Published in final edited form as:
Cancer Res. 2014 September 15; 74(18): 5045–5056. doi:10.1158/0008-5472.CAN-14-0392.

Molecular homology and difference between spontaneous canine mammary cancer and human breast cancer

Deli Liu[1], **Huan Xiong**[1], **Angela E. Ellis**[2], **Nicole C. Northrup**[2], **Carlos O. Rodriguez Jr**[3], **Ruth M. O'Regan**[4], **Stephen Dalton**[1], and **Shaying Zhao**[1,*]

[1]Department of Biochemistry and Molecular Biology, Institute of Bioinformatics, University of Georgia, Athens, GA 30622

[2]College of Veterinary Medicine, University of Georgia, Athens, GA 30622

[3]School of Veterinary Medicine, University of California, Davis, CA 95616

[4]The Winship Cancer Center, Emory School of Medicine, Atlanta, GA 30322

Abstract

Spontaneously occurring canine mammary cancer (MC) represents an excellent model of human breast cancer but is greatly understudied. To better utilize this valuable resource, we performed whole genome sequencing, whole exome sequencing, RNA-seq and/or high density arrays on 12 canine MC cases, including 7 simple carcinomas and four complex carcinomas. Canine simple carcinomas, which histologically match human breast carcinomas, harbor extensive genomic aberrations, many of which faithfully recapitulate key features of human breast cancer. Canine complex carcinomas, which are characterized by proliferation of both luminal and myoepithelial cells and are rare in human breast cancer, appear to lack genomic abnormalities. Instead, these tumors have about 35 chromatin-modification genes downregulated, and are abnormally enriched with active histone modification H4-acetylation while aberrantly depleted with repressive histone modification H3K9me3. Our findings indicate the likelihood that canine simple carcinomas arise from genomic aberrations whereas complex carcinomas originate from epigenomic alterations, reinforcing their unique value. Canine complex carcinomas offer an ideal system to study myoepithelial cells, the second major cell lineage of the mammary gland. Canine simple

[*]**Corresponding author:** Shaying Zhao Department of Biochemistry and Molecular Biology, Institute of Bioinformatics University of Georgia B304B Life Sciences Building, 120 Green Street, Athens, GA 30602-7229 szhao@bmb.uga.edu, 706-542-9147 Tel, 706-542-1738 Fax.

Conflicts of interest: The authors declare no conflicts of interest.

Author Contributions

Conception and design: S. Zhao, S. Dalton, D. Liu
Development of methodology: D. Liu, X. Huan, S. Zhao
Acquisition of data: X. Huan, A.E. Ellis, C.O. Rodriguez Jr, S. Zhao
Analysis and interpretation of data: D. Liu, S. Zhao, X. Huan, A.E. Ellis
Writing, review and/or revision of the manuscript: S. Zhao, D. Liu, A.E. Ellis
Administrative, technical, or material support: A.E. Ellis, N.C. Northrup
Study supervision: S. Dalton, R.M. O'Regan, N.C. Northrup

Data Access

Sequence data have been submitted to the NCBI SRA database with accession numbers SRP023115, SRP023472 and SRP024250. aCGH data have been submitted to the GEO database with accession number GSE54535.

carcinomas, which faithfully represent human breast carcinomas at the molecular level, provide indispensable models for basic and translational breast cancer research.

Keywords

Spontaneous mammary cancer; dog-human comparison; simple versus complex carcinomas; genomic versus epigenomic abnormality; luminal versus myoepithelial cells

Introduction

Spontaneous cancers in pet dogs represent one of the best cancer models (1-8). First, these cancers are naturally-occurring and heterogeneous, capturing the essence of human cancer, unlike genetically-modified or xenograft rodent models. Second, as companion animals, dogs share the same environment as humans and are exposed to many of the same carcinogens. Indeed, environmental toxins, advancing age and obesity are also risk factors for canine cancer (1). Third, dogs better resemble humans in biology, e.g., similar telomere and telomerase activities (9) and frequent spontaneous epithelial cancers (1), unlike mice (10). Fourth, numerous anatomic and clinical similarities are noted for the same types/subtypes of cancer between the two species, and similar treatment schemes are used (2-4). Furthermore, the large population of pet dogs (~70 million estimated in the US) provides a valuable resource facilitating basic and clinical research. Importantly, the dog genome has been sequenced to a >7.6X coverage (11), yielding a genome assembly nearly as accurate as the mouse or rat genome (11, 12), unlike another companion animal, the cat. This makes many genomic analyses possible with the dog but not with the cat.

As in women, mammary cancer (MC) is among the most frequent cancers in female dogs. The annual incidence rate is estimated at 198/100,000 (1), which is comparable to the rate of 125/100,000 for breast cancer in women in the US (13). MC is especially common in dogs that are not spayed or are spayed after the second estrus, with the risk for malignant tumor development expected at 26% (1). However, unlike human breast cancer, canine MC is poorly characterized at the genome-wide level. For example, only five canine MC cases have recently undergone ~2X whole genome sequencing (WGS) (14), and a limited number have been analyzed with gene expression microarray (15-17). This drastically differs from their human counterparts, where thousands of breast cancer genomes and transcriptomes are characterized, with several studies cited here (18-23).

Like other cancer types, many anatomic and clinical similarities are documented between canine MC and human breast cancer (24). Various molecular homologies are also reported, e.g., WNT signaling alteration (15, 17). Meanwhile, canine MC also differs from human breast cancer in certain aspects. For example, dogs have only one or two estrous cycles a year followed by a prolonged luteal phase, which is 63 days for the dog compared to 14 days for the human. During this unusually long luteal phase, the mammary gland is continuously exposed to high levels of progesterone (24). Another variation that may be related to this hormonal difference is described below.

Mammary gland epithelium consists of an inner layer of luminal cells and an outer layer of myoepithelial cells that border the basal lamina. Compared to luminal cells, myoepithelial cells are significantly understudied and poorly characterized (25-30). Although their importance in mammary gland development and pathogenesis has been noted (26, 29, 31), myoepithelial cells have traditionally received far less attention than luminal cells. This is at least partially because they rarely proliferate in human breast cancer (32, 33). However, in canine MC, myoepithelial cell proliferation is much more common, occurring in >20% canine tumors compared to <0.1% human tumors (34). To more effectively utilize this unique feature of canine MC for a better understanding of myoepithelial cells, we set out to comprehensively compare spontaneous canine MCs with and without myoepithelial cell proliferation and to evaluate their molecular similarities to and differences from human breast cancers.

Materials and Methods

Canine Mammary Tissue Samples

Fresh-frozen (FF) and formalin-fixed paraffin-embedded (FFPE) normal and tumor tissue samples of spontaneous canine MC were obtained from the University of California-Davis School of Veterinary Medicine and the Animal Cancer Tissue Repository of the Colorado State University. Samples were collected from client-owned dogs that develop the disease spontaneously, under the guidelines of the Institutional Animal Care and Use Committee and with owner informed consent. The breed, age, histopathologic descriptions and other information are provided in Supplementary Table S1.

Immunohistochemical (IHC) Analyses

IHC experiments were performed following standard protocols with 5μm FFPE sections. Primary antibodies were used as described (35), including those against smooth muscle myosin heavy chain clone ID8 (MAB3568), acetyl-H4 (06-866), H3K9me3 (07-442) and H3-K4/K9-me3 (06-866), all from Millipore; H3K4me3 (Abcam, ab8580), estrogen receptor alpha clone E115 (Abcam, ab32063); and E-cadherin (R&D Systems, AF648). Alexa Fluor® 488, 647 or 594 conjugated secondary antibodies are from Jackson ImmunoResearch. Images were taken with a Zeiss LSM 710 confocal microscope.

Tissue Dissection, DNA and RNA Extraction, and PCR Analyses

Cryosectioning of FF tissues, H&E staining and cryomicrodissection were performed as described (5) to enrich tumor cells for tumor samples and mammary gland epithelial cells for normal samples. Genomic DNA and RNA were then extracted from the dissected tissues using the DNeasy Blood & Tissue Kit (cat. no. 69504), RNeasy Plus Mini Kit (cat. no. 74134) or AllPrep DNA/RNA Mini Kit (cat. no. 80204) from QIAGEN. Only samples with a 260/280 ratio of ~1.8 (DNA) or ~2.0 (RNA) and showing no degradation and other contaminations on the agarose gels were subjected to further analyses. The synthesis of cDNA, primer design, and PCR or qPCR with genomic DNA or cDNA samples were conducted as described (6). Primers used are listed in the Supplementary Methods.

aCGH Analyses

aCGH experiments were conducted at the Florida State University Microarray Facility, with 385K canine CGH array chips from Roche NimbleGen Systems, Inc. CNAs were identified as described (5).

Paired-end WGS, WES and RNA-seq

All three types of sequencing were conducted using the Illumina platform, following the protocols from the manufacture. Paired-end WGS of >12X sequence coverage was performed in collaboration with the Emory Genome Center (50bp or 100bp paired-end sequencing of ~200bp fragments) or the BGI-America (90bp paired-end sequencing of ~500bp fragments). WES was conducted in collaboration with the Hudsonalpha Institute for Biotechnology. First, exome-capturing was achieved by using a solution-based SureSelect kit from Agilent, covering 50Mb canine exons and adjacent regions. Then, paired-end sequences of 50bp of ~200bp fragments were generated from the captured targets to reach the coverage of 134-245X. RNA-seq was performed at Hudsonalpha, yielding 42 to 94 million paired-end sequence reads of 50bp per sample.

Sequence Data Analyses

Sequence read alignment, mutation discovery, translocation and chimeric fusion gene identification, clustering and other analyses are provided in the Supplementary Methods. Briefly, WGS, WES and RNA-seq sequence reads were aligned to the dog reference genome (11). Then, uniquely mapped WES reads were used to detect base substitutions and small indels, and significantly mutated genes were identified as described (20). Uniquely mapped WGS read pairs were used to identify somatic translocations and chimeric fusion genes. Uniquely mapped RNA-seq reads were used to quantify each gene's expression level, as well as to detect chimeric fusion transcripts and sequence mutations.

Results

Canine simple carcinomas have no myoepithelial cell proliferation, whereas canine complex carcinomas have luminal and myoepithelial cells both proliferating

The 12 cases subjected to genome-wide characterization represent two major histologic subtypes of canine MC (34), five with myoepithelial cell proliferation (complex carcinomas) and 7 without (simple carcinomas) (Supplementary Table S1). Tumor cells in simple carcinomas express only luminal markers such as E-cadherin (Figure 1A), and histologically match typical human breast carcinomas (Figures 1A and 1C). Tumors with myoepithelial cell proliferation include four complex carcinomas, a subtype that is rare in humans (32), and one carcinoma with two distinct histologic regions, one considered simple and the other considered complex. Complex carcinomas have prominent expression of both the luminal marker E-cadherin and the myoepithelial marker smooth muscle myosin heavy chain (SMHC) (Figure 1B), indicating dual proliferation of luminal and myoepithelial cells. This is also visible in H&E stained sections (Figure 1D). Besides this histologic difference, the tumors also vary in cancer progression stages (*in situ*, invasive or metastatic to the lung) and

in estrogen receptor (ER) expression (five ER+ tumors and 7 ER− tumors) (Supplementary Table S1).

Copy number abnormalities (CNAs) are frequent in canine simple carcinomas

Reminiscent of human breast cancer, canine simple carcinoma genomes harbor extensive CNAs. First, we observed both focal and broad CNAs totaling from 10Mb to >100Mb and affecting hundreds of genes per tumor (Figures 2A and 2B; Supplementary Tables S2A-S2C), with the extensiveness of CNA in correlation with the tumor progression stage. The only exception to this is an inflammatory carcinoma, where no CNAs were detected. Second, CNAs also occurred in genomic sites recurrently altered in human breast cancer, e.g., amplification of human 8q and dog chromosome 13 that encode genes including *MYC* (Figure 2A). Third, while large deletions were discovered, one resulting in *PTEN* loss (Figure 2A; Supplementary Table S2G), amplifications prevailed over deletions in most tumors. Notably, two large amplicons of >4Mb, harboring 54 and 43 genes respectively, were uncovered (Figure 2C). This led to amplification and overexpression of oncogenes such as *BRAF*, *PIM1* and *CCND3* (Supplementary Tables S2E and S2F).

Translocations and a superamplicon were discovered in a canine simple carcinoma by paired-end WGS

To further explore the two >4Mb amplicons described above, we sequenced the tumor and normal genomes of case 76 (Figure 2A) to a >15X sequence coverage (Supplementary Fig. S1 and Table S2D). For comparison purposes, similar sequencing was performed on the case having the most extensive CNAs (case 406434, with pulmonary metastasis) and another case having hardly any CNAs (case 32510). WGS revealed fewer translocations and inversions than CNAs in these tumors. Furthermore, reminiscent of the human breast cancer MCF7 genome (36), some translocations are associated with amplification, creating a superamplicon with loci from different chromosomes co-localized and co-amplified (Figure 2D).

Based on chimeric sequence reads that span the translocation junctions (Supplementary Table S2H) and PCR confirmation (Supplementary Fig. S1), we propose a mechanism for the superamplicon formation (Figure 2D). First, a circle, consisting of ~1Mb sequences from chromosome 12 and ~0.4Mb from chromosome 16, emerged via two translocations that were likely facilitated by prior sequence amplification. The circle, which harbor oncogene *PIM1* and 17 other genes (Supplementary Tables S2E and S2F), was then further amplified.

The superamplicon harbors a potentially oncogenic fusion gene, *ZFAND3-MGAM*, created via a translocation

The superamplicon also harbors a newly created fusion gene. It consists of the first four exons of *ZFAND3*, a zinc finger gene located on chromosome 12, and the last 22-49 exons of *MGAM*, which encodes maltase-glucoamylase and is located on chromosome 16 (Figure 2E). The fusion gene, termed *ZFAND3-MGAM*, arose from a translocation occurring in introns; transcription and splicing then yielded an in-frame fusion transcript. This was confirmed by the detection of chimeric sequence fusion points via WGS, RNA-seq and PCR analyses (Supplementary Fig. S1 and Table S2H).

As a result of in-frame fusion, the A20-type zinc finger domain of ZFAND3 and the glucoamylase domain of MGAM are kept intact in the fusion product (Figure 2F). This appears to be significant. First, the A20 zinc finger protein has been reported to inhibit tumor necrosis factor-induced apoptosis (37). Second, MGAM, an integral membrane protein with its catalytic domains facing the extracellular environment, is normally expressed in the intestine to digest starch into glucose (38). Indeed, we did not detect *MGAM* expression in normal mammary tissues, unlike *ZFAND3* (Supplementary Fig. S1C). However, in carcinoma 76, both *MGAM* and *ZFAND3-MGAM* are amplified and overexpressed. ZFAND3-MGAM, which lacks the transmembrane domain and becomes intracellular, could promote oncogenesis by accelerating carbohydrate-metabolism via its glucoamylase domain and meanwhile inhibiting apoptosis via its A20-type domain. Of course, whether this is true or not requires further studies.

Somatic sequence mutations are frequent in canine simple carcinomas as revealed by WES

To examine somatic base substitutions and small indels, we performed WES on the matching tumor and normal genomes of four simple carcinoma cases to 134-245X coverage (Supplementary Fig. S2 and Table S3A). This analysis again revealed several dog-human homologies. First, base transitions, particularly C→T/G→A changes, dominate base transversions in most tumors (Figure 3A), indicating similar mutation mechanisms (e.g., deamination of 5mC to T) in both species. The only exception (tumor 406434) has C→A/G→T transversions predominating, which is not an experimental artifact of WES (39) based on our analyses (Supplementary Table S3F), and concurrently harbors an altered *POLD1*. This likewise is consistent with human cancer studies that link C→A/G→T changes to *POLD1* mutations (40). Second, the mutation rate varies greatly among the carcinomas, with tumor 5 having 907 genes significantly mutated, compared to 0-31 genes for tumors of similar or more advanced stages (Figure 3A). This hypermutation is likely linked to defective DNA repair as well, because tumor 5 has as many as 24 DNA repair-associated genes mutated (Supplementary Table S3D). Third, many known human breast cancer genes are also mutated in these canine tumors (Supplementary Tables S3B and S3C), including *BRCA1*, *IGF2R*, *FOXC2*, *DLG2* and *USH2A* as described below.

USH2A is one of the most significantly mutated genes in our study (p = 2.78E-12), having one nonsense-, 12 missense- and three synonymous-mutations (Figure 3B; Supplementary Table S3G). Critically, *USH2A* is also prominently mutated in human breast cancer, ranking as the 21st most significantly mutated gene in the Cancer Genome Atlas (TCGA) study (23). *USH2A* alterations may contribute to MC pathogenesis in both dogs and humans.

Canine complex carcinomas have hardly any genomic CNAs and their sequence mutation rates also appear low

Unlike simple carcinomas, we observed very few genomic CNAs in complex carcinomas, making their genomes appear normal (Figure 2A; Supplementary Table S2A). Their sequence mutation rates are also low, based on our analysis with RNA-seq data. Briefly, to achieve a more accurate mutation-finding, we utilized only regions with 30-300X RNA-seq read coverage, which distribute across the genome and amount to 4.5-9.4 Mb sequence per

sample (Supplementary Table S3I). The analysis indicates that the mutation rates of complex carcinomas are significantly lower than those of simple carcinomas, but are comparable to those of normal mammary gland tissues (Figure 3C).

Numerous chromatin-modification genes are downregulated in canine complex carcinomas

RNA-seq analysis revealed 751 total genes differentially expressed at FDR ≤ 0.2 between simple and complex carcinomas (Figure 3D). Strikingly, among these genes, chromatin modification and transcription regulation are the most significantly enriched functions (Supplementary Tables S4A-S4C). Indeed, a total of 35 known chromatin-modification genes were found to be downregulated in complex carcinomas (Figure 3D; Supplementary Table S4E), and over 40% of them remain so at $p \leq 0.05$ when further compared to normal mammary gland tissues. Moreover, chromatin modification stays as the most significantly enriched function amid genes (327 in total) differentially expressed among the three types of samples (Supplementary Fig. S3, Tables S4H and S4I). The same overall conclusions were reached at FDR ≤ 0.1 (Supplementary Table S4F and S4G).

Amid the 35 chromatin genes downregulated in complex carcinomas, 30 encode histone modifiers, covering methylation and demethylation; acetylation and deacetylation; and ubiquitination and deubiquitination (Figure 3D). Intriguingly, both active and repressive modifiers were noted (see the paragraph follows). Furthermore, the identified histone acetyltransferases and deacetylase modify histones H3, H4 and H2A, influencing not only gene transcription (e.g., CREBBP) but also chromatin packing (e.g., MSL1 on H4K16 acetylation) (41). Besides histone-modification genes, other types of chromatin-remodeling genes were also found downregulated in complex carcinomas (Figure 3D), most of which (e.g., *ARID1B*, *ASF1A* and *DNMT3B*) are known to be mutated in human cancers (42, 43).

To understand the significance of the observed change in chromatin-modification genes, many encoding histone modifiers (Figure 3D), we investigated histone modification. Specifically, we performed IHC experiments to examine H3K9me3, a repressive modification that is associated with gene silencing and heterochromatin and for which six relevant genes are downregulated in complex carcinomas. These include H3K9 methyltransferase genes *SETDB1*, *EHMT1*, *EHMT2* and *SUZ12*, along with demethylase genes *JMJD1C* and *PHF2* (Supplementary Table S4E). Meanwhile, we also examined H4-acetylation because at least 8 of the downregulated genes (*CREBBP*, *CSRP2BP*, *ING3*, *KAT2A*, *MSL1*, *MSL2*, *NCOA3* and *SIRT1*) are involved in histone acetylation or deacetylation. Another active modification, H3K4me3, was studied as well because H3K4 methyltransferase genes *SETD1A*, *MLL2* and *MLL4* are among those downregulated.

In canine normal mammary glands, both active and repressive histone modifications are significantly depleted in myoepithelial cells when compared to luminal cells

To understand the alteration in cancer, we first investigated canine normal mammary glands where both luminal and myoepithelial cells are clearly visible. These include the normal tissue from case 159, where myoepithelial cells form a nearly continuous layer surrounding the luminal cells (159N in Figure 1A), and case 402421, where myoepithelial cells are not as

prominent but are still noticeable (402421N in Figure 1B). Interestingly, in these normal glands, active modifications H4-acetylation and H3K4me3 and repressive modification H3K9me3 are both significantly depleted in myoepithelial cells (Figure 4A; Supplementary Fig. S4), with the intensity reduced by half in most cases (Figure 4B), when compared to luminal cells.

In canine complex carcinomas, active modification H4-acetylation is abnormally enriched while repressive modification H3K9me3 is aberrantly depleted

Compared to normal mammary glands and simple carcinomas, complex carcinomas harbor significantly more myoepithelial cells (Figure 1). Yet unlike normal mammary glands (Figure 4), both myoepithelial and luminal cells in complex carcinomas were found to be equally enriched with active modifications (Figures 5A-5H; Supplementary Fig. S5 and Table S5). This is especially so for H4-acetylation, with the intensity being equal or stronger than luminal cells in normal mammary glands and in simple carcinomas. The repressive modification H3K9me3, to the contrary, becomes significantly more depleted in both cell types in complex carcinomas (Figures 5I-5L; Supplementary Fig. S5 and Table S5).

Redox genes are upregulated in canine ER+ carcinomas, while cell cycle and DNA repair genes are upregulated in canine ER− carcinomas

RNA-seq analyses also revealed a clear difference between canine ER+ and ER− tumors (Figure 6A), with most ER+ tumors being complex carcinomas while most ER− tumors being simple carcinomas. Among 1,350 differentially expressed genes at FDR ≤ 0.2 (Supplementary Table S6A), approximately half are upregulated in ER+ carcinomas and are significantly enriched in redox functions (Figure 6B; Supplementary Tables S6B, S6C and S6E). These genes encode ~25 dehydrogenases or oxidases, and 32 gene products are associated with mitochondria including the electron transport chain. Another half of the 1,350 differentially expressed genes are upregulated in ER− carcinomas, among which ~118 genes are associated with the cell cycle, e.g., mitosis, spindle, microtubule cytoskeleton, etc. (Figure 6B; Supplementary Tables S6D and S6F). Other significant functions comprise DNA repair (38 genes) and protein serine/threonine kinase activity (17 genes). The same overall conclusions were reached at FDR ≤ 0.1 (Supplementary Fig. S6A).

Canine simple carcinomas and the ER− complex carcinoma cluster closely with basal-like human breast carcinomas in PAM50 classification

To directly compare the canine MCs to human breast cancers, we randomly selected 20 human tumors for each subtype among a total of 195 luminal A, 111 luminal B, 53 HER2-enriched and 87 basal-like tumors of the TCGA RNA-seq study (23). This, along with all 7 normal-like tumors in TCGA, amount to 87 human carcinomas covering all five intrinsic subtypes. We then performed PAM50 clustering (44) on these 87 human carcinomas together with our 12 canine carcinomas. This analysis was repeated 100 times, ensuring that each TCGA tumor was sampled at least once. Notably, in 82 out of 100 times, all canine simple carcinomas and the ER− complex carcinoma (ID 518) group with the human basal-like tumors. The remaining canine complex carcinomas (all ER+), however, fail to cluster with any specific human subtypes. One clustering example is shown in Figure 6C and Supplementary Fig. S6B.

Discussion

In this study, we performed an initial comprehensive characterization of the genomes, transcriptomes and epigenomes of two major canine MC histologic subtypes. Even with a small sample size (12 cases), the analysis reveals a remarkable molecular heterogeneity of spontaneous canine MCs. It also emphasizes their unique value and raises a number of important questions that could profoundly impact human breast cancer research.

Canine simple carcinomas, without myoepithelial cell proliferation, harbor extensive genomic aberrations and are molecularly homologous to human breast carcinomas

Canine simple carcinomas investigated have no myoepithelial cell proliferation and are histologically comparable to human *in situ* or invasive ductal or lobular carcinomas. Significantly, these canine cancers faithfully recapitulate key molecular features of human breast cancer. First, analogous to their human counterparts (18-23, 36), the genomes of these canine carcinomas harbor extensive genetic lesions including numerous CNAs, fusion gene-creating translocation, equally complex superamplicon and comparable sequence mutation types. The only exception is an inflammatory carcinoma, whose human equivalent (inflammatory breast cancer) is also devoid of CNAs (21). Second, notable human breast cancer genes (18-20, 23, 45, 46) are altered in these canine cancers as well. Examples include: 1) amplification and/or overexpression of oncogenes *BRAF*, *MYC*, *PIK3CA*, *PIK3R1*, *CCND3* and *TBX3*; 2) deletion and/or underexpression of tumor suppressors *PTEN*, *PTPRD* and *CDH1* (47); and 3) mutations of *BRCA1*, *NF1*, *MAP3K1* and *RUNX1* (Supplementary Table S7B). Third, many of the altered pathways are shared between the two species, e.g., cell adhesion, Wnt-signaling, PI3K signaling and DNA repair (Figure 7C; Supplementary Table S7) (23), consistent with other canine MC studies (15, 17).

These strong molecular homologies make canine simple carcinomas valuable in human breast cancer research. For example, for cancers with large genomic CNAs, we can apply the dog-human comparison strategy for effective driver-passenger discrimination as described (7). Critically, as elegantly discussed in several publications (2-4), these canine cancers, which bridge a gap between traditional rodent models and human clinical trials, can significantly speed up new anticancer drug development. For example, for drugs targeting the PI3K pathway (48) (Figure 7C), their efficacy, toxicity, dosage and schedule can be more accurately evaluated through clinical trials with canine patients, before entering human clinical trials. This will significantly reduce the cost and accelerate the drug discovery process.

Can canine simple carcinomas serve as a much-needed spontaneous cancer model of basal-like human breast cancer?

Canine simple carcinomas cluster with basal-like human tumors with an 82% chance in our PAM50 classification, indicating their closer resemblance to this subtype than other intrinsic subtypes. This may be explained by the observation that none of the canine tumors carrying HER2 amplification or overexpression. Furthermore, many harbor extensive CNAs and are ER− with genes related to DNA repair and cell cycle significantly upregulated, consistent with the basal-like subtype profile (23). This is especially so considering that the ER−

complex carcinoma clusters similarly as well. The only ER+ canine simple carcinoma has a prominent *PTEN* deletion, also a feature of basal-like tumors (23).

Clearly, studies with a larger sample size are needed to determine if canine simple carcinomas as a whole or even just a subset do indeed closely match the basal-like subtype. If confirmed, these canine cancers could serve as a much-needed spontaneous cancer model. Compared to other subtypes, basal-like cancers are aggressive, have a poor prognosis and currently lack effective treatments. Canine MC could make significant contributions towards understanding and treating this worst subtype of human breast cancer.

Canine complex carcinomas, with myoepithelial cell proliferation, appear to originate from epigenomic rather than genomic alterations

Complex carcinomas, featuring dual proliferation of luminal and myoepithelial cells, likely originate from epigenomic, rather than genomic, abnormalities (Figure 7). First, their genomes appear normal without CNAs detected and with sequence mutation rates as low as normal tissues. Thus, it is unlikely that these tumors arise from genetic aberrations, unlike simple carcinomas. Meanwhile, complex carcinomas could acquire genomic changes as they progress to later stages, as shown by tumor 518 (Supplementary Table S2A) and another complex carcinoma with pulmonary metastasis (14). Second, chromatin modification and transcription regulation stand out as the most enriched functions among genes differentially expressed between simple and complex carcinomas, with numerous chromatin-modification genes downregulated in complex carcinomas. Importantly, complex carcinomas are aberrantly enriched with active histone modification H4-acetylation while abnormally depleted with repressive modification H3K9me3. Thus, it is possible that the epigenomes of complex carcinomas are altered, turning on genes that normally should be silenced to promote tumorigenesis. Obviously, more studies are needed to confirm this possibility and to understand the underlying mechanisms.

Myoepithelial cell proliferation is rare in human breast cancer (32, 33). As a result, myoepithelial cells receive far less attention than luminal cells and are poorly understood (25, 27, 28, 30). However, the few laboratories that study myoepithelial cells have noted their importance. For example, myoepithelial cells are thought to be a part of the mammary stem cell niche, mediate the cross-talk between luminal cells and extracellular matrix, contribute to the maintenance of the apicobasal polarity of luminal cells, and serve as a tumor suppressor (26, 29, 31). Canine MC, where myoepithelial cell proliferation is much more common, provides an ideal system to address such functions and to better understand the 2nd major cell lineage of the mammary gland (e.g., by answering questions such as whether a prolonged luteal phase promotes myoepithelial cell proliferation).

Do canine complex carcinomas derive from mammary gland stem cells or luminal/myoepithelial common progenitors?

Several observations indicate the possibility that complex carcinomas arise from mammary gland stem cells or luminal/myoepithelial common progenitors (Figure 7A). First, one of these tumors (ID 518) expresses the pluripotency marker *SOX2*, indicating stem cell property. Second, unlike normal mammary glands which present a clearly different

epigenomic landscape between luminal and myoepithelial cells, no such difference was observed in complex carcinomas. This indicates that proliferating luminal and myoepithelial cells in complex carcinomas may have derived from altered common progenitors. Third, compared to simple carcinomas and normal mammary tissues, glucose metabolic genes are upregulated in complex carcinomas, consistent with this stem cell or progenitor origin theory. For simple carcinomas, we hypothesize that they originate from either luminal progenitors, because of their close resemblance to the basal-like subtype which is reported to have a luminal progenitor origin (49), or differentiated luminal cells because of luminal cell properties (see case 159 in Figure 1). Of course, further studies with a larger sample size are needed to test these hypotheses.

In summary, we performed the first comprehensive characterization of the genomes, transcriptomes and epigenomes of canine simple carcinomas and complex carcinomas, two major histologic subtypes of canine MC. The analysis reveals that canine simple carcinomas, which have no myoepithelial cell proliferation and histologically match typical human breast carcinomas, faithfully recapitulate many molecular features of human breast cancer. Notably, canine simple carcinomas closely cluster with basal-like human breast tumors in PAM50 classification, and thus could serve as a much-needed spontaneous cancer model for the basal-like subtype. Canine complex carcinomas are characterized with dual proliferation of luminal and myoepithelial cells, which is rare in human breast cancer. Our analysis indicates that these canine cancers may arise from epigenomic rather genomic alterations. Canine complex carcinomas hence provide a unique system to investigate the roles of myoepithelial cells, the 2nd major cell lineage of the mammary gland, in normal developmental and pathogenic processes.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank Drs. Timothy D. Read, Shawn Levy, Steven H. Miller, and their respective teams, and the BGI, for their outstanding sequencing or aCGH work; Ms. Teri Guerrero, Ms. Irene Mok and Dr. Susan E Lana for their help on sample collection; Dr. John M. Rosenfeld of Millipore for providing antibodies; Drs. Nancy Manley and Jie Li for their help on IHC and H&E analyses; and Drs. Dong M Shin, J David Puett, Charles M Perou and Malcolm Hayes for their help on the study. Confocal imaging was performed at the UGA Biomedical Microscopy Core.

Financial support: This work was funded by the American Cancer Society, the Georgia Cancer Coalition and the AKC Canine Health Foundation to S. Zhao, as well as by pilot project funds from NCI P50 CA128613 (PI: Dr. Dong M Shin) and GM085354 (PI: S. Dalton).

References

1. Meuten, DJ. Tumors in domestic animals. 4th ed.. Iowa State University Press; Ames, Iowa: 2002.
2. Paoloni M, Khanna C. Translation of new cancer treatments from pet dogs to humans. Nature reviews Cancer. 2008; 8:147–56.
3. Rowell JL, McCarthy DO, Alvarez CE. Dog models of naturally occurring cancer. Trends in molecular medicine. 2011; 17:380–8. [PubMed: 21439907]

4. Gordon I, Paoloni M, Mazcko C, Khanna C. The Comparative Oncology Trials Consortium: using spontaneously occurring cancers in dogs to inform the cancer drug development pathway. PLoS medicine. 2009; 6:e1000161. [PubMed: 19823573]
5. Tang J, Le S, Sun L, Yan X, Zhang M, Macleod J, et al. Copy number abnormalities in sporadic canine colorectal cancers. Genome research. 2010; 20:341–50. [PubMed: 20086242]
6. Youmans L, Taylor C, Shin E, Harrell A, Ellis AE, Seguin B, et al. Frequent alteration of the tumor suppressor gene APC in sporadic canine colorectal tumors. PloS one. 2012; 7:e50813. [PubMed: 23251390]
7. Tang J, Li Y, Lyon K, Camps J, Dalton S, Ried T, et al. Cancer driver-passenger distinction via sporadic human and dog cancer comparison: a proof-of-principle study with colorectal cancer. Oncogene. 2013
8. Parker HG, Shearin AL, Ostrander EA. Man's best friend becomes biology's best in show: genome analyses in the domestic dog. Annual review of genetics. 2010; 44:309–36.
9. Nasir L, Devlin P, McKevitt T, Rutteman G, Argyle DJ. Telomere lengths and telomerase activity in dog tissues: a potential model system to study human telomere and telomerase biology. Neoplasia. 2001; 3:351–9. [PubMed: 11571635]
10. Rangarajan A, Weinberg RA. Opinion: Comparative biology of mouse versus human cells: modelling human cancer in mice. Nature reviews Cancer. 2003; 3:952–9.
11. Lindblad-Toh K, Wade CM, Mikkelsen TS, Karlsson EK, Jaffe DB, Kamal M, et al. Genome sequence, comparative analysis and haplotype structure of the domestic dog. Nature. 2005; 438:803–19. [PubMed: 16341006]
12. Ji X, Zhao S. DA and Xiao-two giant and composite LTR-retrotransposon-like elements identified in the human genome. Genomics. 2008; 91:249–58. [PubMed: 18083327]
13. Siegel R, Naishadham D, Jemal A. Cancer statistics, 2012. CA: a cancer journal for clinicians. 2012; 62:10–29. [PubMed: 22237781]
14. Beck J, Hennecke S, Bornemann-Kolatzki K, Urnovitz HB, Neumann S, Strobel P, et al. Genome aberrations in canine mammary carcinomas and their detection in cell-free plasma DNA. PloS one. 2013; 8:e75485. [PubMed: 24098698]
15. 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. [PubMed: 21062462]
16. Paw Owski KM, Maciejewski H, Dolka I, Mol JA, Motyl T, Krol M. Gene expression profiles in canine mammary carcinomas of various grades of malignancy. BMC veterinary research. 2013; 9:78. [PubMed: 23587192]
17. Rao NA, van Wolferen ME, Gracanin A, Bhatti SF, Krol M, Holstege FC, et al. Gene expression profiles of progestin-induced canine mammary hyperplasia and spontaneous mammary tumors. Journal of physiology and pharmacology : an official journal of the Polish Physiological Society. 2009; 60(Suppl 1):73–84. [PubMed: 19609016]
18. Stephens PJ, Tarpey PS, Davies H, Van Loo P, Greenman C, Wedge DC, et al. The landscape of cancer genes and mutational processes in breast cancer. Nature. 2012; 486:400–4. [PubMed: 22722201]
19. Banerji S, Cibulskis K, Rangel-Escareno C, Brown KK, Carter SL, Frederick AM, et al. Sequence analysis of mutations and translocations across breast cancer subtypes. Nature. 2012; 486:405–9. [PubMed: 22722202]
20. Sjoblom T, Jones S, Wood LD, Parsons DW, Lin J, Barber TD, et al. The consensus coding sequences of human breast and colorectal cancers. Science. 2006; 314:268–74. [PubMed: 16959974]
21. Curtis C, Shah SP, Chin SF, Turashvili G, Rueda OM, Dunning MJ, et al. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. Nature. 2012; 486:346–52. [PubMed: 22522925]
22. Naylor TL, Greshock J, Wang Y, Colligon T, Yu QC, Clemmer V, et al. High resolution genomic analysis of sporadic breast cancer using array-based comparative genomic hybridization. Breast cancer research : BCR. 2005; 7:R1186–98. [PubMed: 16457699]

23. Cancer Genome Atlas N. Comprehensive molecular portraits of human breast tumours. Nature. 2012; 490:61–70. [PubMed: 23000897]
24. Sleeckx N, de Rooster H, Veldhuis Kroeze EJ, Van Ginneken C, Van Brantegem L. Canine mammary tumours, an overview. Reproduction in domestic animals = Zuchthygiene. 2011; 46:1112–31. [PubMed: 21645126]
25. Adriance MC, Inman JL, Petersen OW, Bissell MJ. Myoepithelial cells: good fences make good neighbors. Breast cancer research : BCR. 2005; 7:190–7. [PubMed: 16168137]
26. Gudjonsson T, Adriance MC, Sternlicht MD, Petersen OW, Bissell MJ. Myoepithelial cells: their origin and function in breast morphogenesis and neoplasia. Journal of mammary gland biology and neoplasia. 2005; 10:261–72. [PubMed: 16807805]
27. Lakhani SR, O'Hare MJ. The mammary myoepithelial cell--Cinderella or ugly sister? Breast cancer research : BCR. 2001; 3:1–4. [PubMed: 11250738]
28. Moumen M, Chiche A, Cagnet S, Petit V, Raymond K, Faraldo MM, et al. The mammary myoepithelial cell. The International journal of developmental biology. 2011; 55:763–71. [PubMed: 21948739]
29. Polyak K, Hu M. Do myoepithelial cells hold the key for breast tumor progression? Journal of mammary gland biology and neoplasia. 2005; 10:231–47. [PubMed: 16807803]
30. Sopel M. The myoepithelial cell: its role in normal mammary glands and breast cancer. Folia morphologica. 2010; 69:1–14. [PubMed: 20235044]
31. Hu M, Yao J, Carroll DK, Weremowicz S, Chen H, Carrasco D, et al. Regulation of in situ to invasive breast carcinoma transition. Cancer cell. 2008; 13:394–406. [PubMed: 18455123]
32. Tan PH, Ellis IO. Myoepithelial and epithelial-myoepithelial, mesenchymal and fibroepithelial breast lesions: updates from the WHO Classification of Tumours of the Breast 2012. Journal of clinical pathology. 2013
33. Hayes MM. Adenomyoepithelioma of the breast: a review stressing its propensity for malignant transformation. J Clin Pathol. 2011; 64:477–84. [PubMed: 21307156]
34. Goldschmidt M, Pena L, Rasotto R, Zappulli V. Classification and grading of canine mammary tumors. Veterinary pathology. 2011; 48:117–31. [PubMed: 21266722]
35. Bryson JL, Griffith AV, Hughes B 3rd, Saito F, Takahama Y, Richie ER, et al. Cell-autonomous defects in thymic epithelial cells disrupt endothelial-perivascular cell interactions in the mouse thymus. PloS one. 2013; 8:e65196. [PubMed: 23750244]
36. Volik S, Zhao S, Chin K, Brebner JH, Herndon DR, Tao Q, et al. End-sequence profiling: sequence-based analysis of aberrant genomes. Proceedings of the National Academy of Sciences of the United States of America. 2003; 100:7696–701. [PubMed: 12788976]
37. Lademann U, Kallunki T, Jaattela M. A20 zinc finger protein inhibits TNF-induced apoptosis and stress response early in the signaling cascades and independently of binding to TRAF2 or 14-3-3 proteins. Cell death and differentiation. 2001; 8:265–72. [PubMed: 11319609]
38. Sauer J, Sigurskjold BW, Christensen U, Frandsen TP, Mirgorodskaya E, Harrison M, et al. Glucoamylase: structure/function relationships, and protein engineering. Biochimica et biophysica acta. 2000; 1543:275–93. [PubMed: 11150611]
39. Costello M, Pugh TJ, Fennell TJ, Stewart C, Lichtenstein L, Meldrim JC, et al. Discovery and characterization of artifactual mutations in deep coverage targeted capture sequencing data due to oxidative DNA damage during sample preparation. Nucleic acids research. 2013
40. Palles C, Cazier JB, Howarth KM, Domingo E, Jones AM, Broderick P, et al. Germline mutations affecting the proofreading domains of POLE and POLD1 predispose to colorectal adenomas and carcinomas. Nature genetics. 2012; 45:136–44. [PubMed: 23263490]
41. Fraga MF, Ballestar E, Villar-Garea A, Boix-Chornet M, Espada J, Schotta G, et al. Loss of acetylation at Lys16 and trimethylation at Lys20 of histone H4 is a common hallmark of human cancer. Nature genetics. 2005; 37:391–400. [PubMed: 15765097]
42. Wilson BG, Roberts CW. SWI/SNF nucleosome remodellers and cancer. Nature reviews Cancer. 2011; 11:481–92.
43. Dawson MA, Kouzarides T. Cancer epigenetics: from mechanism to therapy. Cell. 2012; 150:12–27. [PubMed: 22770212]

44. Prat A, Perou CM. Deconstructing the molecular portraits of breast cancer. Molecular oncology. 2011; 5:5–23. [PubMed: 21147047]
45. Futreal PA, Coin L, Marshall M, Down T, Hubbard T, Wooster R, et al. A census of human cancer genes. Nature reviews Cancer. 2004; 4:177–83.
46. Samuels Y, Wang ZH, Bardelli A, Silliman N, Ptak J, Szabo S, et al. High frequency of mutations of the PIK3CA gene in human cancers. Science. 2004; 304:554. [PubMed: 15016963]
47. Wendt MK, Taylor MA, Schiemann BJ, Schiemann WP. Down-regulation of epithelial cadherin is required to initiate metastatic outgrowth of breast cancer. Molecular biology of the cell. 2011; 22:2423–35. [PubMed: 21613543]
48. Gordon V, Banerji S. Molecular pathways: PI3K pathway targets in triple-negative breast cancers. Clinical cancer research : an official journal of the American Association for Cancer Research. 2013; 19:3738–44. [PubMed: 23748695]
49. Molyneux G, Geyer FC, Magnay FA, McCarthy A, Kendrick H, Natrajan R, et al. BRCA1 basal-like breast cancers originate from luminal epithelial progenitors and not from basal stem cells. Cell stem cell. 2010; 7:403–17. [PubMed: 20804975]
50. Lim E, Vaillant F, Wu D, Forrest NC, Pal B, Hart AH, et al. Aberrant luminal progenitors as the candidate target population for basal tumor development in BRCA1 mutation carriers. Nature medicine. 2009; 15:907–13.

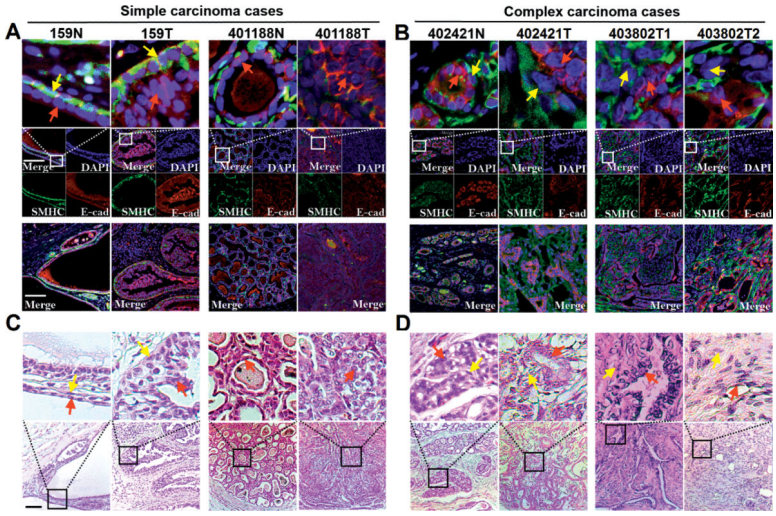

Figure 1. Myoepithelial cell proliferation is absent in canine simple carcinomas but prominent in canine complex carcinomas

A and B, representative images of immunostaining with the myoepithelial marker SMHC and the luminal marker E-cadherin (E-cad) of normal (N) and tumor (T) tissues of two simple carcinoma cases (A), one *in situ* (ID 159) and the other invasive (ID 401188), and two complex carcinomas (B). Top panel shows the enlarged view of the areas indicated below. Red arrows point to luminal cells, while yellow arrows point to myoepithelial cells. Scale bar = 100 μm. **C and D**, H&E staining of the same tissues.

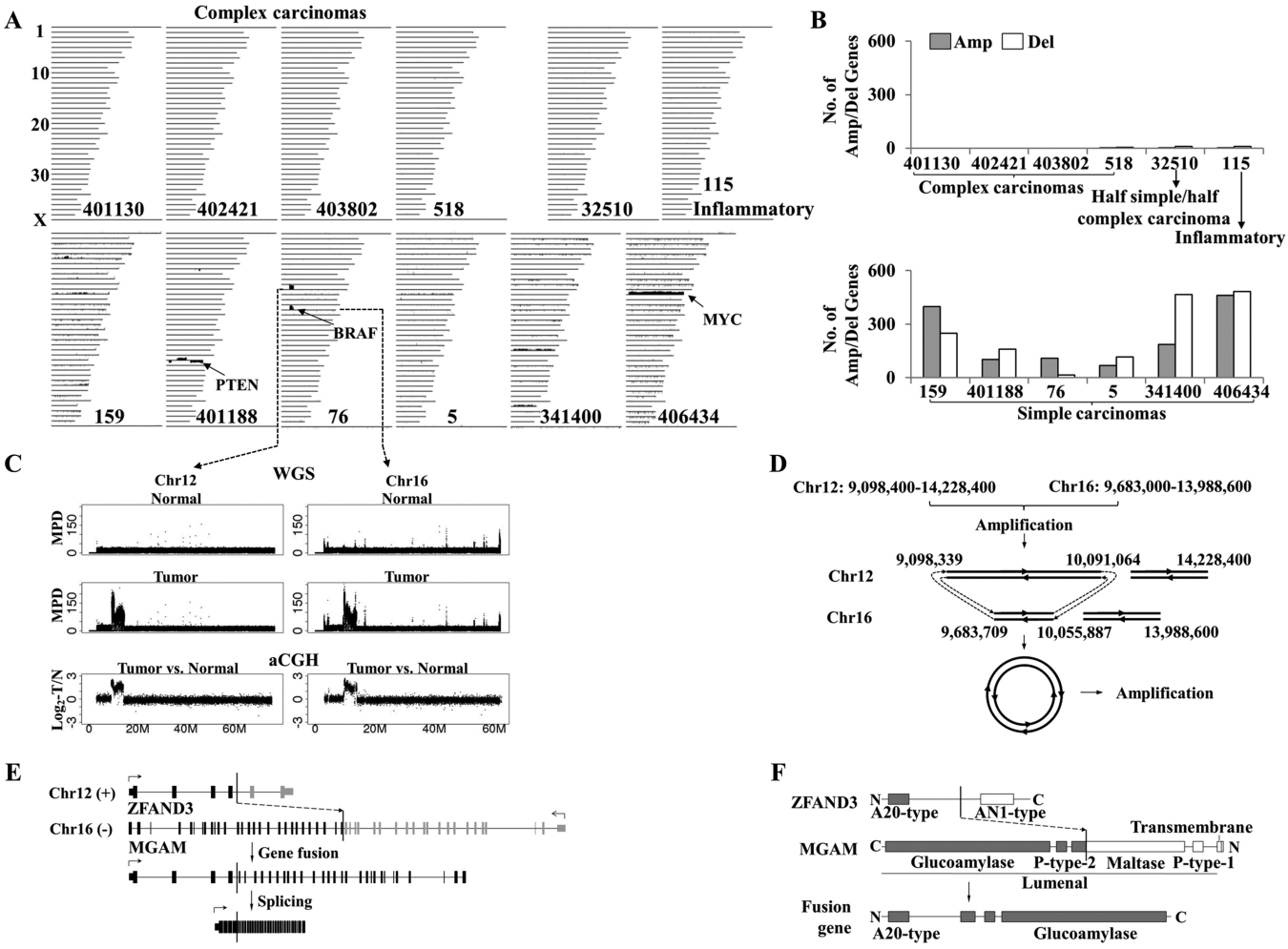

Figure 2. Large scale genomic aberrations are frequent in canine simple carcinomas but rare in canine complex carcinomas

A, CNAs found in four complex carcinomas (labeled), one half complex and half simple carcinoma (ID 32510) and 7 simple carcinomas (which include the inflammatory tumor 115 and the six tumors at the 2nd panel) by aCGH. The images were drawn as described (5), with each line representing a canine chromosome and vertical lines above/below the chromosome indicating amplifications/deletions, respectively. Notable amplified/deleted genes are shown. **B**, the total numbers of amplified (shaded bars) and deleted (empty bars) genes of each carcinoma shown in A.

C, two >4Mb amplicons discovered in simple carcinoma 76 in A, by both WGS and aCGH. The X-axis indicates chromosomal coordinates in Mb, while the Y-axis indicates the mapped read pair density (MPD) values of WGS or the tumor against normal $\log_2$-ratios of aCGH.

D, the proposed mechanism for superamplicon formation. Prior sequence amplifications led to two translocations (represented by the dashed lines), resulting in a circle which was further amplified. The numbers indicate the chromosomal coordinates in bp.

E, a fusion gene created by the 2nd translocation shown in D. The translocation occurred in the intron of both genes as indicated (exons are represented by the vertical bars). An in-frame fusion transcript then emerged via splicing.

F, the A20-type domain of ZFAND3 and the glucoamylase domain of MGAM are preserved in the fusion protein.

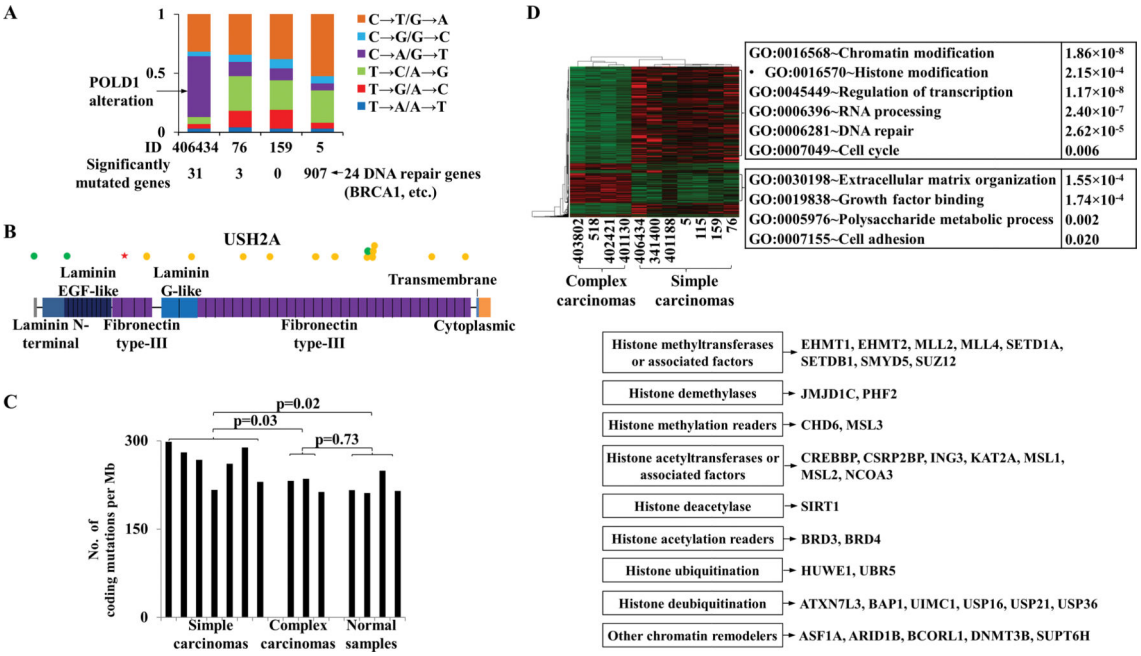

Figure 3. Coding sequence mutations are frequent in canine simple carcinomas; chromatin-modification genes are downregulated in canine complex carcinomas

A, the fractions (the Y-axis) of somatic base substitution types of simple carcinomas (IDs indicated by the X-axis) detected by WES. The total number of significantly mutated genes in each tumor is also shown, and tumor 5 has many DNA repair genes mutated.

B, synonymous (green dots) and non-synonymous substitutions (yellow dots), and a nonsense mutation (red star) uncovered in the *USH2A* gene in tumor 5.

C, the base substitution (compared to the dog reference genome) rates of the three sample types in coding regions with 30-300X RNA-seq read coverage. The p-values were calculated by t-tests.

D, the heatmap of 751 genes differentially expressed at FDR ≤ 0.2 between simple and complex carcinomas (red: upregulation; green: downregulation). The right panel illustrates the enriched functions of each gene cluster indicated, and the 35 chromatin modifiers downregulated in complex carcinomas are specified below.

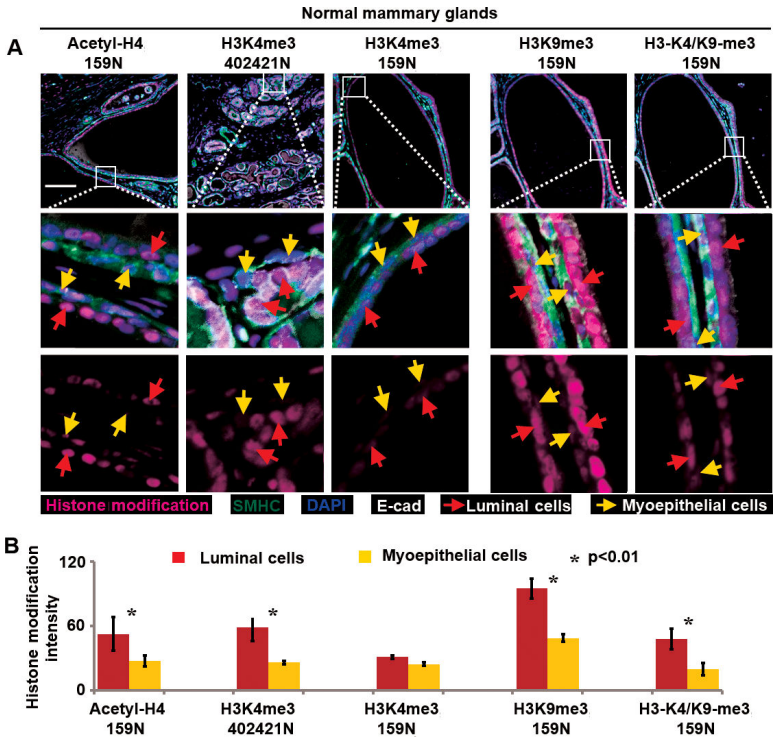

Figure 4. In canine normal mammary glands, active and repressive histone modifications are both depleted in myoepithelial cells when compared to luminal cells

A, representative IHC images of active modifications acetyl-H4 and H3K4me3, repressive modification H3K9me3, and modification H3-K4/K9-me3 (positive only when H3K4me3 and H3K9me3 are present simultaneously). Yellow arrows point to myoepithelial cells, while red arrows point to luminal cells. Scale bar = 100 μm.

B, the intensity of each histone modification was measured from ≥ 10 individual cells of each type from different regions across the tissue section. The p-values were calculated by Wilcoxon tests.

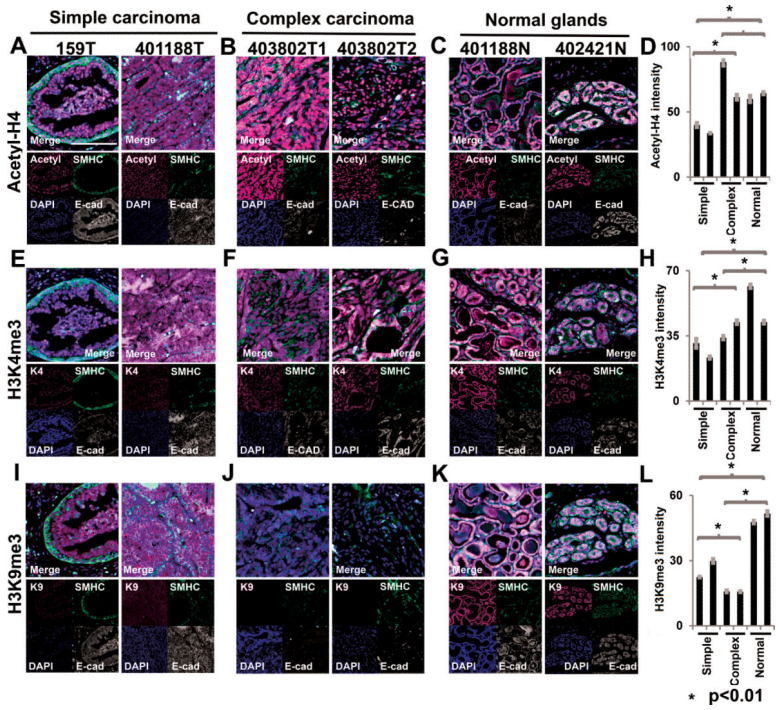

Figure 5. In canine complex carcinomas, active histone modification H4-acetylation is enriched while repressive modification H3K9me3 is depleted in both luminal and myoepithelial cells

A-C, representative IHC images of acetyl-H4 of simple carcinomas (A), complex carcinomas (B), and normal mammary glands (C). The merged (top) and split images (Acetyl: acetyl-H4) are shown. Scale bar = 100 μm.

D, the immunofluorescence staining intensity of acetyl-H4 determined from at least three different areas of the first split image, labeled as “Acetyl”, in A-C. The pairwise comparison p-values were calculated by Wilcoxon tests.

E-H for H3K4me3 (K4) and **I-L** for H3K9me3 (K9) are presented in the same way as A-D. Unlike the 159N sample of Figure 4A, the normal mammary glands in C, G and K consist of mostly luminal cells, with myoepithelial cells either absent (401188N) or very few (402421N).

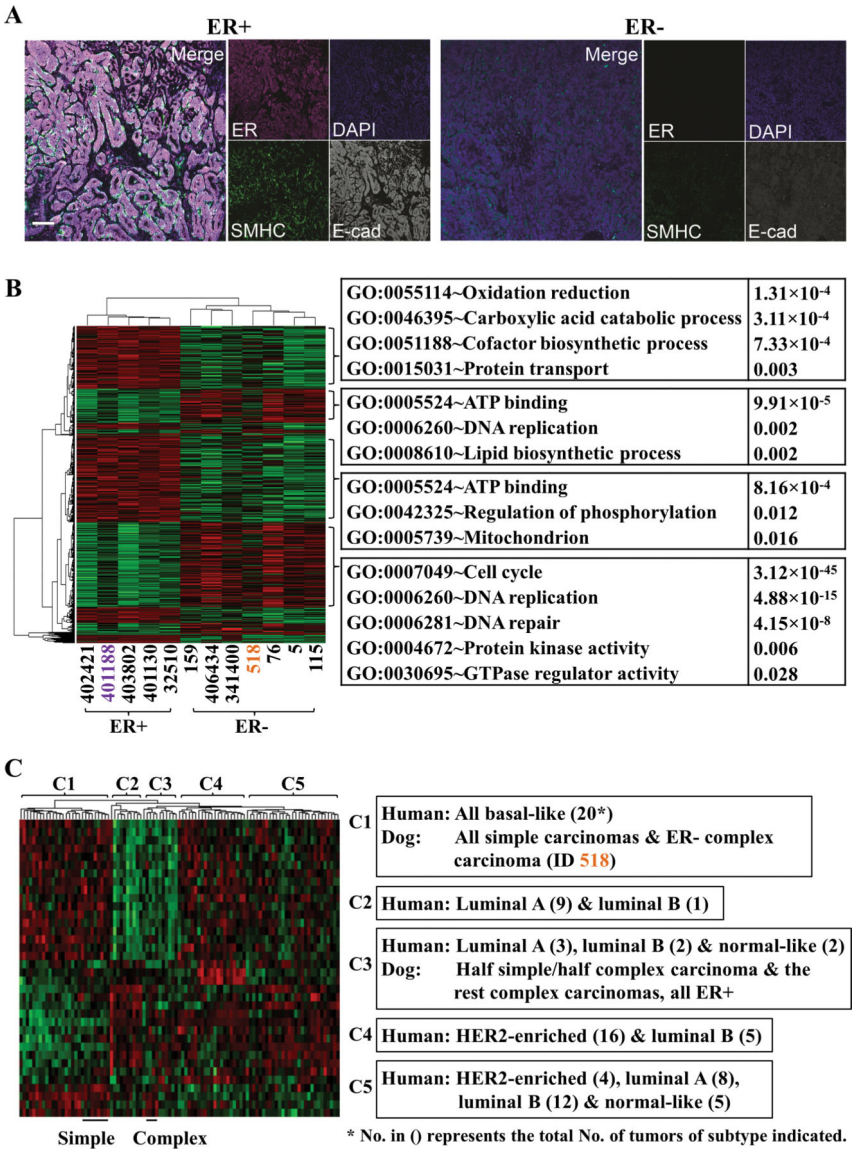

Figure 6. Subtype analysis reveals homology between canine MCs and human breast cancers

A, examples of ER+ (ID 401188) and ER− (ID 341400) canine carcinomas determined by IHC. Scale bar = 100 μm.

B, genes (1,350 total) differentially expressed between ER+ and ER− canine carcinomas at FDR ≤ 0.2. The tumor IDs are indicated. Tumor 401188 (purple) is ER+ but a simple carcinomas, while tumor 518 (orange) is ER− but a complex carcinoma. The right panel illustrates the significantly enriched functions of each gene cluster indicated.

C, Canine simple carcinomas and the ER− complex carcinoma cluster with the basal-like human breast carcinomas in PAM50 classification. The heatmap represents a clustering example of 12 canine tumors and 87 TCGA human tumors (see text). The composition of each cluster specified at the top of the heatmap is explained at the right side.

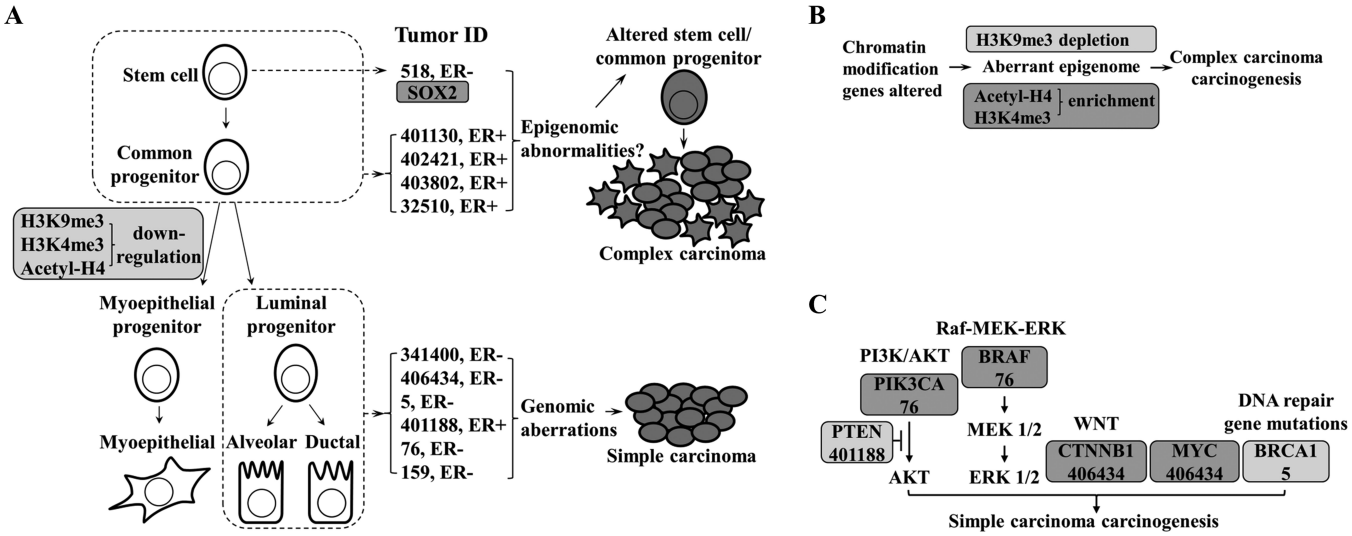

Figure 7. Canine complex carcinomas possibly arise from epigenomic alterations, whereas canine simple carcinomas likely originate from genomic aberrations

A, the proposed carcinogenic mechanism. The mammary gland development hierarchy is modified from a publication (50).

B, epigenomic alterations in complex carcinomas, with histone modifications enriched (darker shading) or depleted (lighter shading).

C, genomic alterations in simple carcinomas, with notable gene and pathway alterations (activation: darker shading; inactivation: lighter shading) indicated in the respective tumors (e.g., PTEN deletion in tumor 401188).