



Comparative analysis of the molecular subtype landscape in canine and human mammary gland tumors

Helga Bergholtz¹ · Tonje Lien¹ · Frode Lingaas² · Therese Sørli^{1,3}

Received: 11 January 2022 / Accepted: 6 July 2022 / Published online: 6 August 2022
© The Author(s) 2022

Abstract

Breast cancers in humans belong to one of several intrinsic molecular subtypes each with different tumor biology and different clinical impact. Mammary gland tumors in dogs are proposed as a relevant comparative model for human breast cancer; however, it is still unclear whether the intrinsic molecular subtypes have the same significance in dogs and humans. Using publicly available data, we analyzed gene expression and whole-exome sequencing data from 158 canine mammary gland tumors. We performed molecular subtyping using the PAM50 method followed by subtype-specific comparisons of gene expression characteristics, mutation patterns and copy number profiles between canine tumors and human breast tumors from The Cancer Genome Atlas (TCGA) breast cancer cohort (n = 1097). We found that luminal A canine tumors greatly resemble luminal A human tumors both in gene expression characteristics, mutations and copy number profiles. Also, the basal-like canine and human tumors were relatively similar, with low expression of luminal epithelial markers and high expression of genes involved in cell proliferation. There were, however, distinct differences in immune-related gene expression patterns in basal-like tumors between the two species. Characteristic HER2-enriched and luminal B subtypes were not present in the canine cohort, and we found no tumors with high-level ERBB2 amplifications. Benign and malignant canine tumors displayed similar PAM50 subtype characteristics. Our findings indicate that deeper understanding of the different molecular subtypes in canine mammary gland tumors will further improve the value of canines as comparative models for human breast cancer.

Keywords Breast cancer · Molecular subtypes · Comparative models · Canine tumor models

Abbreviations

CMGT	Canine Mammary Gland Tumor
ER	Estrogen receptor
ERBB2	Erb-b2 receptor tyrosine kinase 2
ESR1	Estrogen receptor 1
IHC	Immunohistochemistry
HER2	Human Epidermal Growth Factor 2
HOC	Hallmarks of cancer
PCA	Principal Component Analysis
PR	Progesterone receptor
ssGSEA	Single Sample Gene Set Enrichment Analyses

TCGA	The Cancer Genome Atlas
TMB	Tumor Mutation Burden

Introduction

Pre-clinical studies of breast cancer often include experimental rodent models; however, there are important biological differences between humans and rodents with regard to mammary gland tumor development and progression, as well as hormone dependency that complicate the transferability of pre-clinical findings into clinical practice for human cancer [1, 2]. There is, therefore, a need for animal models that mimic human breast cancer more precisely. The dog has emerged as a potential candidate and has been demonstrated to be an efficient model for human cancers [3–6], including breast cancer [7–9].

Mammary gland tumors are common in dogs with an incidence rate twice as high as in humans, but with large differences between breeds [10–12]. In contrast to many rodent models, mammary gland tumors in dogs occur

✉ Therese Sørli
therese.sorlie@rr-research.no

¹ Department of Cancer Genetics, Institute for Cancer Research, Oslo University Hospital, Oslo, Norway

² Department of Preclinical Sciences and Pathology, Faculty of Veterinary Medicine, Norwegian University of Life Sciences, Ås, Norway

³ Institute of Clinical Medicine, Faculty of Medicine, University of Oslo, Oslo, Norway

spontaneously without intentional genetic or chemical manipulation, and importantly, they occur in an intact immune environment. There are many similarities between canine and human mammary gland cancers, such as age of onset, hormonal influence on tumor development, disease course, clinical outcome parameters (such as tumor size, clinical stage, and lymph node invasion), mode of metastatic spread, and numerous molecular markers and genetic risk factors [7, 8, 13, 14]. Importantly, dogs often live together with humans and the two species are therefore exposed to many of the same environmental factors that can influence cancer development. There are also differences between the two species; while human breast cancer is dominated by epithelial tumors, canine mammary gland tumors (CMGTs) frequently also contain myoepithelial components [15, 16]. Tumor mutational burden (TMB) is generally lower in canine tumors, and found to be comparable to TMB in pediatric tumors [17, 18]. In addition, diagnosis generally happens later in the progression course in dogs than in humans, since it depends on discovering palpable lesions as opposed to detection by screening mammography in humans.

In human breast cancer, tumors belong to one of several intrinsic molecular subtypes that have different prognostic and predictive impacts [19, 20]. *Basal-like* tumors are predominantly triple negative i.e., they lack expression of estrogen receptor (ER), progesterone receptor (PR) and human epidermal growth factor 2 (HER2). Basal-like tumors are usually highly proliferative and are associated with a poor prognosis. *HER2-enriched* tumors often show high expression of HER2 (also known as erb-b2 receptor tyrosine kinase 2 and encoded by the *ERBB2* gene) and are also poor prognosis tumors; however, these may benefit from HER2-targeted therapy. *Luminal A* tumors are usually ER-positive tumors that proliferate slowly and have a relatively good prognosis, while *luminal B* tumors are also ER-positive, but are more proliferative, and have a significantly worse prognosis than luminal A tumors [21].

The intrinsic subtypes can be determined by the PAM50 method, a nearest centroid classifier in which gene expression of 50 genes is correlated to previously defined subtype centroids [21–23]. The PAM50 method, including the 50 genes and the associated centroids were identified from studies of human breast cancer, however, this method captures core biological characteristics of tumors that are relevant in both humans and dogs. There are few studies on the applicability and relevance of PAM50 subtyping in canine mammary gland tumors; however, in a study by Graim et al., PAM50 subtyping was performed on CMGTs from 16 canine patients [9]. They showed considerable resemblance between canine and human tumors of the same subtype at both the transcriptional and mutational levels.

In a comprehensive study by Kim et al. including gene expression analyses of 158 CMGTs, clustering was

performed across the PAM50 genes, but without calling the PAM50 subtypes [18, 24]. Their analysis revealed a strong resemblance between CMGTs and human breast tumors including the well-known, clear distinction between basal-like and luminal-like tumors. In the current study, we utilize this unique data resource that includes both exome DNA and whole-transcriptome RNA sequencing data to thoroughly explore the molecular subtype landscape in canine mammary gland tumors. We use the well-established gene expression-based subtyping method PAM50 and assess its applicability as a tool for stratification of tumors when performing canine/human comparative analyses. Using gene expression, mutation and copy number data we performed a comprehensive subtype-specific comparison of CMGT and human breast tumors from The Cancer Genome Atlas (TCGA) and confirmed that the intrinsic subtypes also represent distinct biology in CMGT. We find that luminal A and basal-like are the two main subtypes in CMGT and underline the necessity of stratifying tumors by molecular subtypes when using CMGTs as models for human breast cancer.

Material and Methods

Datasets

Canine Mammary Gland Tumor Datasets

Data used in these analyses were generated by Kim et al. at Yonsei University College of Medicine, Seoul, Republic of Korea [18, 24]. In their studies, CMGTs were subject to DNA (whole-exome) and RNA sequencing. The RNA sequencing dataset, including clinical information, was retrieved from NCBI's Gene Expression Omnibus [25] under the accession number GSE119810 [26]. Clinical characteristics are presented in Supplementary File 1. Variant calling results were downloaded as a VCF file [27] and segmented copy number data were kindly provided by Kim [18]. Details on data pre-processing can be found in Kim et al. and include sequence mapping to the CanFam3.1 reference genome using BWA-MEM2 (DNA) or TopHat (RNA), and variant calling (including filtering out germline variants) using GATK4.0. For RNA sequencing data, FPKM (Fragments Per Kilobase of Transcript per Million) was calculated.

Based on PCA analyses, we identified one sample (CMT.774) as an outlier and hence removed this from the dataset. Mutation and copy number analyses were performed on tumors overlapping with the gene expression dataset. For simplicity, the 13 original histological diagnoses were combined into five main histological categories: *Simple carcinoma*, *complex carcinoma*, *mixed tumor*, *benign* and *other*

histology. Both original and new categories are shown in Supplementary File 1.

Human Breast Cancer Dataset

Data from the TCGA database [28, 29] were downloaded from the UCSC Xena Platform [30]. The TCGA dataset consists of 1097 human breast tumors.

Dataset Merging

To enable cross-species comparison and PAM50 subtyping, the CMGT and TCGA RNAseq datasets were merged. Genes present in both datasets were identified ($n = 13,071$). Next, a Z-score was calculated per gene in each of the two datasets separately before merging. Principal component analysis (PCA) of the merged data showed no distinct separation according to species.

PAM50 Subtyping

Intrinsic molecular subtyping was performed by applying the PAM50 method to the merged dataset using centroids for the four main intrinsic subtypes: basal-like, HER2-enriched, luminal A and luminal B obtained from Parker et al. [22, 23]. We excluded the normal-like subtype from this study due to uncertainty of its value as a bona-fide subtype [22, 31, 32]. We identified canine orthologs to the PAM50 genes using the HGNC Comparison of Orthology Predictions (HCOP) search tool [33]. We identified ER-positive (ER^+) and ER-negative (ER^-) tumors using Estrogen receptor 1 (*ESR1*) gene expression. *ESR1* showed a distinct bimodal distribution in both cohorts (Supplementary Fig. S1), enabling a cut-off to be set across all tumors. Notably, there was high concordance between ER-status determined by *ESR1*-expression and ER-status determined by IHC in TCGA ($\kappa = 0.823$). It is well known that PAM50 subtyping may be affected by different distribution of ER^+ tumors in the training dataset used to calculate the original PAM50 centroids ($\sim 60\% ER^+$) and test datasets (CMGT/TCGA, $\sim 75\% ER^+$) [34, 35]. Hence, we performed gene centering separately for ER^+ and ER^- tumors as described in Lien et al. [32]. We then calculated the Pearson correlation between the gene-centered expression values for the 44 genes and the PAM50 centroids separately for each tumor. The subtype corresponding to the highest correlation coefficient was assigned as the tumor's PAM50 subtype. Subtype assignments for both cohorts are presented in Supplementary File 2. There was high concordance between our subtyping results and the published subtypes for TCGA tumors ($\kappa = 0.81$) [36].

The CMGT RNAseq dataset included expression of 44 of the PAM50 genes (Supplementary File 3). The six genes not

present (*BIRC5*, *CXXC5*, *FOXO1*, *KRT17*, *MIA* and *NAT1*) were therefore removed from the analyses. To investigate the impact of removing six genes from the subtype centroids, we performed subtyping of the TCGA RNAseq data separately with 50 and then with 44 genes. These analyses revealed a high correlation between the subtype correlation coefficients obtained from the two gene lists (Pearson correlations, $r > 0.96$ for all subtypes). We therefore concluded that removing these genes would not notably impact the results of the subtyping. We also compared the results when subtyping the TCGA cohort separately and when merged with the CMGT cohort. Here too, the subtype correlation coefficients were strongly correlated across all subtypes (Pearson correlation, $r > 0.96$ for all subtypes).

Gene Expression Analyses

Proliferation Score

The proliferation score was calculated as the mean of the standardized expression of 10 proliferation-associated genes: *CCNB1*, *CDC20*, *NUF2*, *CEP55*, *NDC80*, *MKI67*, *PTTG1*, *RRM2*, *TYMS*, *UBE2C* [37]. The eleventh gene from the original proliferation signature (*BIRC5*) was not present in the canine dataset.

Immune Score

Immune genes were obtained from Nanostring's Canine IO panel [38]. 650 out of the 800 genes overlapped between the immune gene list and the CMGT expression dataset. The immune score was calculated as the mean of the standardized expressions of the 650 immune genes. Genes representing specific immune cell types were selected from the gene annotation supplied by Nanostring.

Single Sample Gene Set Enrichment Analyses

Single sample gene set enrichment analyses (ssGSEA) were performed on the merged gene expression dataset (CMGT and TCGA) in GenePattern using the ssGSEA module and the Hallmarks of Cancer (HOC) gene sets from the Molecular Signatures Database [39–41].

Independent Validation Data

Two gene expression datasets of canine mammary gland tumors were retrieved from the Gene Expression Omnibus [25] for validation purposes. GSE20718 [42, 43] is gene expression data obtained from the Affymetrix Canine Genome 2.0 array and includes 27 mammary gland tumors; 13 with lymph node metastases and 14 without. 39 of the PAM50 genes were available in this dataset. GSE136197 [9,

[44] consists of 63 benign and malign tumors from 16 dogs and was obtained by RNA sequencing. In this dataset, 43 of the PAM50 genes were available for analysis.

Mutation Analyses

In the CMGT cohort, mutations were annotated using the Ensembl Variant Effect Predictor (VEP) [45] with the CanFam3.1 Genome Assembly [46]. We included in the analyses only non-synonymous mutations of coding regions, and we excluded ten genes characterized as FLAGS (frequently mutated genes) [47]. The tumor mutation burden (TMB) was represented by the number of mutations per tumor. The Cosmic Cancer Gene Census (tier 1 and 2) was downloaded from cancer.sanger.ac.uk [48]. TCGA mutation data was downloaded from the UCSC Xena Platform.

Copy Number Analyses

Segmentation of CMGT copy number data is described in Kim et al. [18]. The cut-off for amplification was set to segment mean > 0.2 and for deletion < -0.2 . A copy number aberration index was calculated separately for each tumor as the proportion of the sequenced genome above or below the cut-off. TCGA copy number data was downloaded from the UCSC Xena Platform.

Statistical Tests

Statistical analyses were conducted in RStudio version 4.0.5 [49, 50]. Fisher's Exact Test was used to test associations between the distribution of subtypes in the canine and human cohorts, to test associations between the canine subtypes and histological grade, histopathological diagnosis, neuter status and lymph node invasion, and to test associations between subtype and mutation of specific genes. Kruskal Wallis tests were performed to test for differences in TMB and copy number aberration index across all subtypes, while Mann Whitney U-tests were performed for pairwise comparisons between subtypes. Mann Whitney U tests were also performed to compare ssGSEA signatures between subtypes and between species. All tests were two-sided. For representation in a heatmap, data were clustered using Euclidean distance as the distance metric and complete linkage as the clustering method. Cohen's kappa coefficient was used for calculating concordance between ER-status based on IHC vs. ESR1 expression in TCGA and for comparing subtype results with subtypes already published for TCGA.

R Packages

The downloaded RNAseq data were annotated with ENSCAF-IDs and refGenome v1.7.7 [51] was used for

gene annotation. VCF files were analyzed using the package vcfR v1.12.0 [52]. Heatmaps were created using Complex Heatmaps v2.7.10 [53] and other plots were created using ggplot2 v3.3.3 [54]. Additional packages used were circlize v0.4.12 [55], factoextra v1.0.7 [56], ggpubr v0.4.0.999 [57], IRanges v2.24.1 [58], vcd v1.4.9 [59] and reshape2 v1.4.4 [60].

Results

PAM50 Subtype Distribution and Clinical Properties of Canine Mammary Gland Tumors

PAM50 subtyping was performed on the merged RNAseq dataset, which included all canine tumor samples (CMGT, $n = 157$) and TCGA human breast tumor samples (TCGA, $n = 1097$). In the canine cohort, all four subtypes were present, and the relative pattern of subtype distribution was comparable to the distribution in the human cohort with minor differences: There was a higher proportion of luminal A and basal-like tumors, and correspondingly fewer luminal B and HER2-enriched tumors in the CMGT compared to the TCGA cohort (Fig. 1a). The proportion of ER⁺ tumors was nearly equal between the species: 75.2% and 76.4% in the canine and human cohorts, respectively. Most luminal A canine tumors were ER⁺, similar to human tumors, while a larger proportion of basal-like canine tumors were ER⁺ compared to basal-like human tumors (Fig. 1a, hatched areas indicate ER⁺).

Next, we investigated the association between CMGTs of different PAM50 subtypes and histopathological diagnosis, histological grade, lymph node invasion and neuter status (Table 1). The histopathology differed significantly between the subtypes (Fisher's Exact Test, $p = 0.002$). Most benign tumors were of the Luminal A subtype, as were most complex carcinomas, while most basal-like tumors were simple carcinomas although many simple carcinomas were Luminal A. When including only the tumors considered malignant by histopathology ($n = 112$), there was still a significant difference between histopathology and subtype (Fisher's Exact Test, $p = 0.004$). Among the malignant tumors there was a significant association between the PAM50 subtype and histological grade (Fisher's Exact Test, $p < 0.001$), with a higher proportion of grade 3 and a lower proportion of grade 1 tumors of the basal-like subtype, while the opposite was true for luminal A tumors. The association between subtype and lymph node invasion was significant (Fisher's Exact Test, $p < 0.001$) with a higher degree of lymph node invasion in basal-like cancers compared to the other subtypes. We found no association between neuter status and subtype (Fisher's Exact Test, $p = 0.89$).

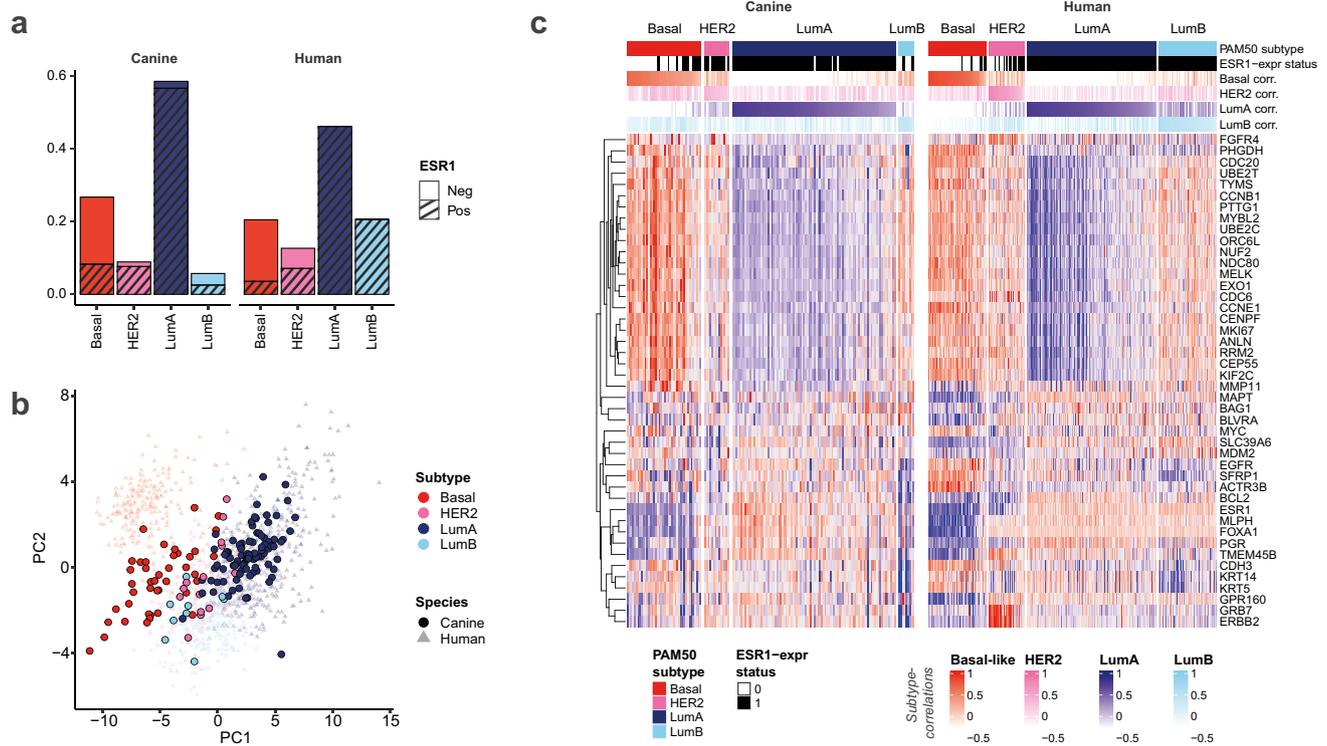


Fig. 1 PAM50 subtypes in canine and human mammary gland tumors **a**: Distribution of PAM50 subtypes in CMGTs ($n=157$) and the TCGA cohort ($n=1097$). Hatched area indicates ER⁺ tumors. **b**: Principal component analyses based on the PAM50 genes. Analysis was performed on the TCGA data with CMGT data overlaid for comparison. First principal component (PC1) on the x-axis; second principal component (PC2) on the y-axis. **c**: Cluster heatmap show-

ing expression of the 44 PAM50 genes present in the CMGT dataset. Tumors are shown in columns (ordered by species, subtype, and subtype correlation) and genes are shown in rows. Clustering of genes was performed using *Euclidean* as distance metric and *complete* as clustering method. Top annotation depicts PAM50 subtype, ER status and correlation coefficients for the four subtype centroids. Data is gene-centered

Subtype Correlation Coefficients Reveal Relevant Tumor Biology Characteristics

PAM50 subtyping is a nearest-centroid method which, in addition to categorical subtype assignment, yields correlations to all centroids that reveal additional insights into tumor biology [61]. Overall, we observed a similar picture in both cohorts; tumors of the basal-like and luminal A subtypes showed a relatively high correlation to their assigned subtype and a relatively low second highest subtype correlation coefficient (Table 2 and Supplementary Fig. S2). The basal-like and luminal A correlation coefficients were negatively correlated in both species, i.e., tumors with high correlation to the basal-like centroid showed correspondingly low correlation to the luminal A centroid. In HER2-enriched and luminal B subtypes, the picture was less distinct, with low correlation to all subtypes. This was particularly evident in the canine dataset. The subtype assignments of HER2-enriched and luminal B canine tumors are therefore unreliable, due to similar correlation coefficients to several centroids. Importantly, even though most basal-like canine

tumors were distinctly basal-like, in general, they showed a lower correlation to the basal-like centroid and were in closer proximity to other centroids compared to the human tumors. The benign tumors in the CMGT cohort did not show lower correlation coefficients to their assigned subtype than the malignant CMGT tumors.

Comparable Gene Expression Landscape in Canine and Human Mammary Gland Tumors

To further compare the gene expression landscape of canine and human mammary gland tumors of different subtypes, we performed principal component analysis based on the 44 overlapping PAM50 genes using the human dataset. Projecting the canine data onto the human principal components (Fig. 1b) revealed a similar picture in the two species, although with a noticeable shift of the canine basal-like tumors towards the luminal tumors. To explore the contribution of the individual genes to the subtyping, we performed a separate PCA analysis on the canine tumors and compared the results from the two species. The first

Table 1 Clinical characteristics of PAM50 subtypes in CMGT. Number of tumors of each subtype (columns) and clinical characteristics (rows) are shown. Total number of tumors and *P*-values obtained

from Fisher's Exact Test for each clinical characteristic are also indicated. Clinical data obtained from Kim et al. [18, 24]

	Basal-like 42 (26.8%)	HER2-enriched 14 (8.9%)	Luminal A 92 (58.6%)	Luminal B 9 (5.7%)	Total 157	<i>P</i>-value
Histopathological diagnosis					157	<i>p</i> =0.002
<i>Benign</i>	6	5	31	3	45	
<i>Mixed tumor</i>	4	1	8	0	13	
<i>Complex carcinoma</i>	6	1	26	0	33	
<i>Simple carcinoma</i>	23	7	25	3	58	
<i>Other histology</i>	3	0	2	3	8	
Histological grade (malignant)					112	<i>p</i> <0.001
<i>1</i>	7	4	48	2	61	
<i>2</i>	9	4	8	2	23	
<i>3</i>	20	1	5	2	28	
Lymph node invasion					157	<i>p</i> <0.001
<i>Absent</i>	30	13	89	9	141	
<i>Present</i>	12	1	3	0	16	
Neuter status					157	<i>p</i> =0.89
<i>Intact</i>	26	10	63	6	105	
<i>Neutered</i>	15	4	28	2	49	
<i>NA</i>	1	0	1	1	3	

principal component represented 49% of the variance in both species and there was high compliance between the species with regard to each gene's contribution to the first principal component (Pearson correlation, $r=0.89$, $p<0.001$) (Supplementary Fig. S3). In both species, the genes with the highest contribution to the first principal component were those involved in cell proliferation, a known distinction between basal-like and luminal human tumors [20].

Cluster analysis of the tumors based on the 44 PAM50 genes revealed similar gene expression patterns in the two cohorts (Fig. 1c). Proliferation-associated genes were distinctly differentially expressed between the subtypes in both species; low expression in luminal A and high expression in basal-like tumors (Fig. 2a). Also, the genes encoding the hormone receptors ER and PR showed a similar profile between the species with generally low expression in basal-like tumors and high expression in

luminal A tumors (Fig. 2b, c). In contrast, *ERBB2* and other genes (such as *GRB7*) typically highly expressed in tumors of the HER2-enriched subtype showed species-specific differences, as the canine HER2-enriched tumors lacked distinct overexpression of *ERBB2* and *GRB7* (Figs. 1c and 2d). A separate analysis of the benign CMGT tumors showed that they displayed gene expression subtype characteristics roughly in line with the malignant CMGT tumors (Supplementary Fig. S4).

To validate these findings, we performed a similar cluster analysis across the PAM50 genes in two separate gene expression datasets obtained from canine mammary gland tumors (Supplementary Fig. S5). In both cases, clustering showed distinct separation between the tumors displaying a basal-like phenotype and the tumors of a luminal-like phenotype, similar to our observations in the main dataset (Fig. 1c). Neither of the datasets contained tumors with high expression levels of *ERBB2*, also in line with the initial results. In one of the datasets (Kloppfleisch et al.) [43], about 50% of the tumors were associated with regional lymph node metastases at the time of resection (Supplementary Fig. S5a) and a majority of these showed a basal-like gene expression profile in line with our main results. The second dataset (Graum et al.) [9] consists of multiple tumors from each patient, including a substantial number of benign tumors, and none of those showed a distinct basal-like phenotype (Supplementary Fig. S5b), also in line with our findings in the main dataset.

Table 2 Subtype correlation coefficients in canine and human tumors for each PAM50 subtype. Interquartile range is shown in parentheses

	Canine	Human
Basal-like	0.45 (0.36–0.57)	0.66 (0.60–0.79)
HER2-enriched	0.25 (0.22–0.31)	0.57 (0.43–0.67)
Luminal A	0.56 (0.38–0.65)	0.55 (0.45–0.70)
Luminal B	0.38 (0.31–0.39)	0.42 (0.33–0.49)

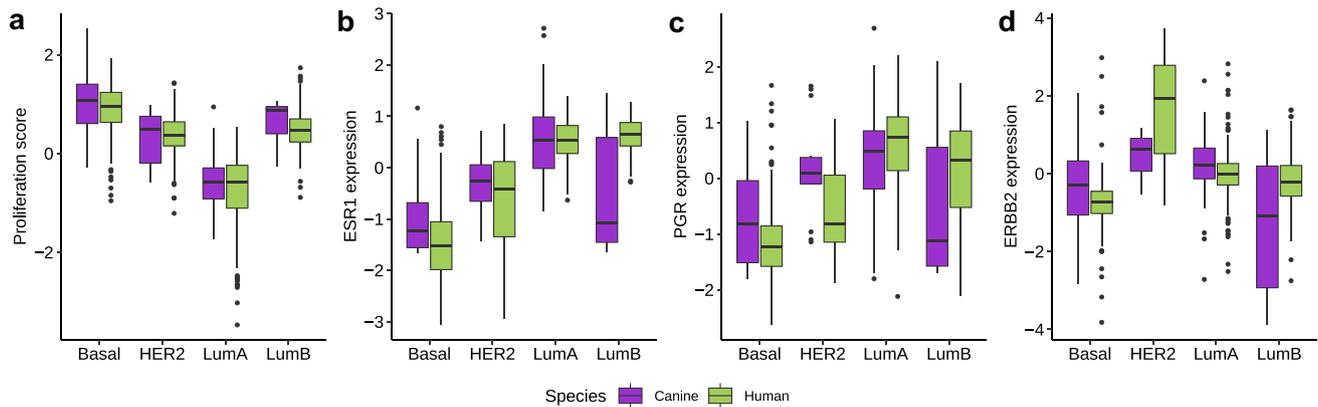


Fig. 2 Gene expression characteristics in CMGTs and TCGA tumors by PAM50 subtype **a**: Proliferation score, **b**: Estrogen receptor 1 (*ESR1*) expression, **c**: Progesterone receptor (*PGR*) expression, **d**:

erb-b2 receptor tyrosine kinase 2 (*ERBB2*) expression (encoding HER2). Boxplots illustrate the median (middle line) and interquartile range (box); whiskers indicate $1.5 \times \text{IQR}$ above and below the box

Variation in Immune Signatures Between Species

To explore in greater depth the biological processes that characterize CMGT subtypes and how they compare to the corresponding human breast tumor subtypes, we performed ssGSEA on the HOC gene sets in the Molecular Signature Database (Supplementary File 4) [40]. Several gene sets showed very similar profiles in canine and human tumors with distinct differences between the subtypes. For instance, gene sets related to cell cycle regulation (e.g., *MYC* targets, G2M checkpoint and E2F targets) and WNT/ β -catenin signaling were significantly more highly expressed in basal-like tumors compared to luminal A tumors in both species (Supplementary Fig. S6a–d, Mann Whitney U tests: $p < 0.001$ for all basal-like vs. luminal A comparisons). Other gene sets, however, showed diverging expression patterns between canine and human tumors, particularly for the basal-like subtype (Supplementary Fig. S6e–h). Interestingly, a majority of these were related to immune responses, such as the TNF α -signaling via NF κ B, Inflammatory response, and Interferon gamma- and Interferon alpha responses, which were all markedly more active in human basal-like tumors compared to canine basal-like tumors (Mann Whitney U tests, $p < 0.001$ for all canine vs. human comparisons).

To elaborate on these findings, we investigated differences in gene expression of 650 immune related genes [38]. An immune score derived from these genes was significantly lower in canine basal-like tumors than human basal-like tumors, while no significant differences were seen between canine and human in tumors of the other subtypes (Supplementary Fig. S7a). Hierarchical clustering of basal-like tumors across the immune genes showed that most canine tumors clustered together with human tumors with low immune score (Supplementary Fig. S7b). There was no correlation between the immune score and correlation

coefficient to the basal-like centroid in basal-like tumors, thus; lower immune score in canine tumors could not be explained by the overall higher correlation to the basal-like centroid in the human tumors (Supplementary Fig. S7c). Genes specific for T-cells (e.g. *CD3D*) and macrophages (*CD68*) were lower expressed in canine compared to human basal-like tumors, while genes highly expressed in mast cells (e.g. *CPA3* and *MS4A2*) were higher expressed in canine basal-like tumors compared to the human counterpart. Moreover, genes expressed in regulatory T cells (*FOXP3*) and cytotoxic cells (*GZMB*) were higher expressed in human basal-like tumors, likewise were genes encoding the immune checkpoints proteins PD-1 (*PDCD1*) and PD-L1 (*CD274*) (Supplementary Fig. S7d).

Immune scores were also calculated for tumors in the two validation datasets and results were similar to our findings in the main dataset. In the data from Klopffleisch et al., there was a strong positive association between high immune score and lymph node metastasis, and the metastatic tumors were more frequent among those with a basal-like profile (Supplementary Fig. S5a). In the data from Graim et al., samples with high immune score were evenly distributed across the basal-like and luminal-like cluster. (Supplementary Fig. 5b).

Similar Association Between Subtype and PIK3CA and TP53 Mutations

In the canine tumor cohort, we identified 2318 non-synonymous mutations in the coding regions of 1873 genes. Twenty-three genes carried mutations in four or more tumors (Fig. 3a). Of these, 11 genes were previously found to be implicated in cancer and present in the Cosmic Cancer Gene Census [48]. *PIK3CA* and *TP53* were the most frequently and second most frequently mutated genes, respectively, in the canine cohort, which was

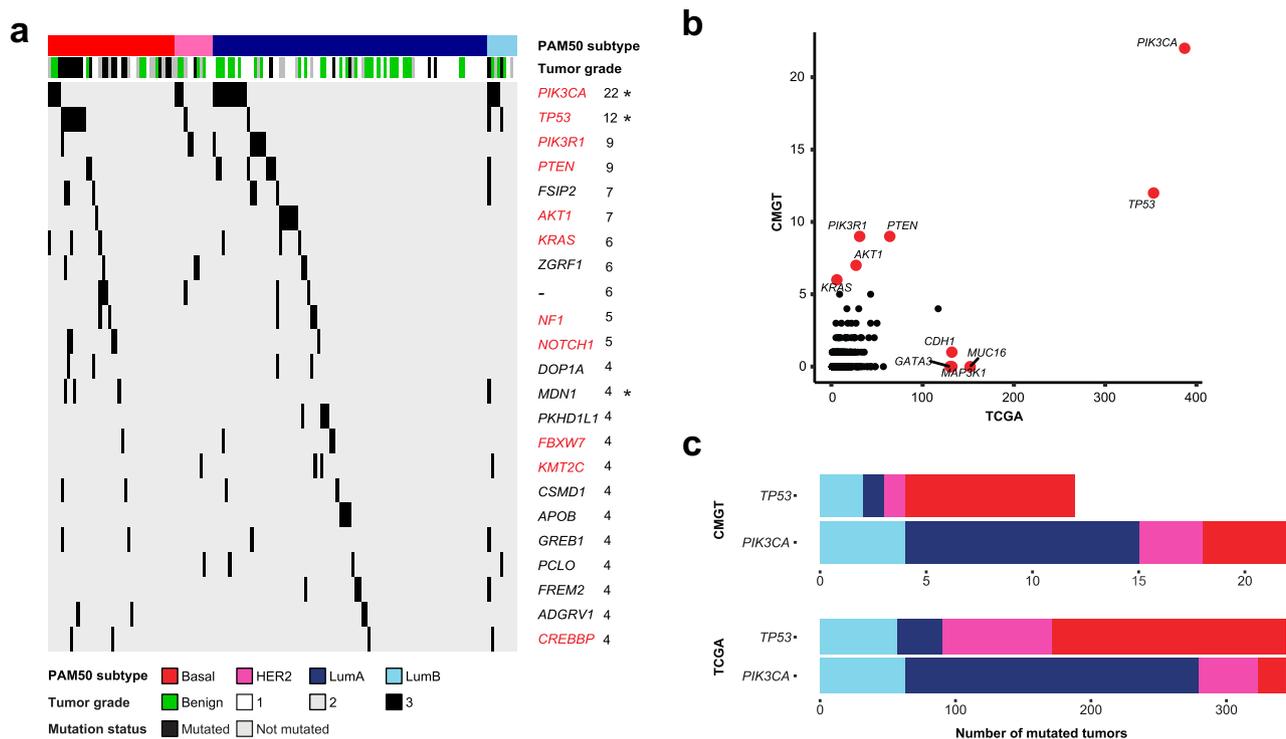


Fig. 3 Subtype specific mutation profiles of CMGTs and TCGA tumors **a**: Frequency and subtype association of all genes with four or more mutations across the CMGT dataset. Tumors are ordered by subtype in columns and genes are shown in rows. Top annotation depicts PAM50 subtype and tumor grade. The number of tumors carrying mutations is noted to the right of the gene symbol. Genes present in the Cosmic Cancer Gene Census are shown in red and genes with significantly different mutation frequencies between the

also the case in the TCGA cohort (Fig. 3b). However, other than these two genes, there were noticeable differences between the species. Some genes with relatively high mutation frequencies in the TCGA cohort (e.g., *GATA3*, *MUC16* and *MAP3K1*) were not mutated in the canine tumors, while genes such as *PIK3R1*, *PTEN*, *AKT1* and *KRAS* were more frequently mutated in canine tumors compared to human (Fig. 3b).

Among the canine tumors, there was a significantly different distribution of tumors with *PIK3CA* and *TP53* mutations between the subtypes (Fisher's Exact Test, *p*-values: *PIK3CA* = 0.04, *TP53* < 0.001) with more *PIK3CA* mutations in luminal tumors and more *TP53* mutations in basal-like tumors (Fig. 3c). The same pattern is found in human tumors (Fisher's Exact Test, *p*-values: *PIK3CA* < 0.001, *TP53* < 0.001) [28]. Interestingly, 27% of the benign tumors carried mutations in *PIK3CA*, while no benign tumors were *TP53* mutated (Fig. 3a).

We found that TMB was higher in human breast tumors compared to canine, in line with previous literature [17, 18], and this was true for all subtypes. In human breast cancer, the TMB varies between the PAM50 subtypes [62]. Among

CMGT subtypes are marked with an asterisk. One gene was uncharacterized with Ensembl-ID ENSCAFG00000018773 and is marked "-". **b**: Mutation frequency in TCGA (x-axis) and CMGT (y-axis) of genes present in the Cosmic Cancer Gene Census. Genes commonly mutated in either TCGA, CMGT or both are highlighted (red). **c**: Subtype distribution of tumors with *TP53* and *PIK3CA* mutations in CMGT and TCGA

the canine tumors, although there was no significant global difference in the mutation burden between the subtypes (Kruskal–Wallis Test, *p* = 0.11), we did identify a significant difference in the TMB between basal-like and luminal A subtypes (Mann Whitney U Test, *p* = 0.015) (Supplementary Fig. S8a). In the human cohort, the global difference in mutation burden was significant between the subtypes (Kruskal–Wallis Test, *p* < 0.001) and the most pronounced difference was seen between basal-like and luminal A tumors (Mann Whitney U Test, *p* < 0.001) (Supplementary Fig. S8b).

Comparable Subtype Specific Copy Number Profiles in Canine and Human

We analyzed segmented copy number profiles from the canine tumors and found distinct differences between the subtypes (Supplementary Fig. S9). The aberration index quantifying the proportion of the genome carrying aberrations was significantly different between the subtypes (Kruskal Wallis test, *p* < 0.001) (Supplementary Fig. S10).

Basal-like tumors were, in general, highly aberrant with changes across the whole genome, while the luminal A tumors carried distinctly fewer changes. The paucity of aberrations in luminal A tumors was not the result of a large proportion of benign tumors, as these had copy number aberrations comparable to malignant luminal A tumors (data not shown). Similar to human breast cancer [63], the canine basal-like and luminal B subtypes carried markedly more aberrations than the luminal A tumors. However, for the HER2-enriched subtype there were noticeably fewer copy number aberrations compared to that seen in human breast cancer. HER2-enriched tumors are characterized by distinct copy number patterns, often including high-level amplification of the *ERBB2* locus on chromosome 17q1.2 in humans [64]. In dogs, the *ERBB2* gene is located on chromosome 9. There were no tumors with high-level *ERBB2* amplification and no overrepresentation of *ERBB2* amplifications in canine HER2-enriched tumors compared to the other subtypes (Fig. 4).

We also investigated other driver genes that are commonly affected by copy number changes in human breast cancer (Fig. 4). *MYC* amplification, for instance, in humans is most commonly found in tumors of the basal-like subtype [65], and has previously been reported in canine mammary gland tumors with a simple carcinoma histology [66, 67]. In the CMGT-dataset we also found amplifications of *MYC* predominantly in tumors of the basal-like subtype. In addition, *CCND1* is commonly amplified in human luminal tumors [68]. We found *CCND1* amplified in some cases, but there was no subtype-specific pattern. Among the known breast cancer-related tumor suppressor genes, we found *PTEN* deletions in CMGTs of all subtypes, *TP53* deletions in basal-like and luminal B tumors and *TP53* amplifications in a few tumors of the HER2-enriched and luminal A subtypes, while

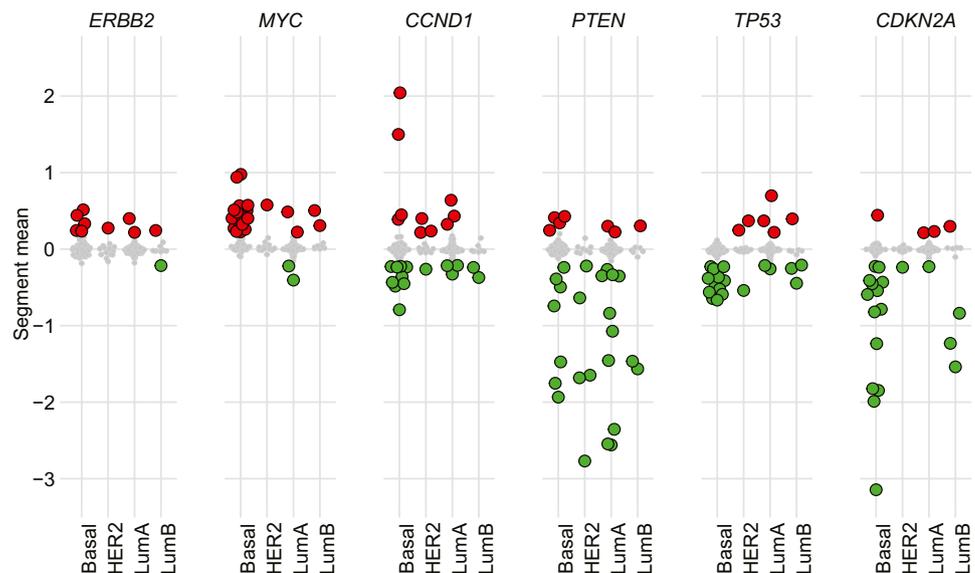
deletions of *CDKN2A* (p16) was mostly seen in basal-like CMGT, similarly to human breast cancer [28].

Discussion

The dog is presumed to be a valuable comparative model for human breast cancer, but there is a need for increased knowledge about the molecular similarities as well as differences between mammary gland tumors in the two species. The PAM50 subtypes are defined for human breast carcinomas; however, in a comparative setting when exploring dogs as models, it is important to determine the relevance of these subtypes also in canine mammary gland tumors using the same method in both species. Here, we have studied the incidence and characteristics of the PAM50 subtypes in a cohort of canine mammary gland tumors. Overall, the PAM50 subtypes captured molecular subtype characteristics in dogs that are well known from human breast cancer. As in human breast cancer, there was a distinct dichotomy between basal-like and luminal canine tumors that could not solely be explained by hormone receptor status.

Nearly 60% of the CMGTs were of the luminal A subtype. These showed high correlation to the luminal A subtype centroid at a level comparable to human luminal A tumors and relatively low correlation to all other subtype centroids. Canine luminal A tumors were, in general, similar to human luminal A tumors with high expression of luminal epithelial genes such as *ESR1*, *PGR*, *FOXA1* and *MLPH* and low expression of genes involved in cell proliferation. *PIK3CA* mutations were more common in canine luminal A tumors than in other subtypes as is also the case in human tumors. The luminal A tumors carried markedly fewer mutations than basal-like tumors. This, combined with fewer copy

Fig. 4 Copy number profiles of cancer-relevant genes in CMGT; Estimated copy number across PAM50 subtypes is illustrated with subtypes on the x-axis and segment mean (representing the copy number) on the y-axis. Tumors with amplifications (segment mean > 0.2) are indicated in red; tumors with deletions (segment mean < -0.2) are indicated in green



number aberrations, indicates a more stable genome in canine luminal A tumors compared to basal-like tumors. The CMGT cohort includes both benign and malignant tumors and most benign tumors were of the luminal A subtype. The benign tumors displayed characteristics in line with their PAM50 subtype, both with regards to proliferation, ESR1 expression, mutations and copy number aberrations. This result is in line with the findings of Sørenmo et al. who proposed benign tumors as precursors of malignant tumors in canines [69].

Tumors of the basal-like subtype constituted 27% of the canine cohort and showed gene expression characteristics well known from analyses of human basal-like tumors. They expressed low levels of luminal epithelial genes and high levels of multiple proliferation-associated genes indicating that these are highly proliferative tumors. The basal-like CMGTs were more often high-grade tumors and of the simple carcinoma histology type compared to luminal A tumors, similar to human breast cancer [70]. However, in contrast to human breast cancer, where basal-like tumors paradoxically are associated with less lymph node invasion than luminal tumors [71], there was a higher number of lymph node-positive cases among canine basal-like tumors compared to the other subtypes. This finding was also evident in one of the independent validation datasets that consisted of approx. 50% locally metastasized tumors. Hence, together with noticeably lower expression of gene signatures related to immune cells in canine basal-like tumors compared to human, this indicates intrinsic differences between canine and human basal-like tumors that might be explained by different immune responses. Exploring species-specific differences in the immune response towards tumors could generate valuable knowledge, especially in view of the emergence of immune therapy in breast cancer treatment [72]. In addition, the basal-like canine tumors had markedly more copy number aberrations than the luminal A tumors in particular, which fits well with what is known about human breast cancer [28]. Canine basal-like tumors also encompassed slightly higher mutation frequencies compared to other subtypes and carried several subtype specific aberrations such as mutations and deletions of *TP53*, amplifications of *MYC* and deletions of *CDKN2A*, all known features of bona-fide basal-like human tumors [28].

In our study, 8.9% of the CMGTs were characterized as HER2-enriched, however, correlation to the HER2-centroid was low for these tumors. *ERBB2*-amplifications in canine tumors were only low level, and were not restricted to tumors of the HER2-enriched subtype. This indicates that the HER2-enriched subtype does not have the same prevalence and relevance in CMGTs, confirming results from several previous studies [18, 66, 73]. Similar to the HER2-enriched subtype, luminal B did not emerge as a definite subtype in the canine cohort. Luminal B tumors were characterized by

high expression of genes involved in cell proliferation, low expression of basal keratins and a high copy number aberration count, but displayed large variation in the expression of luminal epithelial genes, *ERBB2*, as well as in TMB.

The canine samples in this study include mammary gland tumors from predominantly small companion dogs. To obtain a more comprehensive and nuanced overview of the potential of dogs as models for human breast cancer, these findings should be validated in an RNA sequencing dataset including tumors from larger dogs and working breeds; however, such datasets of sufficient size do not yet exist. Nevertheless, we were able to validate our main findings in two smaller independent datasets. For this study, we merged the canine and human cohorts and performed subtyping on the merged datasets. Additionally, subtyping was performed using 44 out of the original 50 genes. These factors, however, do not seem to have affected the subtyping results significantly. Different chromosomal composition in dogs and humans complicates comparative copy number analyses. Nevertheless, our study shows that such analyses need to be performed stratified by subtype.

Conclusions

CMGTs are highly heterogeneous biologically, and represent an unmet potential for modeling human breast cancer. Our study identifies many similarities between mammary gland tumors of dogs and humans, but also discovers important differences and emphasizes that the molecular subtypes should be taken into account when considering dogs as models for breast cancer. Generally, we found a high degree of similarity between canine and human tumors across the four main intrinsic subtypes, but we also pinpointed differences that are important when considering dogs as comparative oncology models. Knowledge obtained from canine/human comparative studies may contribute towards facilitating individualized treatment in dogs suffering from mammary gland tumors and such studies could therefore be of relevance and interest to both the veterinary and human medical communities.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s10911-022-09523-9>.

Acknowledgements We would like to thank Sangwoo Kim and Tae-Min Kim at the Yonsei University College of Medicine, South Korea for making these excellent data publicly available and for providing segmented copy number data.

Authors Contributions All authors contributed to the study conception and design. Helga Bergholtz performed data analysis. Tonje Lien provided feedback on data analysis and contributed statistical guidance. The first draft of the manuscript was written by Helga Bergholtz and all authors commented on the manuscript. All authors read and approved the final manuscript.

Funding Open access funding provided by University of Oslo (incl Oslo University Hospital). Helga Bergholtz is supported by the Norwegian Cancer Society (grant number 214924) and Forskningsfondet kreft hos hund.

Data Availability Data used in this study are publicly available. The gene expression data can be accessed from the NCBI GEO database under accession number GSE119810 [26]. Clinical information and variant calling data can be downloaded from <https://doi.org/10.6084/m9.figshare.c.4543784.v1>. Raw data are available from the Sequence Read Archive IDs (SRP159466 and SRP159481) [74, 75].

Declarations

Conflict of Interests The authors have no relevant financial or non-financial interests to disclose. TS is member of the editorial board of the Journal of Mammary Gland Biology and Neoplasia.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Manning HC, Buck JR, Cook RS. Mouse models of breast cancer: Platforms for discovering precision imaging diagnostics and future cancer medicine. *J Nucl Med*. 2016;57:60S–68S. <https://doi.org/10.2967/jnumed.115.157917>.
- Özdemir BC, Sflomos G, Brisken C. The challenges of modeling hormone receptor-positive breast cancer in mice. *Endocr Relat Cancer*. 2018;25:R319–30. <https://doi.org/10.1530/ERC-18-0063>.
- Khanna C, Lindblad-Toh K, Vail D, et al. The dog as a cancer model. *Nat Biotechnol*. 2006;24:1065–6.
- Cadiou E, Ostrander EA. Canine genetics offers new mechanisms for the study of human cancer. *Cancer Epidemiol Biomark Prev*. 2007;16:2181–3.
- Lindblad-Toh K. What animals can teach us about evolution, the human genome, and human disease. *Upsala J Med Sci*. 2020;125:1–9.
- Ostrander EA, Dreger DL, Evans JM. Canine cancer genomics: lessons for canine and human health. *Annu Rev Anim Biosci*. 2019;7:449–72.
- Abdelmegeed SM, Mohammed S. Canine mammary tumors as a model for human disease (Review). *Oncol Lett*. 2018;15:8195–205.
- Gray M, Meehan J, Martínez-Pérez C, et al. Naturally-occurring canine mammary tumors as a translational model for human breast cancer. *Front Oncol*. 2020;10:1–17. <https://doi.org/10.3389/fonc.2020.00617>.
- Graim K, Gorenshyeyn D, Robinson DG, et al. Modeling molecular development of breast cancer in canine mammary tumors. *Genome Res*. 2021;31:337–47. <https://doi.org/10.1101/GR.256388.119>.
- Moe L. Population-based incidence of mammary tumours in some dog breeds - PubMed. *J Reprod Fertil Suppl*. 2001;57:439–43.
- Dobson JM, Samuel S, Milstein H, et al. Canine neoplasia in the UK: Estimates of incidence rates from a population of insured dogs. *J Small Anim Pract*. 2002;43:240–6. <https://doi.org/10.1111/j.1748-5827.2002.tb00066.x>.
- Siegel RL, Miller KD, Jemal A. Cancer statistics, 2020. *CA Cancer J Clin*. 2020;70:7–30. <https://doi.org/10.3322/caac.21590>.
- Philibert JC, Snyder PW, Glickman N, et al. Influence of host factors on survival in dogs with malignant mammary gland tumors. *J Vet Intern Med*. 2003;17:102–6. <https://doi.org/10.1111/j.1939-1676.2003.tb01330.x>.
- Queiroga FL, Raposo T, Carvalho MI, et al. Canine mammary tumours as a model to study human breast cancer: Most recent findings. *In Vivo*. 2011;25:455–65.
- Bleiweiss IJ (2019) Pathology of breast cancer - UpToDate. <https://www.uptodate.com/contents/pathology-of-breast-cancer>. Accessed 11 Mar 2019.
- Goldschmidt MH, Peña L, Rasotto R, Zappulli V. Classification and grading of canine mammary tumors. *Vet Pathol*. 2011;48:117–31. <https://doi.org/10.1177/0300985810393258>.
- Alsaihati BA, Ho KL, Watson J, et al. Canine tumor mutational burden is correlated with TP53 mutation across tumor types and breeds. *Nat Commun*. 2021;12:1–13. <https://doi.org/10.1038/s41467-021-24836-9>.
- Kim TM, Yang IS, Seung BJ, et al. Cross-species oncogenic signatures of breast cancer in canine mammary tumors. *Nat Commun*. 2020. <https://doi.org/10.1038/s41467-020-17458-0>.
- Sørli T, Perou CM, Tibshirani R, et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci U S A*. 2001;98:10869–74. <https://doi.org/10.1073/pnas.191367098>.
- Perou CM, Sørli T, Eisen MB, et al. Molecular portraits of human breast tumours. *Nature*. 2000;406:747–52. <https://doi.org/10.1038/35021093>.
- Ohnstad HO, Borgen E, Falk RS, et al. Prognostic value of PAM50 and risk of recurrence score in patients with early-stage breast cancer with long-term follow-up. *Breast Cancer Res*. 2017. <https://doi.org/10.1186/s13058-017-0911-9>.
- Wallden B, Storhoff J, Nielsen T, et al. Development and verification of the PAM50-based Prosigna breast cancer gene signature assay. *BMC Med Genomics*. 2015;8:54. <https://doi.org/10.1186/s12920-015-0129-6>.
- Parker JS, Mullins M, Cheang MCU, et al. Supervised risk predictor of breast cancer based on intrinsic subtypes. *J Clin Oncol*. 2009;27:1160–7. <https://doi.org/10.1200/JCO.2008.18.1370>.
- Kim KK, Seung BJ, Kim D, et al. Whole-exome and whole-transcriptome sequencing of canine mammary gland tumors. *Sci Data*. 2019;6:147. <https://doi.org/10.1038/s41597-019-0149-8>.
- Edgar R, Domrachev M, Lash AE. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res*. 2002;30:207–10. <https://doi.org/10.1093/NAR/30.1.207>.
- Gene Expression Omnibus Series GSE119810. <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE119810>. Accessed 19 Apr 2020.
- Whole-exome and whole-transcriptome sequencing of canine mammary gland tumors. https://figshare.com/collections/Whole-exome_and_whole-transcriptome_sequencing_of_canine_mammary_gland_tumors/4543784/1. Accessed 19 Apr 2021.
- Network TCGA, Network TCGA, Cancer T, Atlas G. Comprehensive molecular portraits of human breast tumours. *Nature*. 2012;490:61–70. <https://doi.org/10.1038/nature11412>.
- Hoadley KA, Yau C, Hinoue T, et al. Cell-of-origin patterns dominate the molecular classification of 10,000 tumors from 33 types

- of cancer. *Cell*. 2018;173:291–304.e6. <https://doi.org/10.1016/j.cell.2018.03.022>.
30. Goldman MJ, Craft B, Hastie M, et al. Visualizing and interpreting cancer genomics data via the Xena platform. *Nat Biotechnol*. 2020;38:675–8.
 31. Wirapati P, Sotiriou C, Kunkel S, et al. Meta-analysis of gene expression profiles in breast cancer: toward a unified understanding of breast cancer subtyping and prognosis signatures. *Breast Cancer Res BCR*. 2008;10:R65. <https://doi.org/10.1186/BCR2124>.
 32. Lien TG, Ohnstad HO, Lingjaerde OC, et al (2021) Sample preparation approach influences pam50 risk of recurrence score in early breast cancer. <https://doi.org/10.3390/cancers13236118>.
 33. Wright MW, Eyre TA, Lush MJ, et al. HCOP: The HGNC comparison of orthology predictions search tool. *Mamm Genome*. 2005;16:827–8. <https://doi.org/10.1007/s00335-005-0103-2>.
 34. Kensler KH, Sankar VN, Wang J, et al. PAM50 molecular intrinsic subtypes in the nurses' health study cohorts. *Cancer Epidemiol Biomarkers Prev*. 2019;28:798–806. <https://doi.org/10.1158/1055-9965.EPI-18-0863>.
 35. Sørli T, Borgan E, Myhre S, et al. The importance of gene-centring microarray data. *Lancet Oncol*. 2010;11:719–20. [https://doi.org/10.1016/S1470-2045\(10\)70174-1](https://doi.org/10.1016/S1470-2045(10)70174-1).
 36. Heng YJ, Lester SC, Tse GMK, et al. The molecular basis of breast cancer pathological phenotypes. *J Pathol*. 2017;241:375–91. <https://doi.org/10.1002/PATH.4847>.
 37. Nielsen TO, Parker JS, Leung S, et al. A comparison of PAM50 intrinsic subtyping with immunohistochemistry and clinical prognostic factors in tamoxifen-treated estrogen receptor-positive breast cancer. *Clin Cancer Res*. 2010;16:5222–32. <https://doi.org/10.1158/1078-0432.CCR-10-1282>.
 38. Mason N, Bailey C, Piazza E, et al. Abstract 1693: Multinational, multi-center collaboration to develop a novel gene expression tool for comparative translational immuno-oncology. *Can Res*. 2021;81:1693–1693. <https://doi.org/10.1158/1538-7445.AM2021-1693>.
 39. Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci*. 2005;102:15545–50. <https://doi.org/10.1073/pnas.0506580102>.
 40. Liberzon A, Birger C, Thorvaldsdóttir H, et al. The molecular signatures database hallmark gene set collection. *Cell Syst*. 2015;1:417–25. <https://doi.org/10.1016/J.CELS.2015.12.004>.
 41. Reich M, Liefeld T, Gould J, et al. GenePattern 2.0. *Nat Genet*. 2006;38:500–1. <https://doi.org/10.1038/ng0506-500>.
 42. Gene Expression Omnibus Series GSE20718. <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE20718>. Accessed 9 Jun 2022.
 43. Klopffleisch 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. <https://doi.org/10.1186/1471-2407-10-618>.
 44. Gene Expression Omnibus Series GSE136197. <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE136197>. Accessed 9 Jun 2022.
 45. Hunt SE, McLaren W, Gil L, et al. Ensembl variation resources. *Database*. 2018;2018:1–12. <https://doi.org/10.1093/database/bay119>.
 46. CanFam3.1 - canFam3 - Genome - Assembly - NCBI. https://www.ncbi.nlm.nih.gov/assembly/GCF_000002285.3/. Accessed 19 Apr 2021.
 47. Shyr C, Taraïlo-Graovac M, Gottlieb M, et al. FLAGS, frequently mutated genes in public exomes. *BMC Med Genomics*. 2014;7:64. <https://doi.org/10.1186/s12920-014-0064-y>.
 48. Sondka Z, Bamford S, Cole CG, et al. The COSMIC Cancer Gene Census: describing genetic dysfunction across all human cancers. *Nat Rev Cancer*. 2018;18:696–705.
 49. RStudio Team. RStudio: Integrated Development for R. Boston, MA: RStudio Inc; 2021.
 50. R Core Team, Foundation R, for Statistical Computing. R: A language and environment for statistical computing. Vienna: Austria; 2021.
 51. refGenome: Gene and Splice Site Annotation Using Annotation Data from “Ensembl” and “UCSC” Genome Browsers version 1.7.7 from CRAN. <https://rdrr.io/cran/refGenome/>. Accessed 20 Apr 2021.
 52. Knaus BJ, Grünwald NJ. vcfR: a package to manipulate and visualize variant call format data in R. *Mol Ecol Resour*. 2017;17:44–53. <https://doi.org/10.1111/1755-0998.12549>.
 53. Gu Z, Eils R, Schlesner M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics*. 2016;32:2847–9. <https://doi.org/10.1093/bioinformatics/btw313>.
 54. Wickham H (2016) Ggplot2: elegant graphics for data analysis.
 55. Gu Z, Gu L, Eils R, et al. Circlize implements and enhances circular visualization in R. *Bioinformatics*. 2014;30:2811–2. <https://doi.org/10.1093/bioinformatics/btu393>.
 56. Kassambara A, Mundt F (2020) Extract and Visualize the Results of Multivariate Data Analyses [R package factoextra version 1.0.7].
 57. ggplot2 Based Publication Ready Plots • ggpubr. <https://rpkgs.datanovia.com/ggpubr/>. Accessed 19 Apr 2021.
 58. Lawrence M, Huber W, Pagès H, et al. Software for computing and annotating genomic ranges. *PLoS Comput Biol*. 2013;9:e1003118. <https://doi.org/10.1371/journal.pcbi.1003118>.
 59. Meyer D, Zeileis A, Hornik K (2021) vcd: Visualizing Categorical Data.
 60. Wickham H. Reshaping data with the reshape package. *J Stat Softw*. 2007;21:1–20. <https://doi.org/10.18637/jss.v021.i12>.
 61. Russnes HG, Lingjærde OC, Børresen-Dale AL, Caldas C. Breast cancer molecular stratification: from intrinsic subtypes to integrative clusters. *Am J Pathol*. 2017;187:2152–62.
 62. Budczies J, Bockmayr M, Denkert C, et al. Classical pathology and mutational load of breast cancer – integration of two worlds. *J Pathol Clin Res*. 2015;1:225–38. <https://doi.org/10.1002/CJP2.25>.
 63. Creighton CJ. The molecular profile of luminal B breast cancer. *Biol Targets Ther*. 2012;6:289. <https://doi.org/10.2147/BTT.S29923>.
 64. Ferrari A, Vincent-Salomon A, Pivot X, et al. A whole-genome sequence and transcriptome perspective on HER2-positive breast cancers. *Nat Commun*. 2016;7:1–9. <https://doi.org/10.1038/ncomms12222>.
 65. Xu J, Chen Y, Olopade OI. MYC and breast cancer. *Genes Cancer*. 2010;1:629. <https://doi.org/10.1177/1947601910378691>.
 66. Liu D, Xiong H, Ellis AE, et al. Molecular homology and difference between spontaneous canine mammary cancer and human breast cancer. *Can Res*. 2014;74:5045–56. <https://doi.org/10.1158/0008-5472.CAN-14-0392>.
 67. Borge KS, Nord S, van Loo P, et al. Canine mammary tumours are affected by frequent copy number aberrations, including amplification of myc and loss of pten. *PLoS ONE*. 2015. <https://doi.org/10.1371/journal.pone.0126371>.
 68. Ignatiadis M, Sotiriou C. Luminal breast cancer: from biology to treatment. *Nat Rev Clin Oncol*. 2013;10(9):494–506. <https://doi.org/10.1038/nrclinonc.2013.124>.
 69. Sorenmo KU, Kristiansen VM, Cofone MA, et al. Canine mammary gland tumours; a histological continuum from benign to malignant; clinical and histopathological evidence. *Vet Comp*

- Oncol. 2009;7:162–172. <https://doi.org/10.1111/j.1476-5829.2009.00184.x>.
70. Livasy CA, Karaca G, Nanda R, et al. Phenotypic evaluation of the basal-like subtype of invasive breast carcinoma. *Mod Pathol.* 2005;19:264–71. <https://doi.org/10.1038/modpathol.3800528>.
 71. Crabb SJ, Cheang MCU, Leung S, et al. Basal Breast Cancer Molecular Subtype Predicts for Lower Incidence of Axillary Lymph Node Metastases in Primary Breast Cancer. *Clin Breast Cancer.* 2008;8:249–56. <https://doi.org/10.3816/CBC.2008.N.028>.
 72. Dagher E, Simbault L, Abadie J, et al (2020) Identification of an immune-suppressed subtype of feline triple-negative basal-like invasive mammary carcinomas, spontaneous models of breast cancer. 42. <https://doi.org/10.1177/1010428319901052>.
 73. Abadie J, Nguyen F, Loussouarn D, et al. Canine invasive mammary carcinomas as models of human breast cancer. Part 2: immunophenotypes and prognostic significance. *Breast Cancer Res Treat.* 2018;167:459–68. <https://doi.org/10.1007/s10549-017-4542-8>.
 74. NCBI Sequence Read Archive SRP159466. <https://trace.ncbi.nlm.nih.gov/Traces/sra/?study=SRP159466>. Accessed 6 Oct 2021.
 75. NCBI Sequence Read Archive SRP159481. <https://trace.ncbi.nlm.nih.gov/Traces/sra/?study=SRP159481>. Accessed 6 Oct 2021.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.