Contents lists available at ScienceDirect

Heliyon

Research article

Machine learning determines stemness associated with simple and basal-like canine mammary carcinomas

Pedro L.P. Xavier [a],[*],[1], Maycon Marção [b],[1], Renan L.S. Simões [b],
Maria Eduarda G. Job [a], Ricardo de Francisco Strefezzi [a], Heidge Fukumasu [a],
Tathiane M. Malta [b],[2]

[a] *Laboratory of Comparative and Translational Oncology (LOCT), Department of Veterinary Medicine, Faculdade de Zootecnia e Engenharia de Alimentos, Universidade de São Paulo, Pirassununga, São Paulo, Brazil*
[b] *Cancer Epigenomics Laboratory, Department of Clinical Analysis, Toxicology and Food Sciences, Faculdade de Ciências Farmacêuticas de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto, São Paulo, Brazil*

ARTICLE INFO

ABSTRACT

Simple and complex carcinomas are the most common type of malignant Canine Mammary Tumors (CMTs), with simple carcinomas exhibiting aggressive behavior and poorer prognostic. Stemness is an ability associated with cancer initiation, malignancy, and therapeutic resistance, but is still few elucidated in canine mammary tumor subtypes. Here, we first validated, using CMT samples, a previously published canine one-class logistic regression machine learning algorithm (OCLR) to predict stemness (mRNAsi) in canine cancer cells. Then, using the canine mRNAsi, we observed that simple carcinomas exhibit higher stemness than complex carcinomas and other histological subtypes. Also, we confirmed that stemness is higher and associated with basal-like CMTs and with NMF2 metagene signature, a tumor-specific DNA-repair metagene signature. Using correlation analysis, we selected the top 50 genes correlated with higher stemness, and the top 50 genes correlated with lower stemness and further performed a gene set enrichment analysis to observe the biological processes enriched for these genes. Finally, we suggested two promise stemness-associated targets in CMTs, *POLA2* and *APEX1*, especially in simple carcinomas. Thus, our work elucidates stemness as a potential mechanism behind the aggressiveness and development of canine mammary tumors, especially in simple carcinomas, describing evidence of a promising strategy to target this disease.

1. Introduction

Canine mammary tumors (CMTs) are the most common type of cancer in female dogs and have been widely studied as a

* Corresponding author.
E-mail address: porfirioxavier@usp.br (P.L.P. Xavier).
[1] These authors contributed equally.
[2] Senior author.

comparative model for breast cancer in women [1,2]. Both share similar biological and clinical patterns, including a high incidence, hormonal influence, and histopathological profiles. Additionally, some similarities in molecular and genetic makeup can be observed, such as copy number variation in important breast cancer-associated genes such as *PTEN*, *MYC*, *BRCA1*, and *BRCA2* [3,4]. The histological features of mammary tumors in dogs and humans can be different, with benign tumors being more prevalent in CMTs [5]. Furthermore, CMTs often exhibit mesenchymal tumors, such as fibrosarcomas and carcinosarcomas, and the proliferation of myoepithelial cells in complex adenomas and carcinomas, extremely rare in human breast cancers [6,7].

Simple carcinomas are composed of only one cell population, resembling either luminal epithelial or myoepithelial cells, whereas complex carcinomas resemble two different cell populations: malignant epithelial luminal cells and myoepithelial cells. Dogs with simple carcinomas have poorer prognoses than those with other subtypes [7–10]. However, the mechanisms underlying the association between simple CMTs and poorer prognosis remain unclear, and few studies have investigated this. In recent years, omics-based approaches have helped advance knowledge regarding the genetic and molecular landscape of these types of CMT. Liu et al. [11], using whole-exome sequencing (WES) and transcriptomics (RNA-Seq), identified and suggested that simple CMT probably arises from genomic aberrations, whereas complex CMT originates from epigenomic alterations. More recently, well-designed studies have traced molecular oncogenic signatures in CMT, identifying relevant mechanisms that would lead to a better understanding of the underlying molecular pathogenesis of CMT and comparing it with human breast cancer [12,13].

Stemness is the ability of stem cells (SC) to self-renew and differentiate, ensuring the maintenance of at least one new undifferentiated SC, but also to differentiate into mature cells of different adult organs [14]. It is also a phenotype associated with the cancer stem cell (CSC) concept and is consequently interconnected with intratumoral heterogeneity, initiation, progression, multidrug resistance, and metastasis [15]. Some studies have elucidated the role of stemness in CMT therapies and histogenesis in dogs [16,17]. In addition, several studies have isolated cancer stem cells in different types of canine cancers, including osteosarcoma [18], glioblastoma [19], prostate cancer [20], hepatocellular carcinoma [21], hemangiosarcoma [22], and CMT [23,24]. However, the influence of stemness on different CMT histological and molecular subtypes, such as simple and complex or basal-like and non-basal-like carcinomas, is still a largely unexplored field, and innovative findings could be exploited for the understanding of CMT histogenesis and the development of potential therapeutic approaches.

Here, we applied a one-class logistic regression (OCLR) machine-learning algorithm [16,25] to extract transcriptomic feature sets from a public database and predict stemness in different subtypes of CMT. Our study observed that CMT samples had higher stemness than normal samples, validating our model. Furthermore, simple CMTs exhibited higher stemness than the other histological subtypes. We also confirmed that stemness is higher in basal-like samples and in a "DNA-repair" tumor-specific metagene signature previously determined [12]. Finally, we correlated and suggested two potential targets, *POLA2* and *APEX1*, and target-specific compounds that may inhibit stemness in CMT, especially in the most aggressive subtypes.

2. Results

2.1. Stemness index is higher in malignant CMT in comparison with matched normal samples

In our previous study, we created and validated a canine prediction model for stemness using public data composed of canine non-tumor somatic and pluripotent stem cells. In addition, we applied this stemness model to different canine mammary tumor cell lines established in our laboratory [16]. In the present study, we applied the canine stemness index algorithm to a public RNA-seq dataset
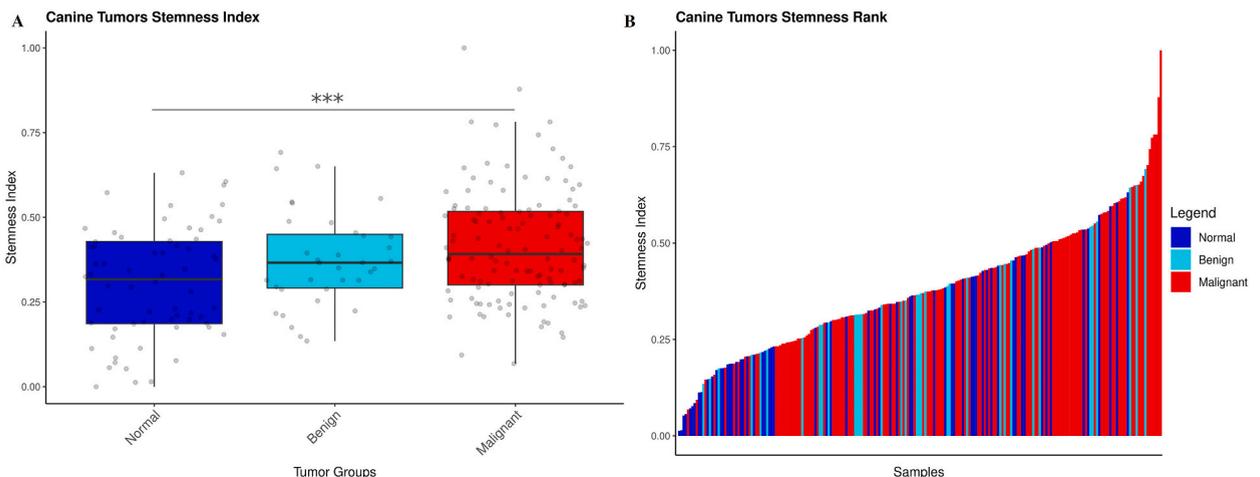
Fig. 1. Stemness is higher in malignant CMTs in comparison with matched normal samples. **(A)** mRNAsi of normal samples and malignant and benign CMTs visualized in a boxplot and stratified by a median. Malignant CMTs exhibited the highest mRNAsi in comparison to pairwise normal samples. *** = p < 0.001 (one-way ANOVA). **(B)** Rank plot corresponding to the mRNAsi predicted to each sample comparing malignant and benign CMTs and normal samples.
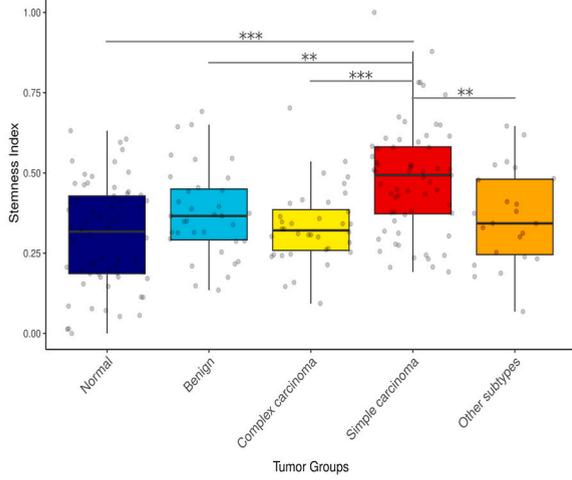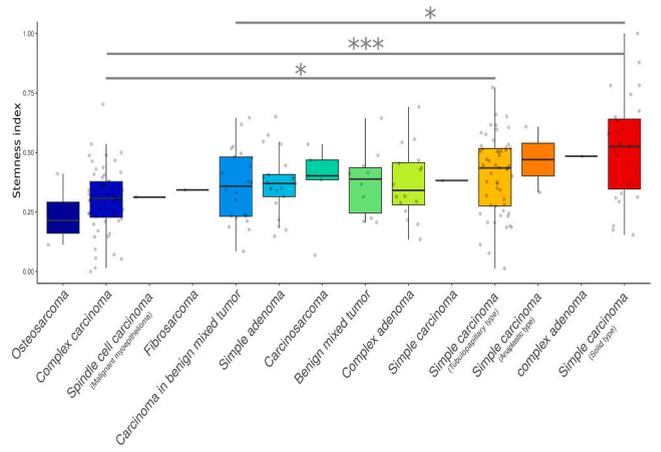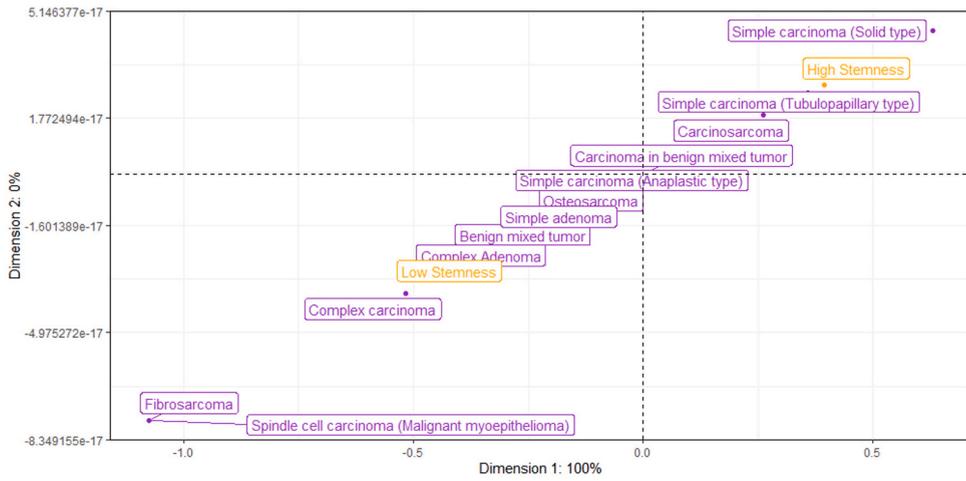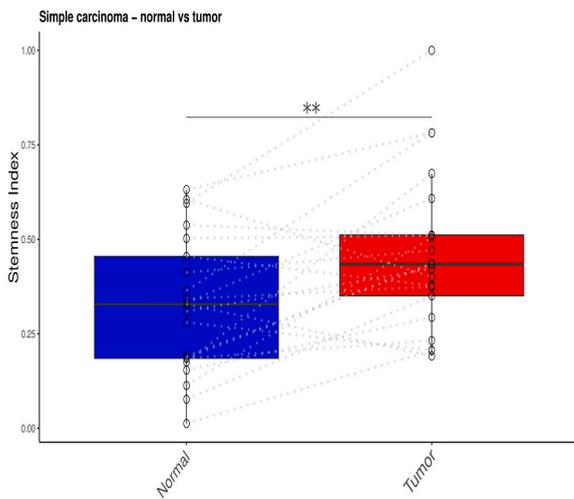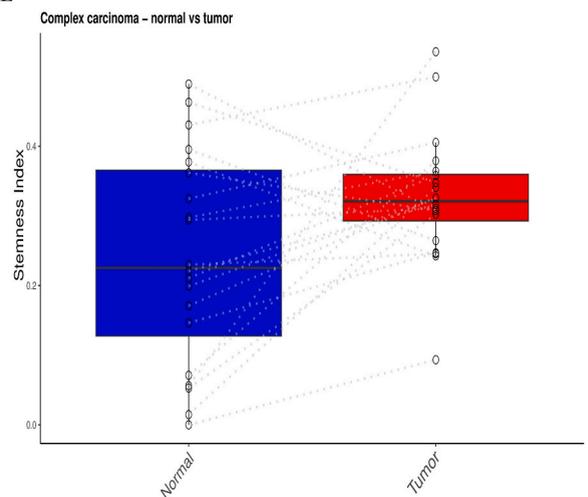
(caption on next page)

Fig. 2. Simple carcinomas exhibit higher stemness than other histologic subtypes. (A) mRNAsi of CMT subtypes visualized in a boxplot and stratified by a median. Simple CMTs exhibited the highest mRNAsi in comparison to normal, benign, complex, and other subtypes of CMT. *** = p < 0.001; ** = p < 0.01 (one-way ANOVA). (B) mRNAsi is higher in solid simple CMTs and tubulopapillary simple CMTs in comparison with other subtypes (one-way ANOVA). (C) A CA two-dimensional perceptual map demonstrating the association between the categories of each categorical variable. Categories in yellow are stemness levels. Categories in purple are CMT Subtypes. Categories that are closely clustered are strongly associated with each other. (D) The pairwise analysis demonstrated that tumor simple CMT exhibits higher stemness than their matched normal samples. ** = p < 0.01 (T-test). (E) The pairwise analysis demonstrated that tumor complex CMT and their matched normal samples exhibit no statistical difference regarding stemness (T-test). (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

comprising CMT and pairwise normal tissues. The dataset was from a cohort containing 157 cases of benign and malignant CMT, and 64 cases of matched normal CMT tissue [12]. As a first step, we rebuilt the prediction model using the overlapping genes between the original stemness prediction model for human [25] and the canine dataset. The efficiency of conversion of canine to human gene ID was approximately 45% (16,405/35,000 transcripts). By comparing the tumor and matched normal sequencing data, we observed that malignant tumors in general presented a higher stemness index (mRNAsi) than normal samples, but not in comparison with benign tumors (Fig. 1A–B, $p < 0.001$). Thus, these results further validated our canine prediction model of stemness in tissue samples and indicated that stemness is a relevant phenotype in malignant CMT.

2.2. Simple carcinomas exhibit higher stemness than other histologic subtypes

Clinically, the presence of myoepithelial cells in complex CMTs is associated with less aggressive biological behavior and better prognosis than simple CMTs [7,9]. To determine whether stemness is related to these clinical phenotypes, we applied mRNAsi in both histological subtypes and others, including carcinoma in benign mixed tumors, carcinosarcoma, fibrosarcoma, osteosarcoma, and spindle cell carcinoma. Interestingly, we observed that simple CMTs presented higher stemness than complex and other subtypes of CMT (Fig. 2A), with both solid and tubulopapillary simple CMTs subtypes exhibiting higher mRNAsi (Fig. 2B). In addition, using correspondence analysis (CA), we demonstrated that high stemness was strongly associated with both solid and tubulopapillary simple carcinomas while low stemness is associated with complex carcinomas (Table 1; Fig. 2C). Finally, pairwise analysis showed that simple carcinomas were the only subtype in which tumor samples exhibited significantly higher mRNAsi levels compared to matched normal tissues (Fig. 2D and E).

Using multiple correspondence analysis (MCA), we also evaluated whether simple and complex carcinomas were associated with clinical prognostic factors available in the database. Simple carcinomas were associated with tumor classification grade 3, negative status for estrogen-receptor expression, PAM50 classification of basal-like tumors, lymphatic invasion, and high stemness predicted with mRNAsi. In contrast, complex carcinoma samples were associated with a positive status for estrogen receptor expression, grade 1, PAM50 classification of non-basal-like tumors, lymphatic invasion, and low stemness predicted with mRNAsi (Table 2; Fig. 3). Thus, these findings suggest that simple carcinomas are associated with poor clinical prognostic factors and that the stemness phenotype may contribute to their malignancy.

2.3. High stemness index is associated with CMT-Basal and NMF2 molecular subtypes

We also investigated whether the stemness phenotype is associated with two different gene expression-based CMT subtypes according to PAM50 and NMF classifications. First, we predicted the mRNAsi of CMT tumors classified as CMT-basal or CMT-non-basal, according to Kim et al. [12]. We found that CMT-basal tumors presented higher stemness than CMT-non-basal tumors and, interestingly, only malignant samples presented this difference, while no significant differences were found in benign samples (Fig. 4A). To further confirm this, we applied simple correspondence analysis and observed that high stemness was strongly associated with

Table 1
Table exhibiting the values of the adjusted standardized residuals. Categories of variables with values higher than 1.96 are considered associated. We could observe a strong association between low stemness and complex CMT (3.552) and between high stemness and solid simple CMT (3.021) and tubulopapillary simple CMT (2.721).

Histological subtypes/Stemness	High Stemness	Low Stemness	Categories are randomly associated?
Benign Mixed Tumor		0.583	No
Carcinoma in Benign Mixed Tumor	0.019		No
Carcinosarcoma	0.459		No
Complex Adenoma		1.144	No
Complex Carcinoma		3.552	Yes
Fibrosarcoma		1.078	No
Osteosarcoma		0.103	No
Simple Adenoma		0.562	No
Simple Carcinoma (Anaplastic)		0.103	No
Simple Carcinoma (Solid)	3.021		Yes
Simple Carcinoma (Tubulopapillary)	2.721		Yes
Spindle Cell Carcinoma		1.078	No

Table 2

Table exhibiting the values of the adjusted standardized residuals. Categories of variables with values higher than 1.96 are associated. We could observe a strong association between simple CMT with poor prognostic variables such as tumor classification grade 3 (3.095), negative status for estrogen-receptor expression (2.55), PAM50 classification of basal-like tumors (2.283), lymphatic invasion (3.128), and high stemness (4.925).

Clinical Prognostic Factors/Histological Subtypes	Complex Carcinoma	Simple Carcinoma	Categories are randomly associated?
Grade 1	2.373		Yes
Grade 2	0.307		No
Grade 3		3.095	Yes
Positive for Estrogen Receptor	2.55		Yes
Negative for Estrogen Receptor		2.55	Yes
Basal-Like Tumor		2.283	Yes
Non-Basal-Like Tumor	2.283		Yes
Positive for Lymphatic Invasion		3.128	Yes
Negative for Lymphatic Invasion	3.128		Yes
High Stemness		4.925	Yes
Low Stemness	4.925		Yes
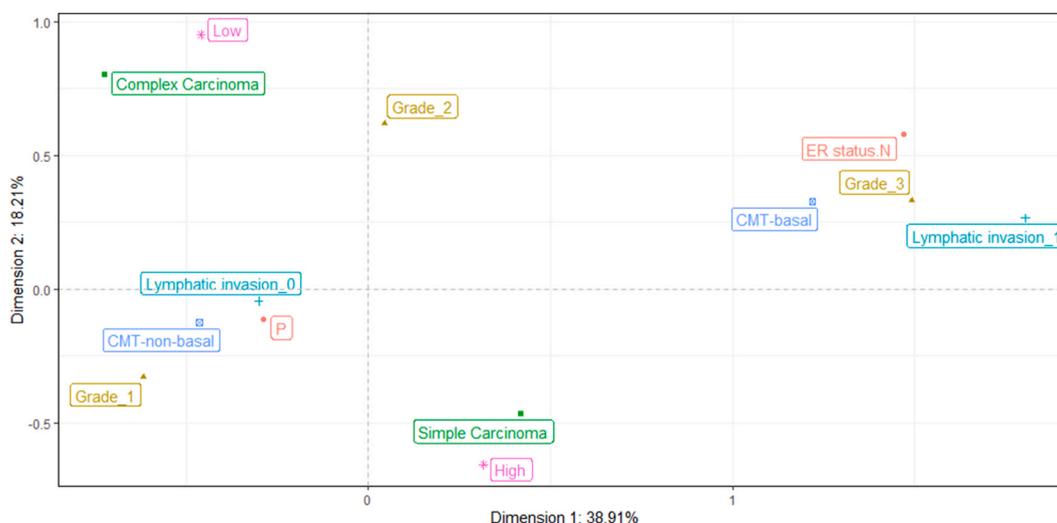
Fig. 3. Simple CMTs are associated with poor prognostic factors. MCA two-dimensional perceptual map demonstrating the association between the categories of each categorical variable. Categories in pink are stemness levels. Categories in brown are tumor classification grades. Categories in red are the status for estrogen-receptor expression. Categories in blue are the PAM50 classification. Categories in light blue are the status of lymphatic invasion. Categories in green are CMT subtypes. Categories that are closely clustered are strongly associated with each other. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

CMT-basal in malignant tumors (Fig. 4B).

Kim et al. identified three gene expression-based CMT subtypes associated with molecular signatures using gene set enrichment analysis. Based on enriched molecular terms, the authors annotated three tumor-specific metagene signatures:1) "mitosis" (NMF1), "DNA repair" (NMF2), and "epithelial-to-mesenchymal transition" (EMT) (NMF3). Here, we assessed whether these metagene signatures were associated with the stemness index. We observed that NMF2 was the metagene signature with the highest stemness, both in malignant and benign tumors (Fig. 4C). We also confirmed these findings using correspondence analysis for malignant tumors (Fig. 4D). Interestingly, 51% of samples classified as NMF2 were simple CMTs (18/35). Although NMF3 CMTs were shown to be a claudin-low subtype, known to be enriched with tumor-initiating cells or stem cells [12], our analysis demonstrated that the metagene signature NMF2 exhibits a higher potential of stemness.

2.4. Correlation analysis identifies potential targets and inhibitors targeting stemness signature

We employed correlation analysis to identify correlations between genes, phenotypes, and compounds/inhibitors and to determine potential markers and targets in CMT, especially in high-stemness simple carcinomas. Firstly, we selected the top 50 genes correlated with higher stemness, and the top 50 genes correlated with lower stemness (Fig. 5) and further performed a gene set enrichment analysis of these genes. Genes correlated with higher stemness were enriched for biological processes, such as synthesis of DNA, cell division, cell cycle, centromere complex assembly, and other terms associated with the cell cycle (Supplemental Table S1). Genes correlated with lower stemness were enriched for biological processes, including the regulation of angiogenesis, regulation of cell migration, kinase activity, and cell adhesion (Supplemental Table S2). In addition, to further investigate other biological processes
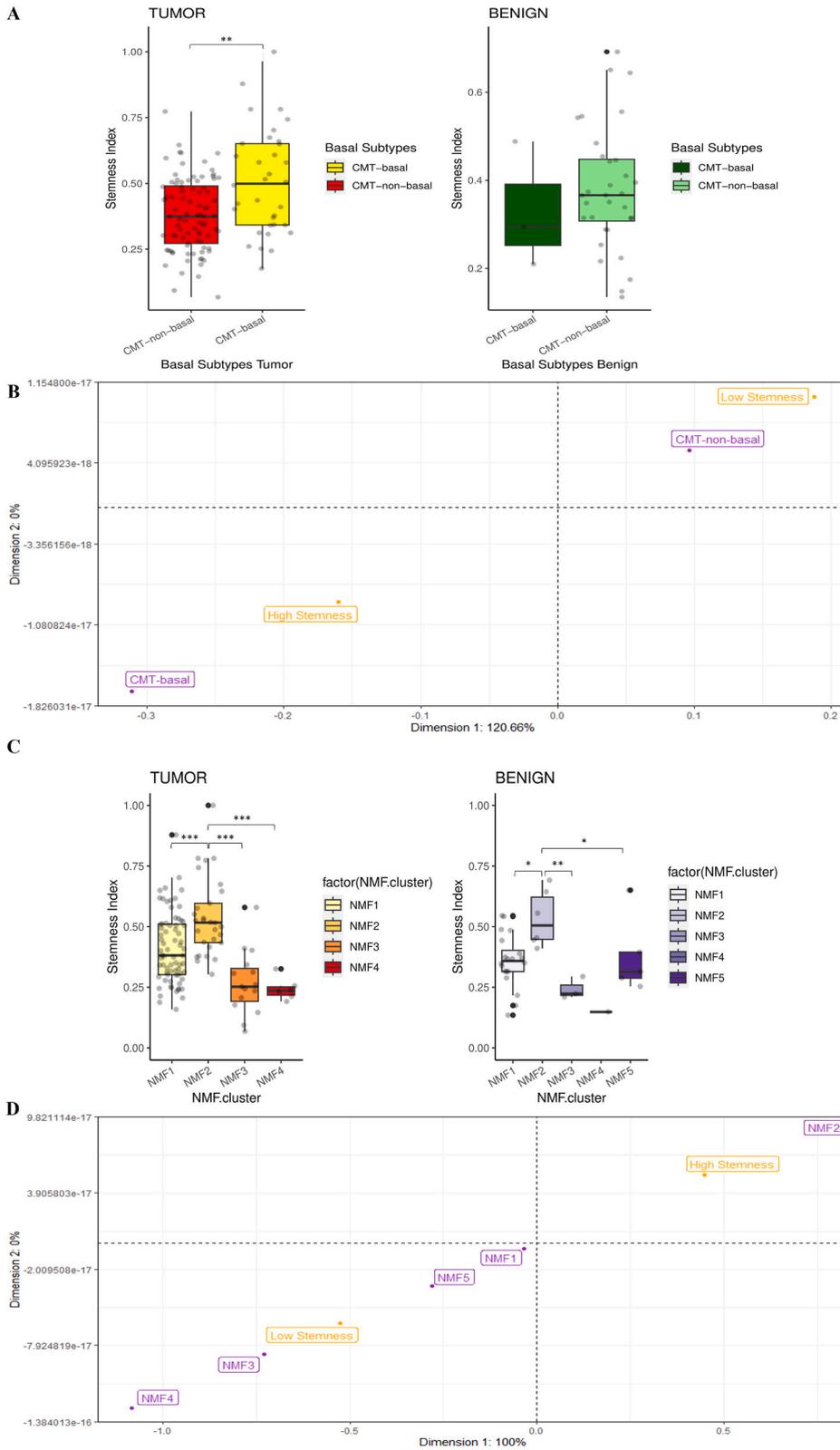
(caption on next page)

Fig. 4. High stemness index is associated with CMT-Basal and NMF2 molecular subtypes. (**A**) mRNAsi visualized in a boxplot and stratified by a median. Basal-like CMTs exhibited the highest mRNAsi in comparison to non-basal subtypes of CMT. ** = p < 0.01 (T-test) (**B**) A CA two-dimensional perceptual map demonstrating the association between the categories of each categorical variable. Categories in yellow are stemness levels. Categories in purple are PAM50 Subtypes. Categories that are closely clustered are strongly associated with each other. (**C**). NMF2 CMTs exhibited the highest mRNAsi in comparison to NMF1, NMF3, and NMF4 subtypes in malignant CMTs. NMF2 exhibited the highest mRNAsi in comparison to NMF1, NMF3, and NMF5 in benign tumors. *** = p < 0.001; ** = p < 0.01; * = p < 0.05 (one-way ANOVA) (**D**) A CA two-dimensional perceptual map demonstrating the association between the categories of each categorical variable. Categories in yellow are stemness levels. Categories in purple are NMF molecular Subtypes. Categories that are closely clustered are strongly associated with each other. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

enriched by genes correlated with high stemness, a gene set enrichment analysis was performed for the top 1000 correlated genes. We also observed genes enriched for DNA repair, mismatch repair, Double-strand break repair via break-induced replication, Base excision repair, and Nucleotide excision repair (Supplemental Table S3).

Two stemness-associated genes, *POLA2* and *APEX1*, are druggable targets for Dacarbazine and Lucanthone small inhibitors, respectively. Dacarbazine is an antineoplastic agent used to treat malignant melanoma and Hodgkin disease in humans [27–30]. Lucanthone inhibits post-radiation DNA repair in tumor cells, especially in brain cancer. Furthermore, Lucanthone has been shown to inhibit stemness in glioma [31]. An anticancer small molecule, APX-3330, also reduced colon cancer stem cell (CCSCs) growth *in vitro*. In addition, in xenograft mice injected subcutaneously with CCSCs, intratumoral administration of APX-3330 increased the tumor response to 5-Fluorouracil delivered intraperitoneally [32].

We also demonstrated that *POLA2* is highly expressed in simple CMTs compared to normal samples (Fig. 6A) and in the high-stemness NMF2 metagene signature (Fig. 6B). Furthermore, *APEX1* was highly expressed in simple carcinomas and other malignant subtypes in comparison with normal samples (Fig. 6C) and the NMF2 metagene signature (Fig. 6D) demonstrating that both genes may be potential therapeutic targets in simple carcinomas and in high-stemness NMF2 samples.

3. Discussion

Stemness is a phenotype associated with cancer stem cells and has been implicated in cancer progression, malignancy, and tumor resistance. However, its roles in canine mammary carcinomas and their different subtypes, such as simple and complex, are poorly defined, partly owing to the lack of models to evaluate and measure stemness. Here, using a one-class logistic regression (OCLR) machine-learning algorithm to predict cancer stemness and generate a stemness index (mRNAsi) in canine cancer [16,25], we found that stemness increased in malignant CMT samples in comparison with normal samples. In addition, we observed that stemness was associated with simple CMT and not with complex CMT, in addition to being associated with basal-like CMTs and the tumor-specific DNA repair metagene signature [12].

Despite the isolation of cancer stem cells from canine tumors in many studies [18–24], further investigation is needed to determine the key role of stemness in canine tumors. Advancements in omics-based technologies and databases have enabled comprehensive phenotypic and molecular studies of canine species [11–13]. Using these public data, we could predict stemness in CMT and normal samples, suggesting interesting influences of stemness, especially in simple CMT. Simple CMT exhibit proliferation of only one cell population, resembling either luminal epithelial or myoepithelial cells, and have a worse prognosis than complex CMTs, but there is little evidence explaining this fact. Simple CMTs can faithfully recapitulate breast cancer in women and are closely clustered with basal-like human breast cancer in the PAM50 classification, suggesting a reason for their malignancy [11,33,34]. Our results also provide evidence to explain the malignancy of simple CMTs based on high stemness and, possibly, the enrichment of cancer stem cells in these tumors. To date, no studies have demonstrated that canine simple CMT is enriched for cancer stem cells in comparison with other subtypes. However, several studies on human breast cancer have demonstrated an association between basal-like tumors, stemness, and cancer stem cells [35–37], contributing to the poor prognosis associated with this subtype [38], although the presence of cancer stem cells is not confined to basal-like tumors [39]. In addition, the stemness index has already been applied to demonstrate that stemness is higher in basal-like breast cancer in humans [25], which is consistent with our study showing that basal-like CMTs exhibited and were associated with higher levels of stemness. To the best of our knowledge, this is the first study to measure molecular CMTs datasets to determine the association between higher stemness and basal-like tumors. Thus, our findings provide the first evidence that stemness may play a key role in simple CMTs. However, further well-designed studies involving the isolation and enrichment of cancer stem cells and validation of stemness markers in simple CMTs are needed to test this hypothesis.

Kim et al. (2020). found that basal-like CMTs were enriched with the NMF3 metagene signature, exhibiting downregulation of claudin-encoding genes and E-cadherin [12], phenotypically similar to the low claudin breast cancer subtype, which is enriched with tumor-initiating cells and stem cells [40]. However, in our study, we did not find an association between high stemness and the NMF3 metagene signature. Instead, the NMF2 subtype, considered a "DNA-repair" subtype, showed a higher mRNAsi and strong association with the stemness phenotype, opening new avenues for discussion and investigation. DNA damage and repair mechanisms play key roles in high stemness cells, such as stem cells and cancer stem cells. Both stem cells and cancer stem cells are endowed with effective DNA repair mechanisms to prevent the accumulation and propagation of genetic damage to daughter cells and to develop mechanisms of resistance to chemotherapy and radiotherapy, respectively [41]. This can explain the enrichment of the NMF2 subtype with high-stemness tumors in our study. In addition, a limitation of our study is that we could not understand the association between higher stemness, NMF metagene signatures, and histological types, since simple CMTs exhibiting higher stemness were more frequently classified in the NMF1 metagene signature [12]. However, how stemness and the NMF metagene signatures correlate with
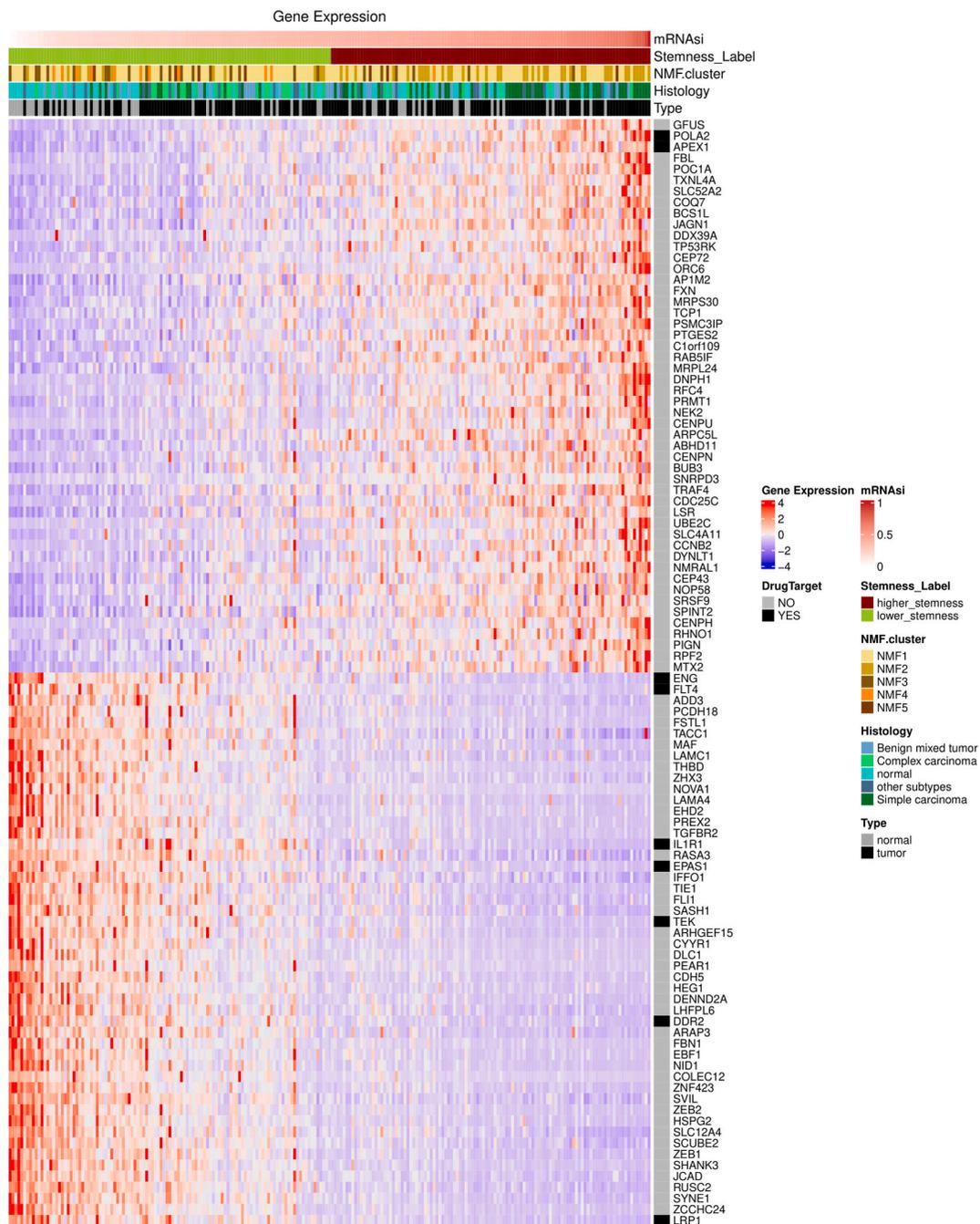
Fig. 5. Correlation analysis identifies potential targets and inhibitors targeting stemness signature. Heatmap of canine mammary carcinoma gene expression data. Samples were sorted in ascending order (from left to right) according to mRNAsi. Above is the annotation of other parameters such as the categorical Stemness Index (high or low), NMF cluster designated for that sample, histological subtype of the tumor, and indication of tumor or normal sample. On the right are listed the 100 genes (50 with the highest positive correlation and 50 with the highest negative correlation) related to the mRNAsi which are anticancer drug targets according to the Open Targets platform.

therapy resistance in CMTs is open for future investigations.

Using a correlation analysis, we identified two possible new druggable targets for CMT. Both *POLA2* and *APEX1* were highly correlated with stemness. In addition, *POLA2* was highly expressed in simple CMTs than in other subtypes and normal samples, whereas *APEX1* expression was increased in general CMTs in comparison with normal samples. These translational and comparative analyses may contribute to drug repositioning and repurposing in CMTs, particularly targeting the stemness phenotype [42].

Our study demonstrates some limitations. We applied a machine learning algorithm to demonstrate stemness as a potential
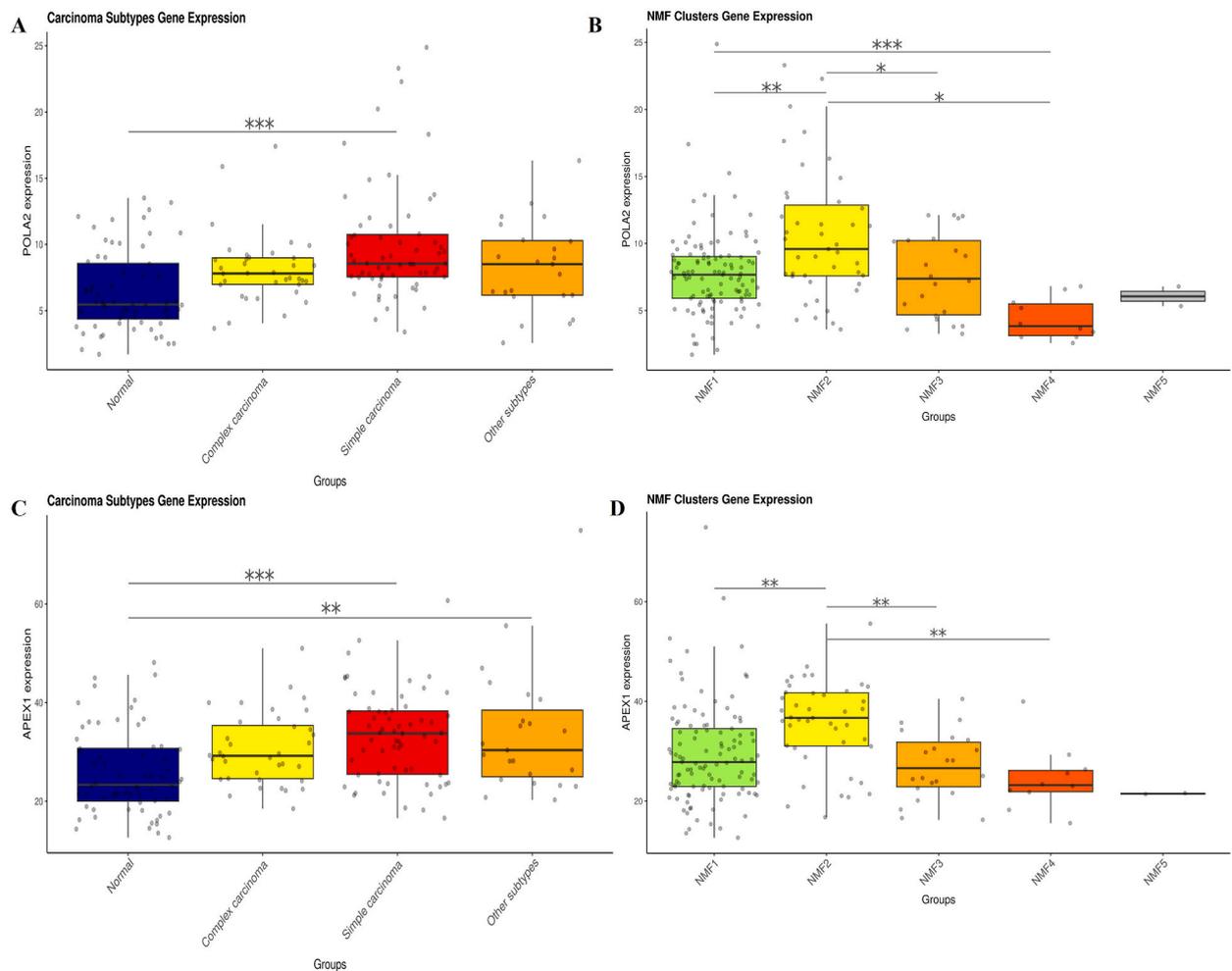
Fig. 6. *POLA2* and *APEX1* expression are higher in simple CMTs and NMF2 metagene signature. (**A**) mRNAsi visualized in a boxplot and stratified by a median. Simple CMTs showed higher *POLA2* gene expression in comparison with normal samples. (**B**) mRNAsi visualized in a boxplot and stratified by a median. *POLA2* expression was higher in the NMF2 metagene molecular subtype. (**C**) mRNAsi visualized in a boxplot and stratified by a median. *APEX1* was higher expressed in simple CMTs and other malignant subtypes in comparison with normal samples. (**D**) *APEX1* expression was higher in the NMF2 molecular subtype. *** = p < 0.001; ** = p < 0.001; * = p < 0.05 (one-way ANOVA).

mechanism behind the development and aggressiveness of CMTs, especially in simple carcinomas. However, a deeper exploration using *in vitro* and *in vivo* approaches is needed to confirm, for example, the enrichment of cancer stem cells and stemness markers in these tumors. Other assays including immunohistochemistry and flow cytometry, using specific stemness and cancer stem cell markers, may further confirm our computational findings. Also, *in vitro* and *in vivo* models will be fundamental to validate *POLA2* and *APEX1* as druggable targets, enhancing drug repositioning and repurposing for CMTs. Overall, this initial computational study provides substantial evidence that stemness is an important phenotype in CMTs development and aggressiveness and deeper studies will be performed to validate these findings.

In conclusion, exploiting the stemness phenotype in CMTs using omics and machine learning approaches holds promise for understanding the development of this heterogeneous disease and its subtypes, and for identifying novel therapeutic targets. Our findings support a role for stemness in behind the aggressiveness and development of CMTs, especially the simple and basal-like carcinomas. This corroborates studies in human breast cancer, exhibiting a strong influence of stemness in more aggressive basal-like tumors. Thus, our results suggest and reinforces stemness is an important phenotype for the development of mammary tumors, highlighting the validity of canine models for comparative and translational studies.

4. Materials and methods

4.1. RNA expression data

RNA-seq data of canine mammary tumors were downloaded from the GEO database (GSE119810) [12]. The database contains 157

tumor tissues and 64 matched, normal, and adjacent tissues. The original samples were classified according to grade, lymphatic invasion, ER status, HER2 score, Histology, NMF cluster, and PAM50 subtypes.

4.2. Stemness index model

To predict the stemness phenotype in canine samples, we applied a machine learning algorithm described by Malta et al. (2018) to generate a one-class logistic regression (OCLR) algorithm employing human stem cell (iPS and ES) data from the Progenitor Cell Biology Consortium (PCBC) to estimate the stemness (mRNAsi) index in human tumors. The stem cell transcriptomic matrix used to train OCLR contained 12,945 gene expression values. Furthermore, this model was validated using leave-one-out cross-validation, which leaves one stem cell out of training, and then applies the model to non-SC cells (their differentiated ecto-, meso-, and endoderm progenitors). Moreover, Malta's model presented a high correlation among other cancer stem cell signatures, providing significant insights into the biological and clinical features of pan-cancer. The workflow to generate the mRNAsi is available at https://github.com/ArtemSokolov/PanCanStem.

4.3. Stemness index prediction

First, we converted the gene IDs from *Canis lupus* familiaris to *Homo sapiens* using tables of Ensembl data downloaded via the BioMart mining tool (https://www.ensembl.org/biomart/martview/). We then mapped ENSEMBL gene ID to gene symbols using the org.Hs.eg.db package from the Bioconductor repository. Genes with no such intersections were excluded. The previously described prediction stemness index model derived from pluripotent cells was applied to the gene expression profiles of canine and human mammary cancer cell lines, leading to their mRNAsi. The indices were scaled to 0–1 (0 being a low stemness score and 1 being a high score) by subtracting the minimum and dividing by the maximum. The top 439 genes of the stemness signature (439 genes) were selected based on the Euclidean distance of the absolute weight of the stemness signature model. From the 12,953 genes in the original model, we applied 11,251 genes owing to the conversion gene IDs output. In relation to stemness prediction in human cell lines and canine stem cells, we applied 12,257 and 7929 genes, respectively, as a result of the number of genes in common among RNA-seq from GEO datasets.

4.4. Correspondence analysis

Simple (CA) and multiple (MCA) correspondence analyses were performed in the RStudio 4.3.1 environment using the packages FactoMineR [43] and cabootcrs (https://CRAN.R-project.org/package=cabootcrs). Contingency tables of the categorical variables were created, and to study the associations between variables, a chi-square test with k degrees of freedom, followed by the analysis of the adjusted standardized residuals, was performed. The Chi-squared test evaluates whether the associations between categorical variables are randomly associated. Adjusted standardized residuals higher than 1.96 indicate an association between variables in the matrix. To perform CA, the categorical variables should not be randomly associated. To create the perceptual map, inertia was determined as the total chi-squared divided by the number of samples, resulting in the number of associations in the dataset. MCA was performed based on the binary and Burt Matrices. Row and column profiles were determined to demonstrate the influence of each category of variables on the others. Matrices were defined based on the row and column profiles. Eigenvalues were then extracted to represent the number of dimensions that could be captured in the analysis. Finally, the x- and y-axis coordinates of the perceptual map were determined, allowing the category of the variables to be represented and established.

4.5. Correlation analysis

To identify genes directly related to the Stemness Index obtained in carcinoma and normal samples, a linear correlation analysis was performed between the Stemness Index and the expression of each gene individually, using the cor.test function of the stats package. Spearman method was applied considering results with "p.value < 0.05" and correlation "-0.5$>$correlation$>$0.5". The 50 genes with the highest positive correlation and the 50 genes with the lowest negative correlation between the Stemness Index and the expression values are indicated in the heatmap.

4.6. Statistical analysis

The stemness Index was compared between multiple groups using one-way ANOVA with post-hoc Tukey's test. An unpaired T-test was used to compare the stemness index between the two groups. Correlation analysis was performed using the Spearman correlation coefficient. Differences were considered significant when the $p < 0.05$.

Data availability statement

All raw and processed RNA sequencing data used in this study was submitted to the NCBI Gene Expression Omnibus (GEO) under accession number GSE119810 [12]. The workflow to generate the mRNAsi is available at https://github.com/ArtemSokolov/PanCanStem.

Funding

This study has been supported by grants from the Sao Paulo Research Foundation (FAPESP), Brazil (2018/00583-0, 2019/14928-1, 2021/00283-9, 2022/06305-7, 2022/09378-5 2023/05099-7, 2023/07358-0).

CRediT authorship contribution statement

Pedro L.P. Xavier: Writing – original draft, Validation, Investigation, Formal analysis, Conceptualization. **Maycon Marção:** Writing – original draft, Visualization, Methodology, Investigation, Formal analysis, Conceptualization. **Renan L.S. Simões:** Visualization, Methodology, Formal analysis, Conceptualization. **Maria Eduarda G. Job:** Formal analysis. **Ricardo de Francisco Strefezzi:** Writing – review & editing, Supervision, Formal analysis. **Heidge Fukumasu:** Writing – review & editing, Supervision, Resources, Funding acquisition. **Tathiane M. Malta:** Writing – review & editing, Validation, Supervision, Resources, Project administration, Investigation, Funding acquisition, Formal analysis, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.heliyon.2024.e26714.

References

[1] S.M. Abdelmegeed, S. Mohammed, Canine mammary tumors as a model for human disease, Oncol. Lett. 15 (2018) 8195–8205, https://doi.org/10.3892/ol.2018.8411 (Review).
[2] M. Gray, J. Meehan, C. Martínez-Pérez, C. Kay, A.K. Turnbull, L.R. Morrison, L.Y. Pang, D. Argyle, Naturally-occurring canine mammary tumors as a translational model for human breast cancer, Front. Oncol. 10 (2020) 617, https://doi.org/10.3389/fonc.2020.00617.
[3] P.J. Rivera, M. Melin, T. Biagi, T. Fall, J. Häggström, K. Lindblad-Toh, H. Von Euler, Mammary tumor development in dogs is associated with BRCA1 and BRCA2, Cancer Res. 69 (2009) 8770–8774, https://doi.org/10.1158/0008-5472.CAN-09-1725.
[4] K.S. Borge, S. Nord, P. Van Loo, O.C. Lingjærde, G. Gunnes, G.I.G. Alnæs, H.K. Solvang, T. Lüders, V.N. Kristensen, A.L. Børresen-Dale, et al., Canine mammary tumours are affected by frequent copy number aberrations, Including amplification of myc and loss of pten, PLoS One 10 (2015) 1–21, https://doi.org/10.1371/journal.pone.0126371.
[5] Y. Salas, A. Márquez, D. Diaz, L. Romero, Epidemiological study of mammary tumors in female dogs diagnosed during the period 2002-2012: a growing animal health problem, PLoS One 10 (2015) 1–15, https://doi.org/10.1371/journal.pone.0127381.
[6] G.D. Cassali, H. Gobbi, C. Malm, F.C. Schmitt, Evaluation of accuracy of fine needle aspiration cytology for diagnosis of canine mammary tumours: comparative features with human tumours, Cytopathology 18 (2007) 191–196, https://doi.org/10.1111/j.1365-2303.2007.00412.x.
[7] M.H. Goldschmidt, L. Peña, R. Rasotto, V. Zappulli, Classification and grading of canine mammary tumors, Vet. Pathol. 48 (2011) 117–131, https://doi.org/10.1177/0300985810393258.
[8] G.D. Cassali, G.E. Lavalle, A.B. de Nardi, E. Ferreira, A.C. Bertagnolli, A. Estrela-Lima, A.C. Alessi, C.R. Daleck, B.S. Salgado, C.G. Fernandes, et al., Consensus for the diagnosis, prognosis and treatment of canine mammary tumors, Brazilian J. Vet. Pathol. 4 (2011).
[9] M. Karayannopoulou, E. Kaldrymidou, T.C. Constantinidis, A. Dessiris, Histological grading and prognosis in dogs with mammary carcinomas: application of a human grading method, J. Comp. Pathol. 133 (2005) 246–252, https://doi.org/10.1016/j.jcpa.2005.05.003.
[10] R. Rasotto, D. Berlato, M.H. Goldschmidt, V. Zappulli, Prognostic significance of canine mammary tumor histologic subtypes: an observational cohort study of 229 cases, Vet. Pathol. 54 (2017) 571–578, https://doi.org/10.1177/0300985817698208.
[11] D. Liu, H. Xiong, A.E. Ellis, N.C. Northrup, C.O. Rodriguez, R.M. O'Regan, S. Dalton, S. Zhao, Molecular homology and difference between spontaneous canine mammary cancer and human breast cancer, Cancer Res. 74 (2014) 5045–5056, https://doi.org/10.1158/0008-5472.CAN-14-0392.
[12] T.M. Kim, I.S. Yang, B.J. Seung, S. Lee, D. Kim, Y.J. Ha, M. kyoung Seo, K.K. Kim, H.S. Kim, J.H. Cheong, et al., Cross-species oncogenic signatures of breast cancer in canine mammary tumors, Nat. Commun. 11 (2020), https://doi.org/10.1038/s41467-020-17458-0.
[13] K. Graim, D. Gorenshteyn, D.G. Robinson, N.J. Carriero, J.A. Cahill, R. Chakrabarti, M.H. Goldschmidt, A.C. Durham, J. Funk, J.D. Storey, et al., Modeling molecular development of breast cancer in canine mammary tumors, Genome Res. 31 (2021) 337–347, https://doi.org/10.1101/GR.256388.119.
[14] J.D. Lathia, H. Liu, Overview of cancer stem cells and stemness for community oncologists, Targeted Oncol. 12 (2017) 387–399, https://doi.org/10.1007/s11523-017-0508-3.
[15] A.Z. Ayob, T.S. Ramasamy, Cancer stem cells as key drivers of tumour progression, J. Biomed. Sci. 25 (2018) 1–18, https://doi.org/10.1186/s12929-018-0426-4.
[16] M. Marção, S. Müller, P.L.P. Xavier, T.M. Malta, Stemness inhibition by (+)-JQ1 in canine and human mammary cancer cells revealed by machine learning, Front. Drug Discov 2 (2022) 1–11, https://doi.org/10.3389/fddsv.2022.953988.
[17] M. Michishita, Understanding of tumourigenesis in canine mammary tumours based on cancer stem cell research, Vet. J. 265 (2020) 105560, https://doi.org/10.1016/j.tvjl.2020.105560.
[18] Wilson, H., Huelsmeyer, M., Chun, R., Young, K.M., Friedrichs, K., and Argyle, D.J. Isolation and characterisation of cancer stem cells from canine osteosarcoma. Vet. J. 208. 175, 69–75. 10.1016/j.tvjl2007.07.025..
[19] G. Stoica, G. Lungu, H. Martini-Stoica, S. Waghela, J. Levine, R. Smith, Identification of cancer stem cells in dog glioblastoma, Vet. Pathol. 46 (2009) 391–406, 10.1354/vp.08-VP-0218-S-FL.
[20] M. Moulay, W. Liu, S. Willenbrock, K.A. Sterenczak, R. Carlson, A. Ngezahayo, H.M. Escobar, I. Nolte, Evaluation of stem cell marker gene expression in canine prostate carcinoma- and prostate cyst-derived cell lines, Anticancer Res. 33 (2013) 5421–5431.
[21] M. Michishita, S. Ezaki, K. Ogihara, Y. Naya, D. Azakami, T. Nakagawa, N. Sasaki, T. Arai, T. Shida, K. Takahashi, Identification of tumor-initiating cells in a canine hepatocellular carcinoma cell line, Res. Vet. Sci. 96 (2014) 315–322, https://doi.org/10.1016/j.rvsc.2014.01.004.

[22] K. Aoshima, Y. Fukui, K.C.M. Gulay, O. Erdemsurakh, A. Morita, A. Kobayashi, T. Kimura, Notch2 signal is required for the maintenance of canine hemangiosarcoma cancer stem cell-like cells, BMC Vet. Res. 14 (2018) 1–16, https://doi.org/10.1186/s12917-018-1624-8.

[23] M. Michishita, R. Akiyoshi, H. Suemizu, T. Nakagawa, N. Sasaki, H. Takemitsu, T. Arai, K. Takahashi, Aldehyde dehydrogenase activity in cancer stem cells from canine mammary carcinoma cell lines, Vet. J. 193 (2012) 508–513, https://doi.org/10.1016/j.tvjl.2012.01.006.

[24] C. Cocola, P. Anastasi, S. Astigiano, E. Piscitelli, P. Pelucchi, L. Vilardo, G. Bertoli, M. Beccaglia, M.C. Veronesi, S. Sanzone, et al., Isolation of canine mammary cells with stem cell properties and tumour-initiating potential, Reprod. Domest. Anim. 44 (2009) 214–217, https://doi.org/10.1111/j.1439-0531.2009.01413.x.

[25] T.M. Malta, A. Sokolov, A.J. Gentles, T. Burzykowski, L. Poisson, J.N. Weinstein, B. Kamińska, J. Huelsken, L. Omberg, O. Gevaert, et al., Machine learning identifies stemness features associated with oncogenic dedifferentiation, Cell 173 (2018) 338–354.e15, https://doi.org/10.1016/j.cell.2018.03.034.

[27] P.B. Chapman, L.H. Einhorn, M.L. Meyers, S. Saxman, A.N. Destro, K.S. Panageas, C.B. Begg, S.S. Agarwala, L.M. Schuchter, M.S. Ernstoff, et al., Phase III multicenter randomized trial of the dartmouth regimen versus dacarbazine in patients with metastatic melanoma, J. Clin. Oncol. 17 (1999) 2745, https://doi.org/10.1200/JCO.1999.17.9.2745.

[28] A.A. Tarhini, S.S.A. Garwala, Cutaneous melanoma: available therapy for metastatic disease, Dermatol. Ther. 19 (2006) 19–25, https://doi.org/10.1111/j.1529-8019.2005.00052.x.

[29] P. Torka, E. Przespolewski, A.M. Evens, Treatment strategies for advanced classical Hodgkin lymphoma in the times of dacarbazine shortage, JCO Oncol. Pract. 18 (2022) 491–497, https://doi.org/10.1200/OP.21.00890.

[30] J.S. Abramson, E. Bengston, R. Redd, J.A. Barnes, T. Takvorian, L. Sokol, F. Lansigan, P. Armand, B. Shah, E. Jacobsen, et al., Brentuximab vedotin plus doxorubicin and dacarbazine in nonbulky limited-stage classical Hodgkin lymphoma, Blood Adv 7 (2023) 1130–1136, https://doi.org/10.1182/bloodadvances.2022008420.

[31] D.P. Radin, G. Smith, V. Moushiaveshi, A. Wolf, R. Bases, S.E. Tsirka, Lucanthone targets lysosomes to Perturb glioma proliferation, Chemoresistance and stemness, and Slows tumor growth in vivo, Front. Oncol. 12 (2022) 1–14, https://doi.org/10.3389/fonc.2022.852940.

[32] D. Lou, L. Zhu, H. Ding, H.Y. Dai, G.M. Zou, Aberrant expression of redox protein Ape1 in colon cancer stem cells, Oncol. Lett. 7 (2014) 1078–1082, https://doi.org/10.3892/ol.2014.1864.

[33] H. Bergholtz, T. Lien, F. Lingaas, T. Sørlie, Comparative analysis of the molecular subtype landscape in canine and human mammary gland tumors, J. Mammary Gland Biol. Neoplasia 27 (2022) 171–183, https://doi.org/10.1007/s10911-022-09523-9.

[34] J. Watson, T. Wang, K.-L. Ho, Y. Feng, K.K. Dobbin, Human basal-like breast cancer is represented by one of the two mammary tumor subtypes in dogs, Breast Cancer Res. 25 (2023), https://doi.org/10.1186/s13058-023-01705-5.

[35] T. Kim, S.J. Yang, D. Hwang, J. Song, M. Kim, S. Kyum Kim, K. Kang, J. Ahn, D. Lee, M.Y. Kim, et al., A basal-like breast cancer-specific role for SRF-IL6 in YAP-induced cancer stemness, Nat. Commun. 6 (2015), https://doi.org/10.1038/ncomms10186.

[36] G. Honeth, P.O. Bendahl, M. Ringnér, L.H. Saal, S.K. Gruvberger-Saal, K. Lövgren, D. Grabau, M. Fernö, Å. Borg, C. Hegardt, The CD44+/CD24-phenotype is enriched in basal-like breast tumors, Breast Cancer Res. 10 (2008) 1–12, https://doi.org/10.1186/bcr2108.

[37] J. Zhou, Q. Chen, Y. Zou, H. Chen, L. Qi, Y. Chen, Stem cells and cellular origins of breast cancer: updates in the rationale, controversies, and therapeutic implications, Front. Oncol. 9 (2019) 1–12, https://doi.org/10.3389/fonc.2019.00820.

[38] K. Da Yu, R. Zhu, M. Zhan, A.A. Rodriguez, W. Yang, S. Wong, A. Makris, B.D. Lehmann, X. Chen, I. Mayer, et al., Identification of prognosis-relevant subgroups in patients with chemoresistant triple-negative breast cancer, Clin. Cancer Res. 19 (2013) 2723–2733, https://doi.org/10.1158/1078-0432.CCR-12-2986.

[39] J. Kim, R. Villadsen, T. Sørlie, L. Fogh, S.Z. Grønlund, A.J. Fridriksdottir, I. Kuhn, F. Rank, V.T. Wielenga, H. Solvang, et al., Tumor initiating but differentiated luminal-like breast cancer cells are highly invasive in the absence of basal-like activity, Proc. Natl. Acad. Sci. U.S.A. 109 (2012) 6124–6129, https://doi.org/10.1073/pnas.1203203109.

[40] A. Prat, J.S. Parker, O. Karginova, C. Fan, C. Livasy, J.I. Herschkowitz, X. He, C.M. Perou, Phenotypic and molecular characterization of the claudin-low intrinsic subtype of breast cancer, Breast Cancer Res. 12 (2010), https://doi.org/10.1186/bcr2635.

[41] I. Vitale, G. Manic, R. De Maria, G. Kroemer, L. Galluzzi, DNA damage in stem cells, Mol. Cell. 66 (2017) 306–319, https://doi.org/10.1016/j.molcel.2017.04.006.

[42] A. Giuliano, R.S. Horta, R.A.M. Vieira, K.R. Hume, J. Dobson, Repurposing drugs in small animal oncology, Animals 13 (2023) 1–22, https://doi.org/10.3390/ani13010139.

[43] S. Le, J. Josse, F.FactoMineR. Husson, An R package for multivariate analysis, J. Stat. Software 25 (2008), https://doi.org/10.18637/jss.v025.i01.