR2
CTLA4
TIGIT
CXCL13
TOX
TNFRSF9
LAG3
Subgroup/Subtype
ER−PR− cBLMT
ER−PR+ cBLMT
ER+PR+ cBLMT
cNBLMT
B
C
ER-PR- cBLMT
ER-PR+ cBLMT
ER+PR+ cBLMT
cNBLMT
PDCD1
HAVCR2
CTLA4
TIGIT
p > 0.05
p < 0.01
p < 0.01
p < 0.01
p < 0.05
p < 0.05
p < 0.01
p < 0.05
p < 0.05
p < 0.05
p > 0.05
p > 0.05
p < 0.05
PGR
ESR1

Figure 2.6: PGR correlates with several T cell exhaustion signature genes in mRNA expression in ER-PR+ cBLMTs.

A. Dot plots of log_2(TPM) values of IFNG and 8 canonical T cell exhaustion marker genes in each cBLMT subgroup and cNBLMT. P-values are from Wilcoxon tests. **: $p<0.01$; ***: $p<0.001$; ****: $p<0.0001$.

B & C. Pearson (top) and Spearman (bottom) correlation analysis between PGR (B) or ESR1 (C) and PDCD1, HAVCR2, CTLA4, or TIGIT in mRNA expression in each subgroup and subtype shown. P-values of only significant Pearson and/or Spearman correlations are indicated.

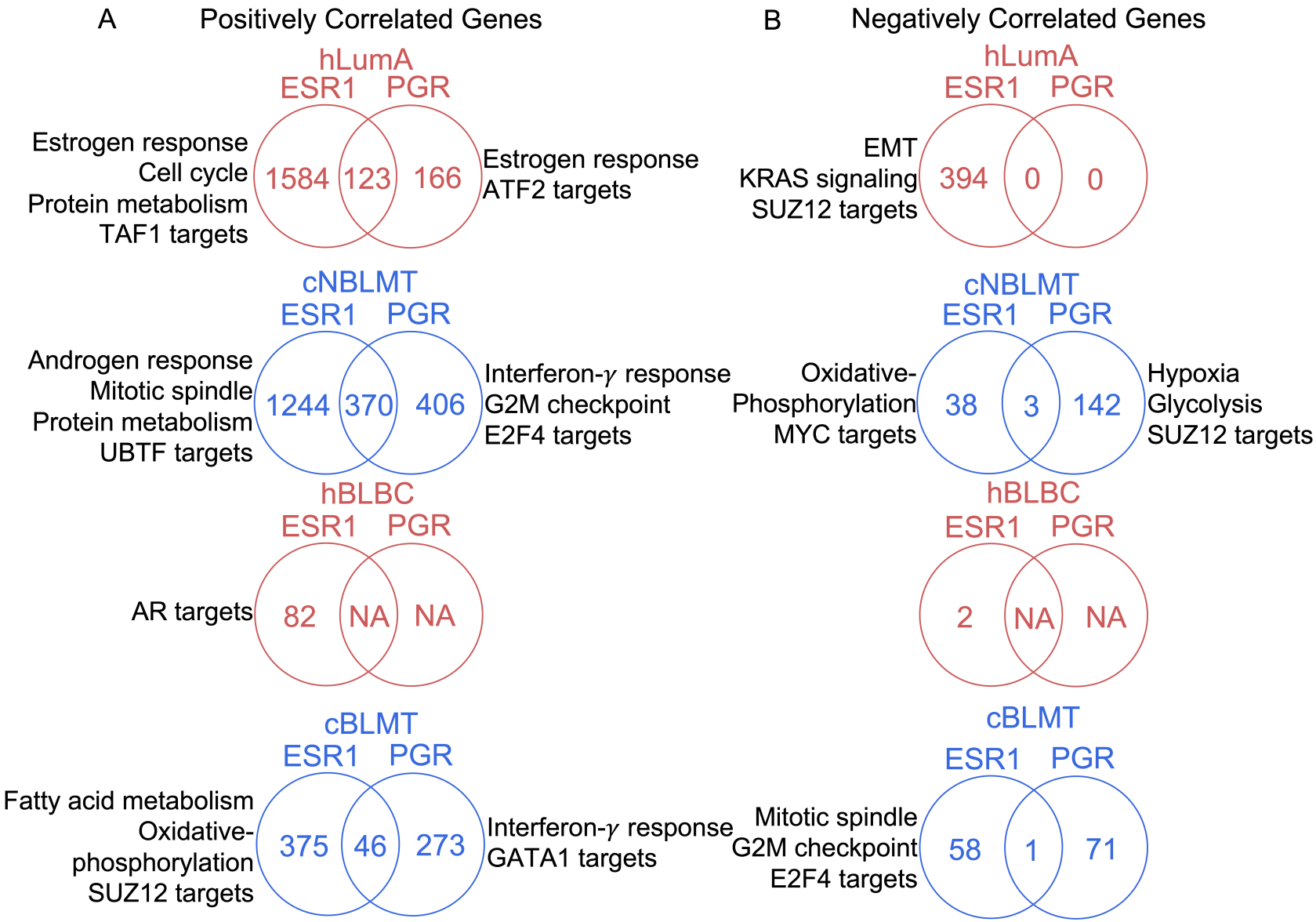

A
Positively Correlated Genes
hLumA
ESR1
PGR
Estrogen response
Cell cycle
Protein metabolism
TAF1 targets
1584
123
166
Estrogen response
ATF2 targets
cNBLMT
ESR1
PGR
Androgen response
Mitotic spindle
Protein metabolism
UBTF targets
1244
370
406
Interferon-γ response
G2M checkpoint
E2F4 targets
hBLBC
ESR1
PGR
AR targets
82
NA
NA
cBLMT
ESR1
PGR
Fatty acid metabolism
Oxidative-
phosphorylation
SUZ12 targets
375
46
273
Interferon-γ response
GATA1 targets
B
Negatively Correlated Genes
hLumA
ESR1
PGR
EMT
KRAS signaling
SUZ12 targets
394
0
0
cNBLMT
ESR1
PGR
Oxidative-
Phosphorylation
MYC targets
38
3
142
Hypoxia
Glycolysis
SUZ12 targets
hBLBC
ESR1
PGR
2
NA
NA
cBLMT
ESR1
PGR
Mitotic spindle
G2M checkpoint
E2F4 targets
58
1
71

Figure 2.7: PGR is associated with gene silencing in canine tumors but not in human tumors.

A. Venn diagrams indicating the number of genes that are positively associated with ESR1 and/or PGR in mRNA expression in each human or canine subtype specified. These genes were identified as those having correlation coefficient $\mathrm{R} \geq 0.3$ and BH-adjusted $\mathrm{p} \leq 0.05$ in both Pearson and Spearman correlation analyses. Indicated are also the top enriched functions of genes that are correlated with only ESR1 or PGR.

B. Venn diagrams for genes negatively correlated with ESR1 and/or PGR, identified with $\mathrm{R} \leq -0.3$ and other cutoffs and presented as described in A.

CHAPTER 3

SUPPLEMENTAL FIGURES

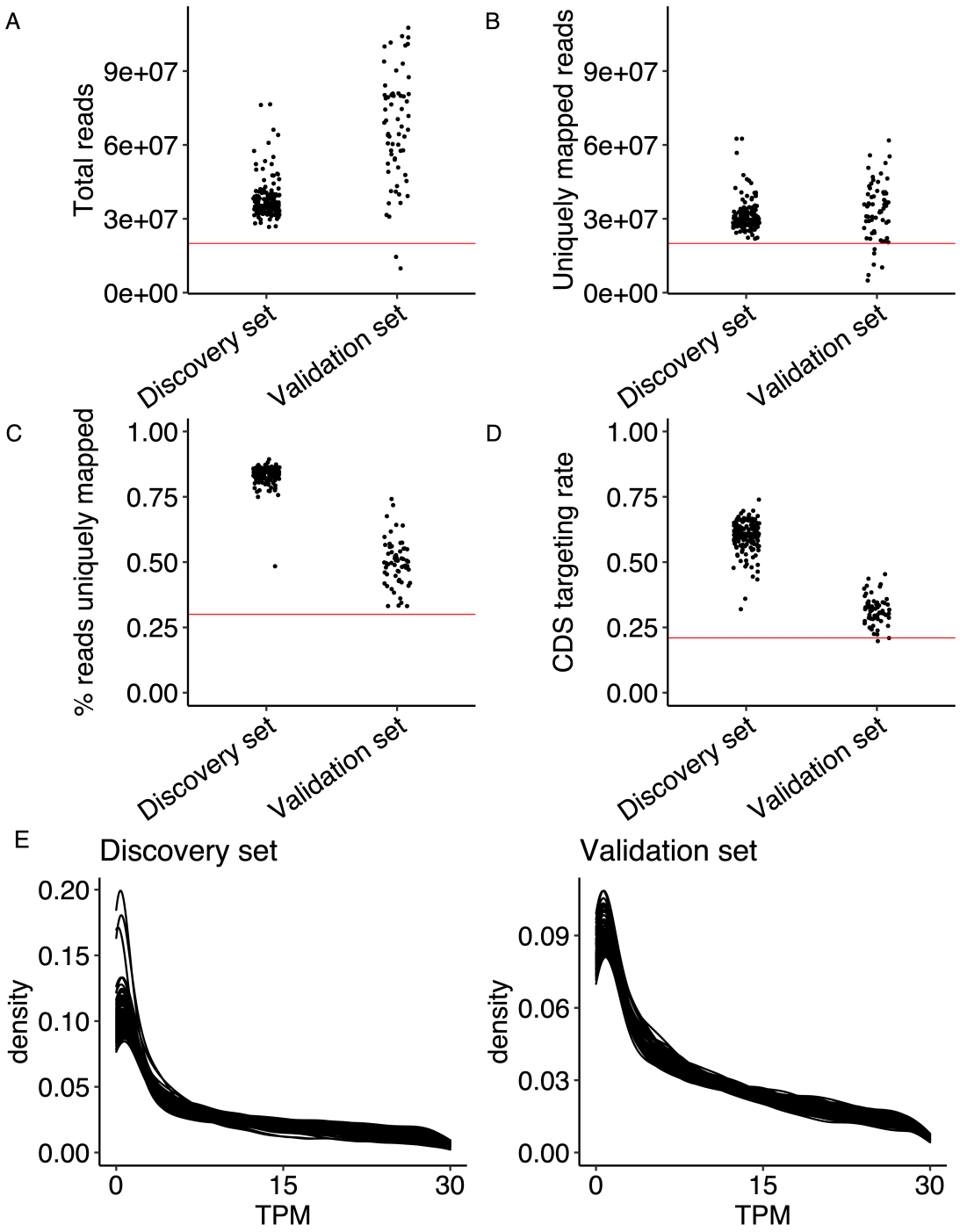


Figure 3.1: Figure S1. RNA-seq quality control; related to Figure 2.1

A. Distributions of sequencing amounts (the number of reads per sample) of the discovery and validation sets. Each dot represents a sample. Red line specifies the cutoff (20 million total reads) and samples below the red line were excluded.

B-C. Distributions of per sample total amount (B) and rate (C) of reads that are uniquely (and concordantly for the discovery set) mapped to the canFam3 genome. Red line specifies the cutoff (20 million reads for B; 0.3 for C) and samples below the red line were excluded.

D. Distributions of per sample CDS-targeting rate. Red line specifies the cutoff (0.21) and samples below the red line were excluded.

E. Gene expression distributions in each sample. Each line represents a sample, with the TPM value of each of all ~20,000 protein-coding genes plotted.

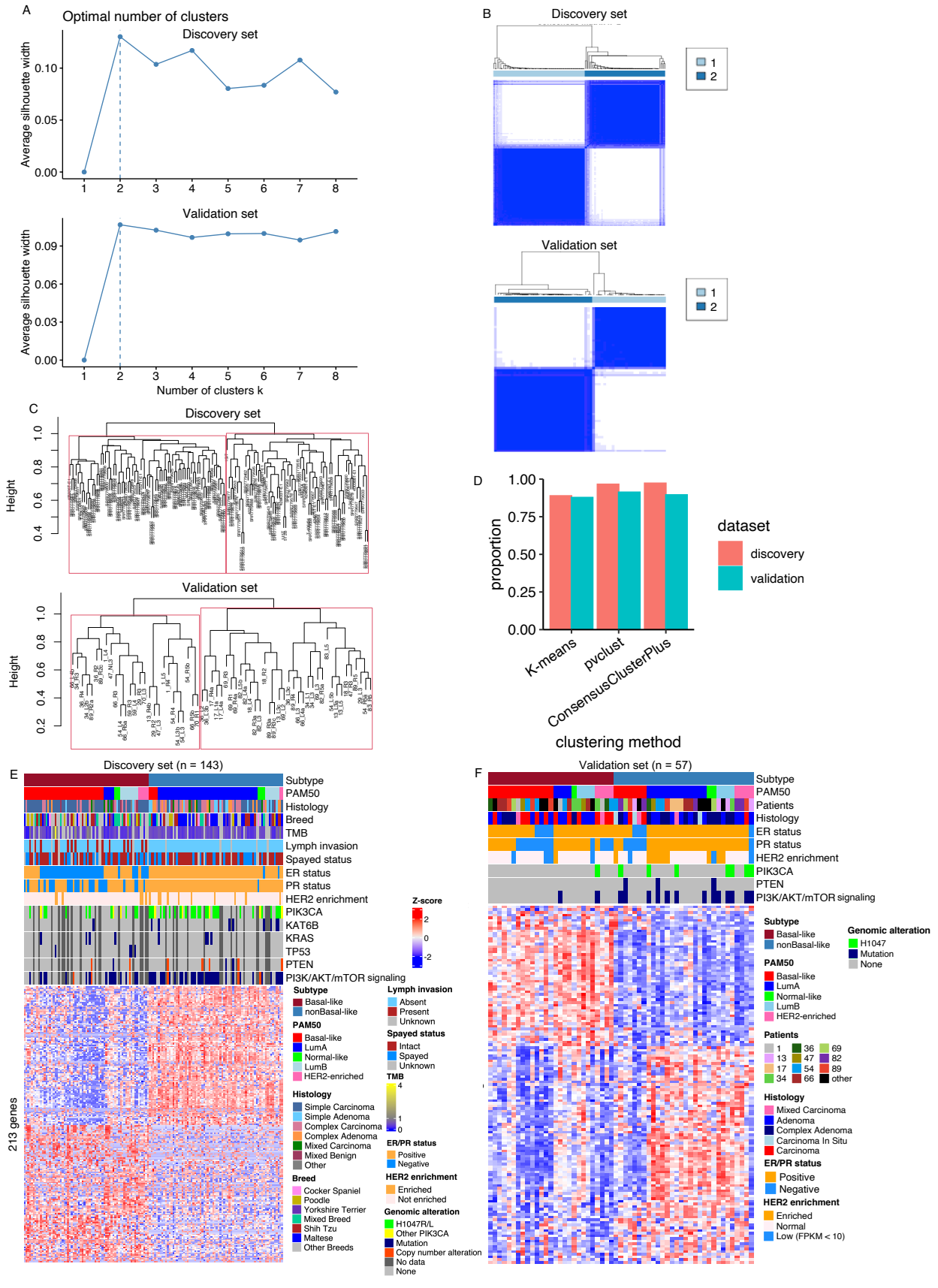

A
Optimal number of clusters
Discovery set
Average silhouette width
Validation set
Number of clusters k
B
Discovery set
Validation set
C
Discovery set
Validation set
Height
D
proportion
dataset
discovery
validation
K-means
pvclust
ConsensusClusterPlus
clustering method
E
Discovery set (n = 143)
Subtype
PAM50
Histology
Breed
TMB
Lymph invasion
Spayed status
ER status
PR status
HER2 enrichment
PIK3CA
KAT6B
KRAS
TP53
PTEN
PI3K/AKT/mTOR signaling
Z-score
213 genes
Basal-like
nonBasal-like
LumA
Normal-like
LumB
HER2-enriched
Simple Carcinoma
Simple Adenoma
Complex Carcinoma
Complex Adenoma
Mixed Carcinoma
Mixed Benign
Other
Cocker Spaniel
Poodle
Yorkshire Terrier
Mixed Breed
Shih Tzu
Maltese
Other Breeds
Absent
Present
Unknown
Intact
Spayed
Positive
Negative
Enriched
Not enriched
H1047R/L
Other PIK3CA
Mutation
Copy number alteration
No data
None
F
Validation set (n = 57)
Patients
Mixed Carcinoma
Adenoma
Complex Adenoma
Carcinoma In Situ
Carcinoma
Normal
Low (FPKM < 10)
H1047
other

Figure 3.2: Figure S2. Validation of canine mammary tumor subtyping results shown in Figure 2.1 using different strategies and data set.

A-C. K-means (A), consensus clustering (ConsensusClusterPlus) (B), and permutation-based hierarchical clustering (pvclust) (C) were applied using the top 10% most variably expressed genes in samples of the discovery set (top) or the validation set (bottom). These approaches all yield two subtypes, the same as the NMF strategy shown in Figure 2.1.

D. The proportions of samples assigned to the same subtype by each approach indicated in A-C as the NMF strategy shown in Figure 2.1.

E-F. Heatmaps of the discovery (E) and validation (F) sets as presented in Figure 2.1, with NMF clustering conducted on the top 2,000 most variably expressed genes within each cohort.

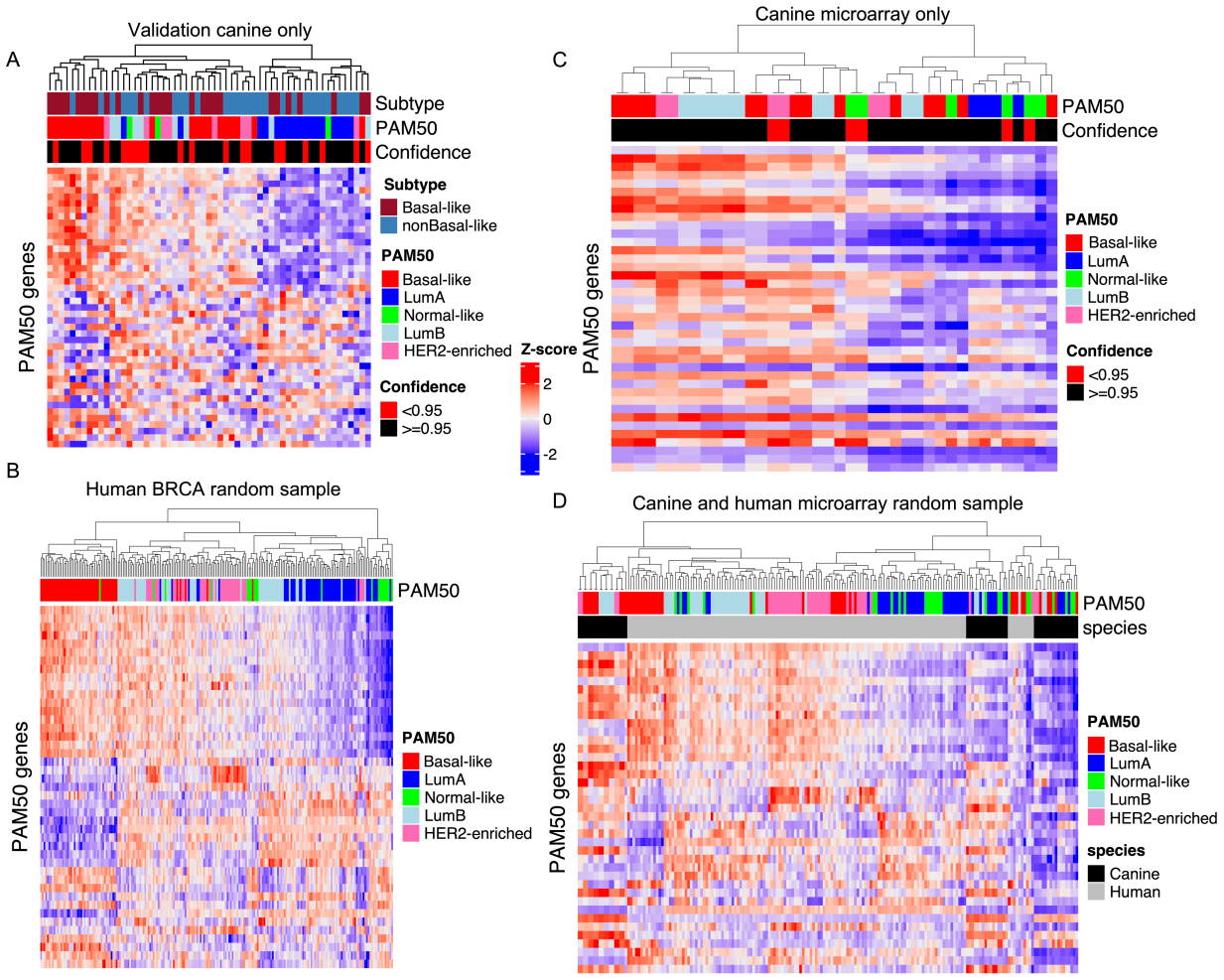


Figure 3.3: Figure S3. PAM50 classification with canine tumors only, and with canine and human combined samples; related to Figure 2.2.

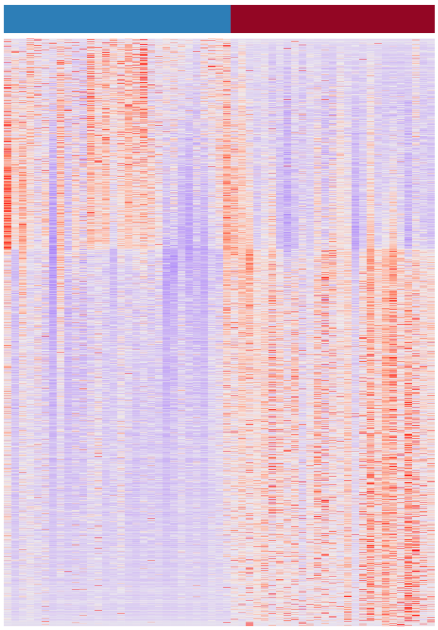
A. PAM50 classification of the 57 tumors of the validation set, presented as described in Figure 2.2A.

B. PAM50 classification of 267 human breast tumors sampled from TCGA as described in the Methods section. The 50 PAM50 genes were used and the tumor subtypes shown were from cBioportal.

C. PAM50 classification of 40 canine tumors from two gene expression microarray studies [70, 71]. Log_2-transformed expression values of 40 out of the 50 PAM50 genes were used (see Methods).

D. An example of cross-species PAM50 classification using canine and human microarray data. Human tumors (30 tumors per PAM50 subtype) were randomly sampled from a microarray data study [69] as described in the Methods section. The figure is presented as described for Figure 2.2B.

A

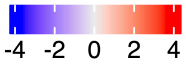


cNBLMT

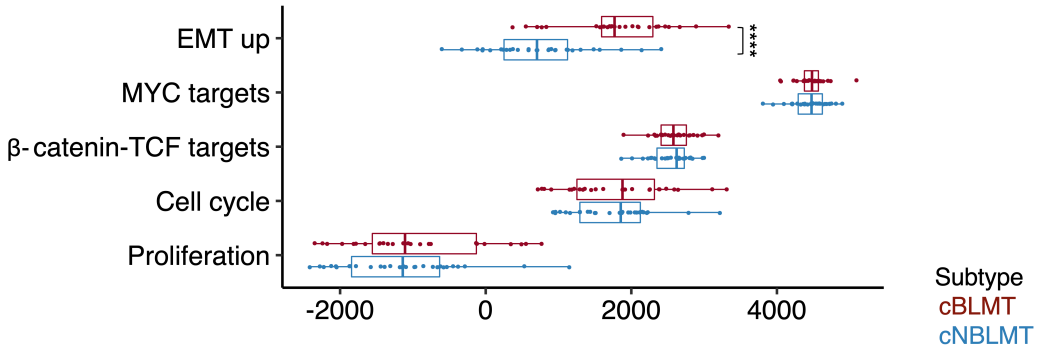
Gene sets	Q-value
Basal-like down	4.27E-21
ESR1 targets	1.59E-07

cBLMT

Breast cancer ductal invasive up	1.22E-68
Core matrisome	9.73E-66
EMT	2.64E-44
Locomotion	2.15E-3



B



C

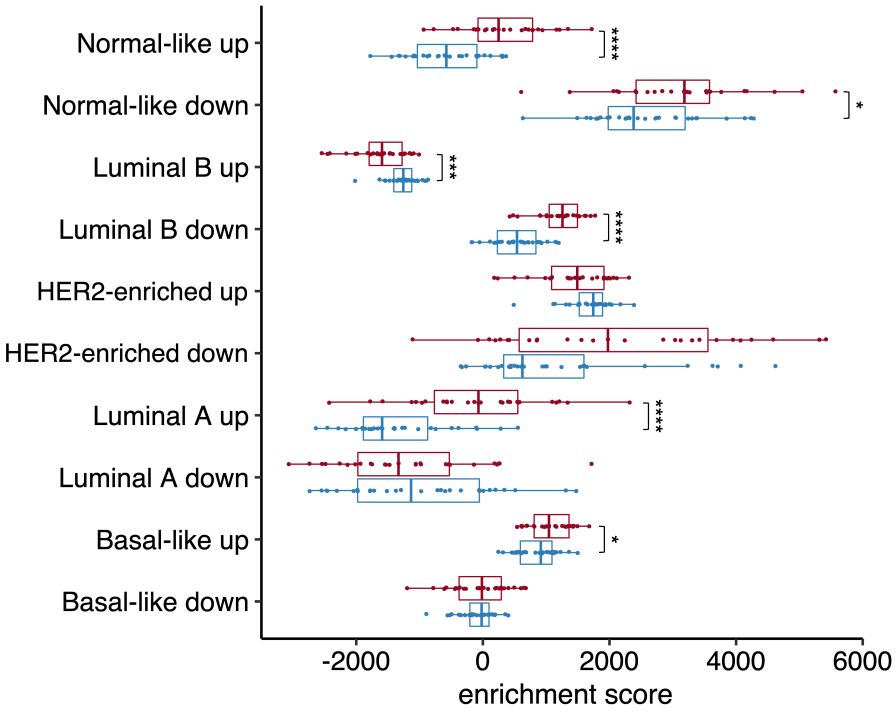


Figure 3.4: Figure S4. Differentially expressed (DE) gene analysis indicates the enrichment of hBLBC signatures in cBLMT of the validation set; related to Figure 2.3.

A. Heatmap of the row scaled log_2(TPM) values of the 761 DE genes between cBLMT and cNBLMT of the validation set, presented as described for Figure 2.3A.

B & C. Distributions of ssGSEA scores, presented as described for Figures 3B-C.

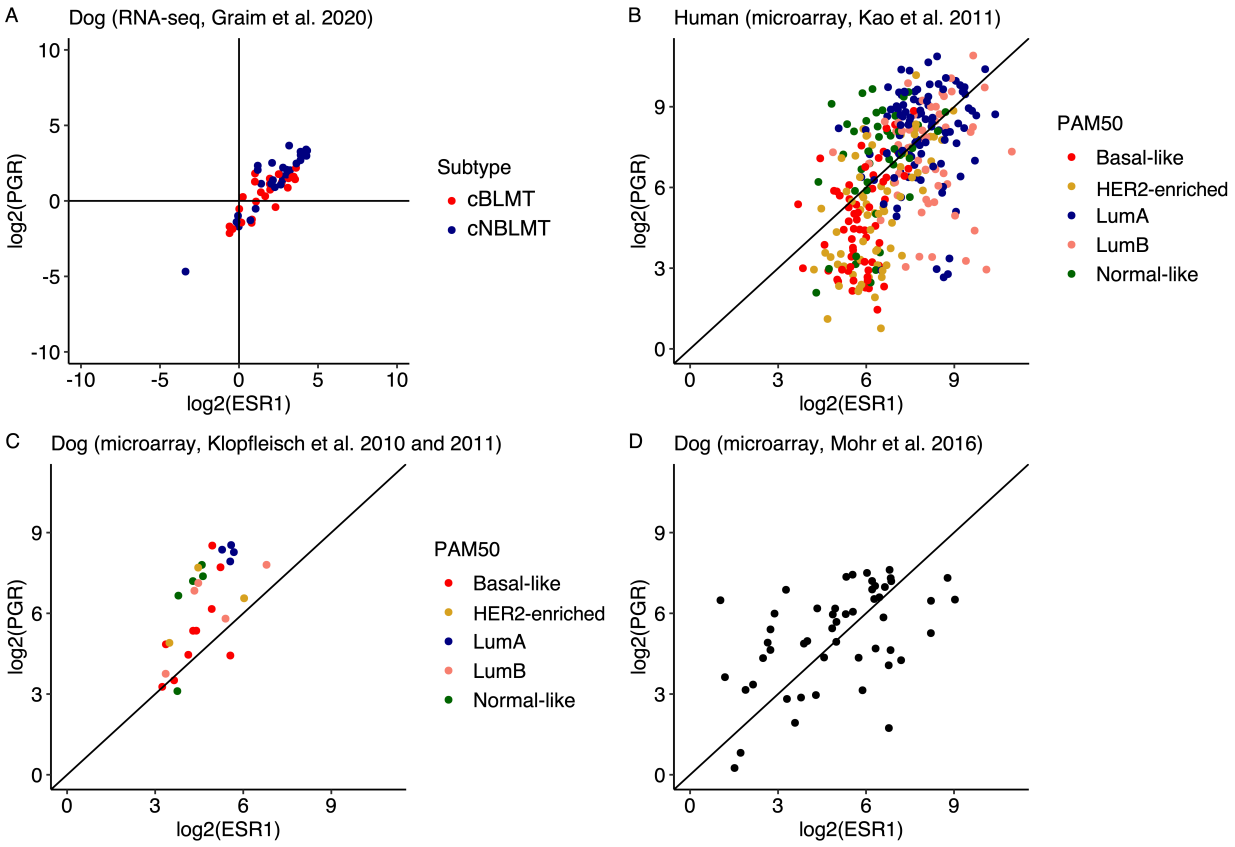


Figure 3.5: Figure S5. Canine tumors, especially cBLMTs, express PGR more abundantly than hBLBCs; related to Figure 2.4.

A. Scatter plot showing the $\log_2$(FPKM) values of ESR1 and PGR for canine tumors of the validation set (n=57) [67].

B-D. Scatter plots showing the $\log_2$-tranformed expression values of ESR1 and PGR from microarray studies of human breast cancers (n=327) [69] (B), as well as of canine mammary tumors by one group (n=40) (C) [70, 71] and by another group (n=51) (D) [95].

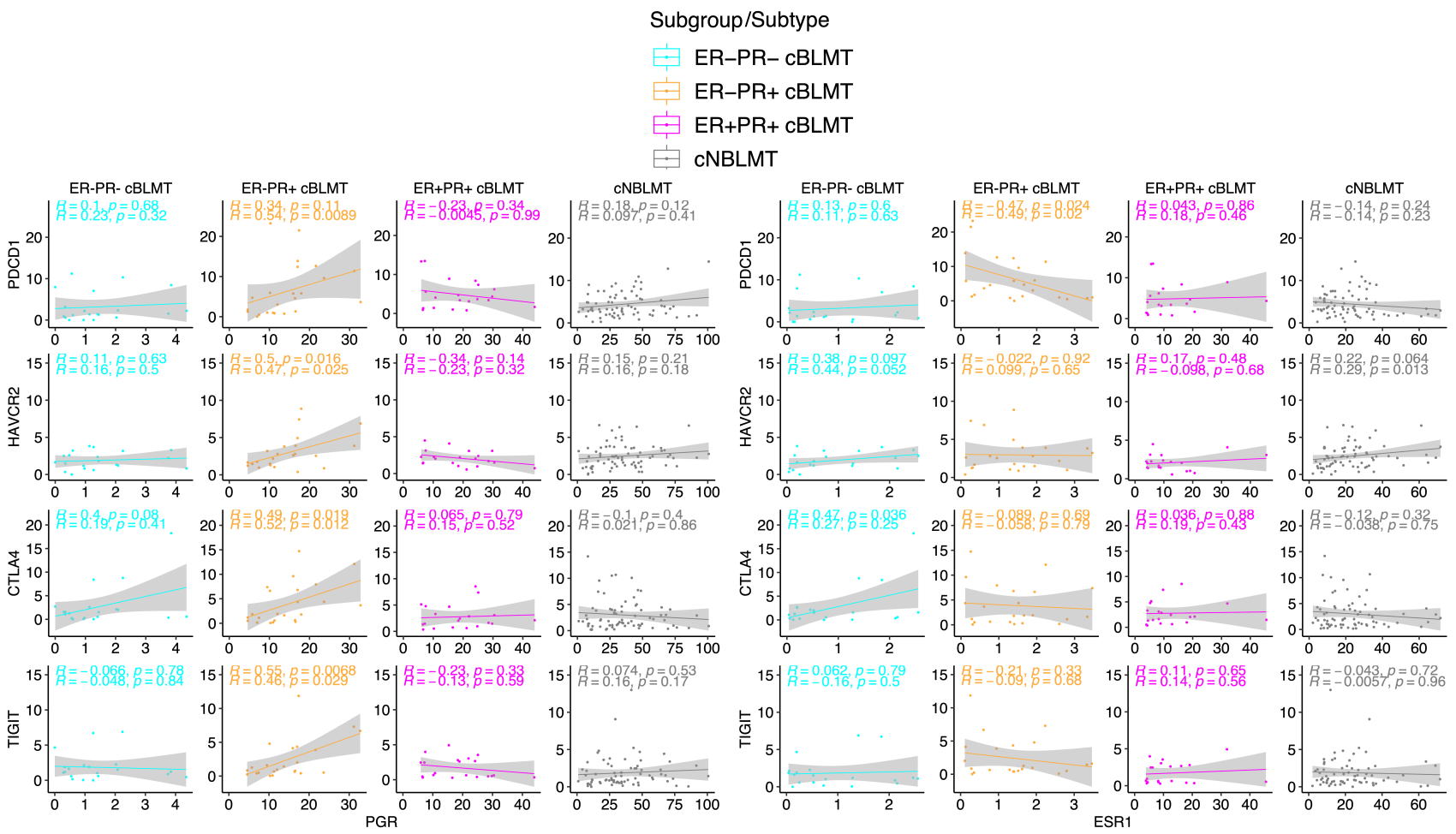


Figure 3.6: Figure S6. PGR correlates with a number of T-cell exhaustion signature genes in mRNA expression in ER-PR+ cBLMTs; related to Figure 2.6.

Pearson (top) and Spearman (bottom) correlation analysis between PGR (left) or ESR1 (right) and PDCD1, HAVCR2, CTLA4, or TIGIT in mRNA expression in each subgroup and subtype shown. Correlation coefficients, p-values, and linear regression line are indicated.

CHAPTER 4

CONCLUSIONS AND FUTURE DIRECTIONS

This research represents the first large-scale mRNA expression subtyping study developed independent of previous subtyping techniques in canine mammary tumors. The principal technique used to develop these subtypes, NMF, has the distinct advantage of being designed specifically for developing clusters based on gene expression data. The two subtypes of canine mammary tumors (cBLMT and cNBLMT) determined by NMF were also consistent across different subtyping strategies, indicating that these subtypes are inherent. cBLMTs match hBLBC in a plethora of features, such as histology (enriched in simple carcinomas, which better reflect hBLBC and human breast cancers in general), mutation pattern (more TP53 and less PIK3CA mutation), and functions of expression patterns (such as enriched cell cycle, proliferation, and gene sets expected to be upregulated in hBLBC). Most tellingly, cBLMT and hBLBC cluster together in a cross-species PAM50 analysis, away from any other human or canine subtype. To the contrary, cNBLMT does not match any human subtypes, in both histology (enriched in complex carcinomas, which are extremely rare in human breast cancer) and PAM50 analysis.

The results of this research are supported by a previous study by Liu et al. 2014, but contrast those of another by Bergholtz et al. 2022, which found that canine mammary tumors assigned to the Luminal A subtype were a better match for hLumA than canine tumors assigned to Basal-like were to hBLBC [33, 53]. While the current

study was done with RNA-seq data used in both studies, we addressed several factors that were not accounted for previously to support the validity of our conclusions. First, the histological difference between cNBLMT and any human breast cancer is significant, as cNBLMT are enriched in complex tumors. Second, this research directly compared the PAM50 expression pattern of human and dog tumors, and it is highly telling that cBLMT and hBLBC cluster together away from other subtypes, even when some PAM50 genes are missing from the canine reference genome. It is possible that the PAM50 subtype assignment may be different, as this study used the original assignment algorithm while Bergholtz et al. 2022 used an in-house algorithm [53]. This is unlikely, however, as all of the canine tumors assigned to the Luminal A subtype in this study were assigned the same in Bergholtz et al. 2022, but with differing results [53].

An unexpected result in this research was the high prevalence of ER-PR+ cBLMT tumors. Less than 4% of human breast cancers have this expression pattern and have only recently been accepted as a real phenotype as opposed to a sequencing or immunohistochemistry error [105-108]. When they do occur, ER-PR+ human breast cancers are usually assigned to the Basal-like subtype [105]. To the contrary, this expression pattern was common in our cBLMT, at around 1/3rd of cases, and this pattern was also present in other unrelated datasets. Previous works on the role of canine mammary tumors as animal models for human breast cancers have not addressed this difference. As previously mentioned, the ability of serum progesterone to induce PGR (and the long luteal phase of the dog) likely are the cause of this

expression pattern [52]. Supporting this conclusion is the fact that intact smaller breeds are more common in these datasets, as smaller dog breeds have more frequent estrus cycles compared to larger dogs (and thus undergoing more long-term progesterone exposure). While the estrus cycle stage was not noted in the samples used here, it presents the opportunity for a possible future study. Information on the estrus stage during canine mammary tumor development could answer this question.

The ER-PR+ cBLMTs maintained the same basal-like features of ER-PR- cBLMTs, such as PAM50 subtype assignment and enriched molecular features compared to cNBLMTs. Compared to ER-PR- cBLMTs, the ER-PR+ cBLMTs are enriched in a Wnt signaling and pluripotency pathway. Active Wnt signaling is a well-known feature of hBLBC [60, 96]. In particular, one of the genes in this pathway is WNT5B, which has been shown to drive a BLBC phenotype, both in vitro and in vivo, by activating both canonical and non-canonical Wnt signaling [19]. As this gene is specifically upregulated in ER-PR+ and ER+PR+ cBLMTs, this may be one of the chief reasons for these subgroups maintaining basal-like features, even with the difference in receptor expression.

The Wnt signaling pathway offers a valuable avenue to hBLBC treatment through small molecule inhibitors [109]. Targeting this pathway in both human breast cancer and canine mammary tumors has shown some efficacy in inhibiting cell cycle progression by the anti-parasitic drug Ivermectin [110, 111]. Two other small molecule inhibitors targeting Wnt signaling in hBLBC have been researched in canine mammary stem/progenitor cells: LGK-974 and iCRT3 [112]. LGK-974 is a phase I drug made for

targeting the Wnt pathway, and has been shown to inhibit Wnt5b (protein of WNT5B) secretion, while iCRT3 inhibits cell proliferation and increases apoptosis by antagonizing Wnt signaling [19, 109]. Both drugs have been shown to reduce the number of mammary stem/progenitor cells (which may be related to cancer stem cells) in canine mammary cell culture, but have not been used in vivo as tumor treatments [112]. Another treatment with some research in canine mammary tumors is the combination of Auranofin (arthritis medicine) and ICG-001 (Wnt-β catenin inhibitor) [113]. Combined use of Auranofin and ICG-001 has been shown to induce apoptosis in canine mammary cancer cell lines and mouse xenograft, but this study did not include a test on a spontaneous tumor model [113]. Auranofin has been used in combination with either vitamin C or immune checkpoint inhibitors in TNBC [114]. A potential future study could be to test the combination therapy of Auranofin and ICG-001 in cBLMTs as a proof of concept for its use in hBLBC.

A notable finding from this study is the strong positive correlation between PGR and the immune checkpoint genes PDCD1 (encoding PD-1) and CTLA4 only in ER-PR+ cBLMTs. This is a potential indicator of T cell exhaustion, a state of chronic overwork in T cells which makes them less able to kill their targeted cells [21]. T cell exhaustion is commonly detected in hBLBC, and importantly, targeting these immune checkpoint genes is one of the few treatments aside from chemotherapy in hBLBC [8, 97]. Canine cancers may offer a useful model for immunotherapy research, as evidenced by recent work in neoantigen detection in canine cancers [115]. The evidence in this current study is interesting and suggests the presence of T cell exhaustion in cBLMTs, but as

the only data used was WES and bulk RNA-seq, it cannot be considered totally conclusive. A more focused study on the immune landscape and microenvironment in cBLMTs could better elucidate its role as research model for T cell exhaustion.

Unexpectedly, PGR, unlike ESR1, was independently positively and negatively correlated with larger numbers of genes than in human. The negatively correlated genes are potentially more interesting, as no genes were significantly correlated with PGR in human breast cancers. These genes negatively correlated with PGR expression could indicate an ER-independant role of PGR in gene silencing. The previously mentioned pattern of progesterone inducing PGR expression could be involved in an ER-independent role of PGR [52].

While there were no enriched functions of genes negatively correlated with PGR in cBLMT, genes involved in hypoxia and glycolysis were negatively correlated with PGR in cNBLMT. PGR and especially ESR1 are highly expressed in cNBLMT, and ER is degraded by hypoxia [116]. Hypoxia is therefore unlikely in cNBLMT, and it correlates with metastasis and a worse outcome overall [117, 118]. To the contrary, hypoxia signatures are upregulated in hBLBC and are accordingly enriched in cBLMT, suggesting another feature that cBLMT has in common with hBLBC [118].

Many genes correlated with ESR1 and/or PGR were targets of the transcription factor SUZ12, a component gene of the polycomb repressor complex 2 (PRC2). PRC2 is a protein complex that trimethylates lysine 27 of histone 3 (H3K27me3), which acts as a transcriptional silencer [119]. Interestingly, PGR is a known target of SUZ12, while ESR1 is not [120]. The role of PRC2 in canine mammary tumors is currently not well

known, but there is some work hinting at its role. Liu et al. 2014 found 35 downregulated chromatin modifier genes in complex carcinomas, and suggests an epigenetic basis for the development of complex tumors [33, 121]. Choi et al. found EZH2 (the catalytic subunit of PRC2) expression increased with malignancy in canine mammary tumors, albeit without information on any receptor expression [122]. In these canine cancers, EZH2 immunohistochemistry staining was not significantly different between complex and simple carcinomas [122]. Increased EZH2 is associated with worse prognosis in multiple human cancers, including BLBC [119]. EZH2 in hBLBC represses tissue inhibitor metalloproteinases, which allows more activity of matrix metallopeptidases MMP2 and 9 [119]. These two genes promote migration and metastatic potential of hBLBC [119]. Further studies on PRC2 in both hBLBC and canine mammary tumors could determine if it has a significant role in PGR-associated gene silencing.

In conclusion, this study detected two independent canine mammary tumor subtypes which are consistent across subtyping strategies. One of these subtypes, cBLMT, is a good molecular match for hBLBC in factors such as histology, functions of upregulated genes, and PAM50 gene expression. The other subtype, cNBLMT, is mostly PAM50 subtyped as Luminal A, but does not match human LumA cancers on the molecular level. Furthermore, some cBLMTs have an ER-PR+ expression pattern, but these tumors still maintain many other functions expected of hBLBC. ER-PR+ cBLMTs may be a valuable spontaneous cancer model in which to study Wnt signaling and T cell

exhaustion. Canine mammary tumors may also offer a model in which to study an ER-independent role of PR in gene silencing.

REFERENCES

1. Siegel RL, Miller KD, Wagle NS, Jemal A: **Cancer statistics, 2023**. *CA Cancer J Clin* 2023, **73**(1):17-48.

2. Parker JS, Mullins M, Cheang MC, Leung S, Voduc D, Vickery T, Davies S, Fauron C, He X, Hu Z *et al*: **Supervised risk predictor of breast cancer based on intrinsic subtypes**. *J Clin Oncol* 2009, **27**(8):1160-1167.

3. Yin L, Duan JJ, Bian XW, Yu SC: **Triple-negative breast cancer molecular subtyping and treatment progress**. *Breast Cancer Res* 2020, **22**(1):61.

4. Prat A, Adamo B, Cheang MC, Anders CK, Carey LA, Perou CM: **Molecular characterization of basal-like and non-basal-like triple-negative breast cancer**. *Oncologist* 2013, **18**(2):123-133.

5. Howlader N, Cronin KA, Kurian AW, Andridge R: **Differences in Breast Cancer Survival by Molecular Subtypes in the United States**. *Cancer Epidemiol Biomarkers Prev* 2018, **27**(6):619-626.

6. Milioli HH, Tishchenko I, Riveros C, Berretta R, Moscato P: **Basal-like breast cancer: molecular profiles, clinical features and survival outcomes**. *BMC Med Genomics* 2017, **10**(1):19.

7. Alluri P, Newman LA: **Basal-like and triple-negative breast cancers: searching for positives among many negatives**. *Surg Oncol Clin N Am* 2014, **23**(3):567-577.

8. Burguin A, Diorio C, Durocher F: **Breast Cancer Treatments: Updates and New Challenges**. *J Pers Med* 2021, **11**(8).

9. Kreike B, van Kouwenhove M, Horlings H, Weigelt B, Peterse H, Bartelink H, van de Vijver MJ: **Gene expression profiling and histopathological characterization of triple-negative/basal-like breast carcinomas**. *Breast Cancer Res* 2007, **9**(5):R65.

10. Botti G, Cantile M, Collina F, Cerrone M, Sarno S, Anniciello A, Di Bonito M: **Morphological and pathological features of basal-like breast cancer**. *Transl Cancer Res* 2019, **8**(Suppl 5):S503-S509.

11. Xiao W, Zheng S, Yang A, Zhang X, Zou Y, Tang H, Xie X: **Breast cancer subtypes and the risk of distant metastasis at initial diagnosis: a population-based study**. *Cancer Manag Res* 2018, **10**:5329-5338.

12. Limsakul P, Choochuen P, Charupanit G, Charupanit K: **Transcriptomic Analysis of Subtype-Specific Tyrosine Kinases as Triple Negative Breast Cancer Biomarkers**. *Cancers (Basel)* 2023, **15**(2).

13. Cancer Genome Atlas N: **Comprehensive molecular portraits of human breast tumours**. *Nature* 2012, **490**(7418):61-70.

14. Risom T, Langer EM, Chapman MP, Rantala J, Fields AJ, Boniface C, Alvarez MJ, Kendsersky ND, Pelz CR, Johnson-Camacho K *et al*: **Differentiation-state plasticity is a targetable resistance mechanism in basal-like breast cancer**. *Nat Commun* 2018, **9**(1):3815.

15. Pohl SG, Brook N, Agostino M, Arfuso F, Kumar AP, Dharmarajan A: **Wnt signaling in triple-negative breast cancer**. *Oncogenesis* 2017, **6**(4):e310.

16. Klemm F, Bleckmann A, Siam L, Chuang HN, Rietkotter E, Behme D, Schulz M, Schaffrinski M, Schindler S, Trumper L *et al*: **beta-catenin-independent WNT signaling in basal-like breast cancer and brain metastasis**. *Carcinogenesis* 2011, **32**(3):434-442.

17. Marotta LL, Almendro V, Marusyk A, Shipitsin M, Schemme J, Walker SR, Bloushtain-Qimron N, Kim JJ, Choudhury SA, Maruyama R *et al*: **The JAK2/STAT3 signaling pathway is required for growth of CD44(+)CD24(-) stem cell-like breast cancer cells in human tumors**. *J Clin Invest* 2011, **121**(7):2723-2735.

18. Sarrio D, Rodriguez-Pinilla SM, Hardisson D, Cano A, Moreno-Bueno G, Palacios J: **Epithelial-mesenchymal transition in breast cancer relates to the basal-like phenotype**. *Cancer Res* 2008, **68**(4):989-997.

19. Jiang S, Zhang M, Zhang Y, Zhou W, Zhu T, Ruan Q, Chen H, Fang J, Zhou F, Sun J *et al*: **WNT5B governs the phenotype of basal-like breast cancer by activating WNT signaling**. *Cell Commun Signal* 2019, **17**(1):109.

20. Pereira B, Chin SF, Rueda OM, Vollan HK, Provenzano E, Bardwell HA, Pugh M, Jones L, Russell R, Sammut SJ *et al*: **The somatic mutation profiles of 2,433 breast cancers refines their genomic and transcriptomic landscapes**. *Nat Commun* 2016, **7**:11479.

21. Wherry EJ: **T cell exhaustion**. *Nat Immunol* 2011, **12**(6):492-499.

22. Li X, Zhao L, Chen C, Nie J, Jiao B: **Can EGFR be a therapeutic target in breast cancer?** *Biochim Biophys Acta Rev Cancer* 2022, **1877**(5):188789.

23. Jung J, Kang E, Gwak JM, Seo AN, Park SY, Lee AS, Baek H, Chae S, Kim EK, Kim SW: **Association between basal-like phenotype and BRCA1/2 germline mutations in Korean breast cancer patients**. *Curr Oncol* 2016, **23**(5):298-303.

24. Weigman VJ, Chao HH, Shabalin AA, He X, Parker JS, Nordgard SH, Grushko T, Huo D, Nwachukwu C, Nobel A *et al*: **Basal-like Breast cancer DNA copy number losses identify genes involved in genomic instability, response to therapy, and patient survival**. *Breast Cancer Res Treat* 2012, **133**(3):865-880.

25. O'Meara TA, Tolaney SM: **Tumor mutational burden as a predictor of immunotherapy response in breast cancer**. *Oncotarget* 2021, **12**(5):394-400.

26. Zeng L, Li W, Chen CS: **Breast cancer animal models and applications**. *Zool Res* 2020, **41**(5):477-494.

27. Dow S: **A Role for Dogs in Advancing Cancer Immunotherapy Research**. *Front Immunol* 2019, **10**:2935.

28. Na D, Moon HG: **Patient-Derived Xenograft Models in Breast Cancer Research**. *Adv Exp Med Biol* 2021, **1187**:283-301.

29. Gray M, Meehan J, Martinez-Perez C, Kay C, Turnbull AK, Morrison LR, Pang LY, Argyle D: **Naturally-Occurring Canine Mammary Tumors as a Translational Model for Human Breast Cancer**. *Front Oncol* 2020, **10**:617.

30. Knapp DW, Dhawan D, Ostrander E: **"Lassie," "Toto," and fellow pet dogs: poised to lead the way for advances in cancer prevention**. *Am Soc Clin Oncol Educ Book* 2015:e667-672.

31. Meuten DJ: **Tumors in domestic animals**, 4th edn. Ames, Iowa: Iowa State University Press; 2002.

32. Siegel R, Naishadham D, Jemal A: **Cancer statistics, 2012**. *CA Cancer J Clin* 2012, **62**(1):10-29.

33. Liu D, Xiong H, Ellis AE, Northrup NC, Rodriguez CO, Jr., O'Regan RM, Dalton S, Zhao S: **Molecular homology and difference between spontaneous canine mammary cancer and human breast cancer**. *Cancer Res* 2014, **74**(18):5045-5056.

34. Kwon JY, Moskwa N, Kang W, Fan TM, Lee C: **Canine as a Comparative and Translational Model for Human Mammary Tumor**. *J Breast Cancer* 2023.

35. Pinho SS, Carvalho S, Cabral J, Reis CA, Gartner F: **Canine tumors: a spontaneous animal model of human carcinogenesis**. *Transl Res* 2012, **159**(3):165-172.

36. Lindblad-Toh K, Wade CM, Mikkelsen TS, Karlsson EK, Jaffe DB, Kamal M, Clamp M, Chang JL, Kulbokas EJ, 3rd, Zody MC *et al*: **Genome sequence, comparative analysis and haplotype structure of the domestic dog**. *Nature* 2005, **438**(7069):803-819.

37. Beck J, Ren L, Huang S, Berger E, Bardales K, Mannheimer J, Mazcko C, LeBlanc A: **Canine and murine models of osteosarcoma**. *Vet Pathol* 2022, **59**(3):399-414.

38. Amin SB, Anderson KJ, Boudreau CE, Martinez-Ledesma E, Kocakavuk E, Johnson KC, Barthel FP, Varn FS, Kassab C, Ling X *et al*: **Comparative**

Molecular Life History of Spontaneous Canine and Human Gliomas. *Cancer Cell* 2020, **37**(2):243-257 e247.

39. Naik S, Galyon GD, Jenks NJ, Steele MB, Miller AC, Allstadt SD, Suksanpaisan L, Peng KW, Federspiel MJ, Russell SJ *et al*: **Comparative Oncology Evaluation of Intravenous Recombinant Oncolytic Vesicular Stomatitis Virus Therapy in Spontaneous Canine Cancer**. *Mol Cancer Ther* 2018, **17**(1):316-326.

40. Nagaya T, Okuyama S, Ogata F, Maruoka Y, Knapp DW, Karagiannis SN, Fazekas-Singer J, Choyke PL, LeBlanc AK, Jensen-Jarolim E *et al*: **Near infrared photoimmunotherapy targeting bladder cancer with a canine anti-epidermal growth factor receptor (EGFR) antibody**. *Oncotarget* 2018, **9**(27):19026-19038.

41. Goldschmidt M, Pena L, Rasotto R, Zappulli V: **Classification and grading of canine mammary tumors**. *Vet Pathol* 2011, **48**(1):117-131.

42. Alsaihati BA, Ho KL, Watson J, Feng Y, Wang T, Dobbin KK, Zhao S: **Canine tumor mutational burden is correlated with TP53 mutation across tumor types and breeds**. *Nat Commun* 2021, **12**(1):4670.

43. Kim TM, Yang IS, Seung BJ, Lee S, Kim D, Ha YJ, Seo MK, Kim KK, Kim HS, Cheong JH *et al*: **Cross-species oncogenic signatures of breast cancer in canine mammary tumors**. *Nat Commun* 2020, **11**(1):3616.

44. Borge KS, Nord S, Van Loo P, Lingjaerde OC, Gunnes G, Alnaes GI, Solvang HK, Luders T, Kristensen VN, Borresen-Dale AL *et al*: **Canine Mammary Tumours Are Affected by Frequent Copy Number Aberrations, including Amplification of MYC and Loss of PTEN**. *PLoS One* 2015, **10**(5):e0126371.

45. Nakagaki KY, Nunes MM, Garcia APV, Nunes FC, Schmitt F, Cassali GD: **Solid Carcinoma of the Canine Mammary Gland: a Histological Type or Tumour Cell Arrangement?** *J Comp Pathol* 2022, **190**:1-12.

46. Tan PH, Ellis IO: **Myoepithelial and epithelial-myoepithelial, mesenchymal and fibroepithelial breast lesions: updates from the WHO Classification of Tumours of the Breast 2012**. *J Clin Pathol* 2013, **66**(6):465-470.

47. Hayes MM: **Adenomyoepithelioma of the breast: a review stressing its propensity for malignant transformation**. *J Clin Pathol* 2011, **64**(6):477-484.

48. Grenier K, Altinel G, Dastani Z, Omeroglu A: **Epithelial-Myoepithelial Carcinoma of the Breast with Rhabdoid Features**. *Case Rep Pathol* 2020, **2020**:8879035.

49. Oliver R, Pillarisetty LS: **Anatomy, Abdomen and Pelvis, Ovary Corpus Luteum**. In: *StatPearls.* edn. Treasure Island (FL); 2022.

50. Papa PC, Kowalewski MP: **Factors affecting the fate of the canine corpus luteum: Potential contributors to pregnancy and non-pregnancy**. *Theriogenology* 2020, **150**:339-346.

51. Graf KJ: **Serum oestrogen, progesterone and prolactin concentrations in cyclic, pregnant and lactating beagle dogs**. *J Reprod Fertil* 1978, **52**(1):9-14.

52. Diep CH, Ahrendt H, Lange CA: **Progesterone induces progesterone receptor gene (PGR) expression via rapid activation of protein kinase pathways required for cooperative estrogen receptor alpha (ER) and progesterone receptor (PR) genomic action at ER/PR target genes**. *Steroids* 2016, **114**:48-58.

53. Bergholtz H, Lien T, Lingaas F, Sorlie T: **Comparative analysis of the molecular subtype landscape in canine and human mammary gland tumors**. *J Mammary Gland Biol Neoplasia* 2022, **27**(2):171-183.

54. Kim KK, Seung BJ, Kim D, Park HM, Lee S, Song DW, Lee G, Cheong JH, Nam H, Sur JH *et al*: **Whole-exome and whole-transcriptome sequencing of canine mammary gland tumors**. *Sci Data* 2019, **6**(1):147.

55. Trepanier LA, Ray K, Winand NJ, Spielberg SP, Cribb AE: **Cytosolic arylamine N-acetyltransferase (NAT) deficiency in the dog and other canids due to an absence of NAT genes**. *Biochem Pharmacol* 1997, **54**(1):73-80.

56. Polyak K: **Heterogeneity in breast cancer**. *J Clin Invest* 2011, **121**(10):3786-3788.

57. Ciriello G, Gatza ML, Beck AH, Wilkerson MD, Rhie SK, Pastore A, Zhang H, McLellan M, Yau C, Kandoth C *et al*: **Comprehensive Molecular Portraits of Invasive Lobular Breast Cancer**. *Cell* 2015, **163**(2):506-519.

58. Curtis C, Shah SP, Chin SF, Turashvili G, Rueda OM, Dunning MJ, Speed D, Lynch AG, Samarajiwa S, Yuan Y *et al*: **The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups**. *Nature* 2012, **486**(7403):346-352.

59. Fu Z, Chen S, Zhu Y, Zhang D, Xie P, Jiao Q, Chi J, Xu S, Xue Y, Lu X *et al*: **Proteolytic regulation of CD73 by TRIM21 orchestrates tumor immunogenicity**. *Sci Adv* 2023, **9**(1):eadd6626.

60. Sharma M, Castro-Piedras I, Rodgers AD, Pruitt K: **Genomic profiling of DVL-1 and its nuclear role as a transcriptional regulator in triple negative breast cancer**. *Genes Cancer* 2021, **12**:77-95.

61. Liao L, Zhang YL, Deng L, Chen C, Ma XY, Andriani L, Yang SY, Hu SY, Zhang FL, Shao ZM *et al*: **Protein Phosphatase 1 Subunit PPP1R14B Stabilizes STMN1 to Promote Progression and Paclitaxel Resistance in Triple-Negative Breast Cancer**. *Cancer Res* 2023, **83**(3):471-484.

62. Reddy TP, Rosato RR, Li X, Moulder S, Piwnica-Worms H, Chang JC: **A comprehensive overview of metaplastic breast cancer: clinical features and molecular aberrations**. *Breast Cancer Res* 2020, **22**(1):121.

63. Bu W, Liu ZY, Jiang WY, Nagi C, Huang SX, Edwards DP, Jo E, Mo QX, Creighton CJ, Hilsenbeck SG *et al*: **Mammary Precancerous Stem and Non-Stem Cells Evolve into Cancers of Distinct Subtypes**. *Cancer Research* 2019, **79**(1):61-71.

64. Thamm DH: **Canine Cancer: Strategies in Experimental Therapeutics**. *Front Oncol* 2019, **9**:1257.

65. Sassi F, Benazzi C, Castellani G, Sarli G: **Molecular-based tumour subtypes of canine mammary carcinomas assessed by immunohistochemistry**. *BMC Vet Res* 2010, **6**:5.

66. Sorenmo KU, Kristiansen VM, Cofone MA, Shofer FS, Breen AM, Langeland M, Mongil CM, Grondahl AM, Teige J, Goldschmidt MH: **Canine mammary gland tumours; a histological continuum from benign to malignant; clinical and histopathological evidence**. *Vet Comp Oncol* 2009, **7**(3):162-172.

67. Graim K, Gorenshteyn D, Robinson DG, Carriero NJ, Cahill JA, Chakrabarti R, Goldschmidt MH, Durham AC, Funk J, Storey JD *et al*: **Modeling molecular development of breast cancer in canine mammary tumors**. *Genome Res* 2020.

68. Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, Sun Y, Jacobsen A, Sinha R, Larsson E *et al*: **Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal**. *Sci Signal* 2013, **6**(269):pl1.

69. Kao KJ, Chang KM, Hsu HC, Huang AT: **Correlation of microarray-based breast cancer molecular subtypes and clinical outcomes: implications for treatment optimization**. *BMC Cancer* 2011, **11**:143.

70. Klopfleisch R, Lenze D, Hummel M, Gruber AD: **The metastatic cascade is reflected in the transcriptome of metastatic canine mammary carcinomas**. *Vet J* 2011, **190**(2):236-243.

71. Klopfleisch R, Lenze D, Hummel M, Gruber AD: **Metastatic canine mammary carcinomas can be identified by a gene expression profile that partly overlaps with human breast cancer profiles**. *BMC Cancer* 2010, **10**:618.

72. Gautier L, Cope L, Bolstad BM, Irizarry RA: **affy—analysis of Affymetrix GeneChip data at the probe level**. *Bioinformatics* 2004, **20**(3):307-315.

73. Wang J, Wang T, Sun Y, Feng Y, Kisseberth WC, Henry CJ, Mok I, Lana SE, Dobbin K, Northrup N *et al*: **Proliferative and Invasive Colorectal Tumors in Pet Dogs Provide Unique Insights into Human Colorectal Cancer**. *Cancers (Basel)* 2018, **10**(9).

74. Wang T, Kwon SH, Peng X, Urdy S, Lu Z, Schmitz RJ, Dalton S, Mostov KE, Zhao S: **A Qualitative Change in the Transcriptome Occurs after the First**

Cell Cycle and Coincides with Lumen Establishment during MDCKII Cystogenesis. *iScience* 2020, **23**(10):101629.

75. Feng Y, Hess, P.R, Tompkins, S.M., Hildebrand, W.H, Zhao, S: **A Kmer-based paired-end read de novo assembler and genotyper for canine MHC class I genotyping**. *iScience* 2023, **doi: https://doi.org/10.1016/j.isci.2023.105996.**

76. Kim D, Paggi JM, Park C, Bennett C, Salzberg SL: **Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype**. *Nat Biotechnol* 2019, **37**(8):907-915.

77. Ewels P, Magnusson M, Lundin S, Kaller M: **MultiQC: summarize analysis results for multiple tools and samples in a single report**. *Bioinformatics* 2016, **32**(19):3047-3048.

78. Liao Y, Smyth GK, Shi W: **featureCounts: an efficient general purpose program for assigning sequence reads to genomic features**. *Bioinformatics* 2014, **30**(7):923-930.

79. Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg SL, Wold BJ, Pachter L: **Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation**. *Nat Biotechnol* 2010, **28**(5):511-515.

80. Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD: **The sva package for removing batch effects and other unwanted variation in high-throughput experiments**. *Bioinformatics* 2012, **28**(6):882-883.

81. Zhang Y, Parmigiani G, Johnson WE: **ComBat-seq: batch effect adjustment for RNA-seq count data**. *NAR Genom Bioinform* 2020, **2**(3):lqaa078.

82. Gaujoux R, Seoighe C: **A flexible R package for nonnegative matrix factorization**. *BMC Bioinformatics* 2010, **11**:367.

83. Suzuki R, Shimodaira H: **Pvclust: an R package for assessing the uncertainty in hierarchical clustering**. *Bioinformatics* 2006, **22**(12):1540-1542.

84. Wilkerson MD, Hayes DN: **ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking**. *Bioinformatics* 2010, **26**(12):1572-1573.

85. R Core Team R, Team RC: **R: a language and environment for statistical computing. R Foundation for Statistical Computing; 2020**. In.; 2021.

86. Galili T: **dendextend: an R package for visualizing, adjusting and comparing trees of hierarchical clustering**. *Bioinformatics* 2015, **31**(22):3718-3720.

87. Coombes KR: **Genetic algorithms for feature selection**. 2017.

88. Love MI, Huber W, Anders S: **Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2**. *Genome Biol* 2014, **15**(12):550.

89. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES *et al*: **Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles**. *Proc Natl Acad Sci U S A* 2005, **102**(43):15545-15550.

90. Sherman BT, Hao M, Qiu J, Jiao X, Baseler MW, Lane HC, Imamichi T, Chang W: **DAVID: a web server for functional enrichment analysis and functional annotation of gene lists (2021 update)**. *Nucleic Acids Res* 2022.

91. Smid M, Wang Y, Zhang Y, Sieuwerts AM, Yu J, Klijn JG, Foekens JA, Martens JW: **Subtypes of breast cancer show preferential site of relapse**. *Cancer Res* 2008, **68**(9):3108-3114.

92. Kuehn H, Liberzon A, Reich M, Mesirov JP: **Using GenePattern for gene expression analysis**. *Curr Protoc Bioinformatics* 2008, **Chapter 7**:Unit 7 12.

93. Xie Z, Bailey A, Kuleshov MV, Clarke DJB, Evangelista JE, Jenkins SL, Lachmann A, Wojciechowicz ML, Kropiwnicki E, Jagodnik KM *et al*: **Gene Set Knowledge Discovery with Enrichr**. *Curr Protoc* 2021, **1**(3):e90.

94. Perou CM, Sorlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, Pollack JR, Ross DT, Johnsen H, Akslen LA *et al*: **Molecular portraits of human breast tumours**. *Nature* 2000, **406**(6797):747-752.

95. Mohr A, Luder Ripoli F, Hammer SC, Willenbrock S, Hewicker-Trautwein M, Kielbowicz Z, Murua Escobar H, Nolte I: **Hormone Receptor Expression Analyses in Neoplastic and Non-Neoplastic Canine Mammary Tissue by a Bead Based Multiplex Branched DNA Assay: A Gene Expression Study in Fresh Frozen and Formalin-Fixed, Paraffin-Embedded Samples**. *PLoS One* 2016, **11**(9):e0163311.

96. Khramtsov AI, Khramtsova GF, Tretiakova M, Huo DZ, Olopade OI, Goss KH: **Wnt/beta-Catenin Pathway Activation Is Enriched in Basal-Like Breast Cancers and Predicts Poor Outcome**. *Am J Pathol* 2010, **176**(6):2911-2920.

97. Guo L, Cao C, Goswami S, Huang X, Ma L, Guo Y, Yang B, Li T, Chi Y, Zhang X *et al*: **Tumoral PD-1hiCD8+ T cells are partially exhausted and predict favorable outcome in triple-negative breast cancer**. *Clin Sci (Lond)* 2020, **134**(7):711-726.

98. Shutan Xu YF, Shaying Zhao: **Proteins with evolutionarily hypervariable domains are associated with immune response and better survival of basal-like breast cancer patients**. *Computational and Structural Biotechnology Journal* 2019.

99. Zheng L, Qin S, Si W, Wang A, Xing B, Gao R, Ren X, Wang L, Wu X, Zhang J *et al*: **Pan-cancer single-cell landscape of tumor-infiltrating T cells**. *Science* 2021, **374**(6574):abe6474.

100. Lapidus RG, Nass SJ, Davidson NE: **The loss of estrogen and progesterone receptor gene expression in human breast cancer**. *J Mammary Gland Biol Neoplasia* 1998, **3**(1):85-94.

101. Burstein HJ: **Systemic Therapy for Estrogen Receptor-Positive, HER2-Negative Breast Cancer**. *N Engl J Med* 2020, **383**(26):2557-2570.

102. Concannon PW: **Reproductive cycles of the domestic bitch**. *Anim Reprod Sci* 2011, **124**(3-4):200-210.

103. Sleeckx N, de Rooster H, Veldhuis Kroeze EJ, Van Ginneken C, Van Brantegem L: **Canine mammary tumours, an overview**. *Reprod Domest Anim* 2011, **46**(6):1112-1131.

104. Chow A, Perica K, Klebanoff CA, Wolchok JD: **Clinical implications of T cell exhaustion for cancer immunotherapy**. *Nat Rev Clin Oncol* 2022, **19**(12):775-790.

105. Beltjens F, Molly D, Bertaut A, Richard C, Desmoulins I, Loustalot C, Charon-Barra C, Courcet E, Bergeron A, Ladoire S *et al*: **ER-/PR+ breast cancer: A distinct entity, which is morphologically and molecularly close to triple-negative breast cancer**. *Int J Cancer* 2021, **149**(1):200-213.

106. Zheng H, Ge C, Lin H, Wu L, Wang Q, Zhou S, Tang W, Zhang X, Jin X, Xu X *et al*: **Estrogen receptor-negative/progesterone receptor-positive and her-2-negative breast cancer might no longer be classified as hormone receptor-positive breast cancer**. *Int J Clin Oncol* 2022, **27**(7):1145-1153.

107. Foley NM, Coll JM, Lowery AJ, Hynes SO, Kerin MJ, Sheehan M, Brodie C, Sweeney KJ: **Re-Appraisal of Estrogen Receptor Negative/Progesterone Receptor Positive (ER-/PR+) Breast Cancer Phenotype: True Subtype or Technical Artefact?** *Pathol Oncol Res* 2018, **24**(4):881-884.

108. Schroth W, Winter S, Buttner F, Goletz S, Faisst S, Brinkmann F, Saladores P, Heidemann E, Ott G, Gerteis A *et al*: **Clinical outcome and global gene expression data support the existence of the estrogen receptor-negative/progesterone receptor-positive invasive breast cancer phenotype**. *Breast Cancer Res Treat* 2016, **155**(1):85-97.

109. Zheng S, Liu J, Wu Y, Huang TL, Wang G: **Small-molecule inhibitors of Wnt signaling pathway: towards novel anticancer therapeutics**. *Future Med Chem* 2015, **7**(18):2485-2505.

110. Tang M, Hu X, Wang Y, Yao X, Zhang W, Yu C, Cheng F, Li J, Fang Q: **Ivermectin, a potential anticancer drug derived from an antiparasitic drug**. *Pharmacol Res* 2021, **163**:105207.

111. Diao H, Cheng N, Zhao Y, Xu H, Dong H, Thamm DH, Zhang D, Lin D: **Ivermectin inhibits canine mammary tumor growth by regulating cell cycle progression and WNT signaling**. *BMC Vet Res* 2019, **15**(1):276.

112. Bussche L, Rauner G, Antonyak M, Syracuse B, McDowell M, Brown AMC, Cerione RA, Van de Walle GR: **Microvesicle-mediated Wnt/beta-Catenin Signaling Promotes Interspecies Mammary Stem/Progenitor Cell Growth**. *J Biol Chem* 2016, **291**(47):24390-24405.

113. Lin Z, Lin Z, Zhao Y, Cheng N, Zhang D, Lin J, Zhang H, Lin D: **Auranofin and ICG-001 Emerge Synergistic Anti-tumor Effect on Canine Breast Cancer by Inducing Apoptosis via Mitochondrial Pathway**. *Front Vet Sci* 2021, **8**:772687.

114. Hatem E, Azzi S, El Banna N, He T, Heneman-Masurel A, Vernis L, Baille D, Masson V, Dingli F, Loew D *et al*: **Auranofin/Vitamin C: A Novel Drug Combination Targeting Triple-Negative Breast Cancer**. *J Natl Cancer Inst* 2018.

115. Feng Y, Hess PR, Tompkins SM, Hildebrand WH, Zhao S: **A Kmer-based paired-end read de novo assembler and genotyper for canine MHC class I genotyping**. *iScience* 2023, **26**(2):105996.

116. Fuady JH, Gutsche K, Santambrogio S, Varga Z, Hoogewijs D, Wenger RH: **Estrogen-dependent downregulation of hypoxia-inducible factor (HIF)-2alpha in invasive breast cancer cells**. *Oncotarget* 2016, **7**(21):31153-31165.

117. Zhang Y, Zhang H, Wang M, Schmid T, Xin Z, Kozhuharova L, Yu WK, Huang Y, Cai F, Biskup E: **Hypoxia in Breast Cancer-Scientific Translation to Therapeutic and Diagnostic Clinical Applications**. *Front Oncol* 2021, **11**:652266.

118. Ye IC, Fertig EJ, DiGiacomo JW, Considine M, Godet I, Gilkes DM: **Molecular Portrait of Hypoxia in Breast Cancer: A Prognostic Signature and Novel HIF-Regulated Genes**. *Mol Cancer Res* 2018, **16**(12):1889-1901.

119. Martin CJ, Moorehead RA: **Polycomb repressor complex 2 function in breast cancer (Review)**. *Int J Oncol* 2020, **57**(5):1085-1094.

120. Ben-Porath I, Thomson MW, Carey VJ, Ge R, Bell GW, Regev A, Weinberg RA: **An embryonic stem cell-like gene expression signature in poorly differentiated aggressive human tumors**. *Nat Genet* 2008, **40**(5):499-507.

121. Borges BDN: **Epigenetic alterations in canine mammary cancer**. *Genet Mol Biol* 2022, **45**(3 Suppl 1):e20220131.

122. Choi HJ, Jang S, Ryu JE, Lee HJ, Lee HB, Ahn WS, Kim HJ, Lee HJ, Lee HJ, Gong GY *et al*: **Significance of EZH2 expression in canine mammary tumors**. *BMC Vet Res* 2016, **12**(1):164.