

Yale University

EliScholar – A Digital Platform for Scholarly Publishing at Yale

Public Health Theses

School of Public Health

January 2022

A Nonsense Mutation In Gsdmc Is Associated With Cancer Mortality In Dogs

Hang Li
dfjrsayf@gmail.com

Follow this and additional works at: <https://elischolar.library.yale.edu/ysphtdl>

Recommended Citation

Li, Hang, "A Nonsense Mutation In Gsdmc Is Associated With Cancer Mortality In Dogs" (2022). *Public Health Theses*. 2175.

<https://elischolar.library.yale.edu/ysphtdl/2175>

This Open Access Thesis is brought to you for free and open access by the School of Public Health at EliScholar – A Digital Platform for Scholarly Publishing at Yale. It has been accepted for inclusion in Public Health Theses by an authorized administrator of EliScholar – A Digital Platform for Scholarly Publishing at Yale. For more information, please contact elischolar@yale.edu.

Title: A nonsense mutation in GSDMC is associated with cancer mortality in dogs

Name: Hang Li

Year Completed: 2022

Year Degree Awarded: 2022

Degree Awarded: Master of Public Health

Department: Should be School of Public Health

Advisor/Committee Chair: Josephine Hoh

Committee Members: Andrew Dewan

Abstract

Over 400 distinct canine breeds today were result from artificial selection during the last centuries. And due to the selection pressure from human interest and the founder effects, the difference among individual genomes from one same breed is less than the difference between breeds. 446 samples were included in this study to recognize the cancer-associated nonsense mutations in canine genome. 17 of the 446 samples were lab-recruited dogs, and the other 429 samples were downloaded from Sequence Read Archive. One nonsense mutation in GSDMC protein is found significantly associated with cancer risk in dogs. Since nonsense mutations could provide premature proteins or shortened mRNA lifespan, given that the potential function of GSDMC in cell death pathways, our study results provide clues for pathological function of GSDMC in cancer.

Table of Contents

Introduction	4
Background.....	5
Research Design	5
Whole genome sequencing.....	5
Hypothesis generation.....	6
Result	6
Discussion.....	8
Position of the nonsense mutation in GSDMC	8
The potential function of GSDMC	9
Strength of the study	10
Limitation of the study.....	10
Conclusion	11

List of Tables

- Table 1 Ratio of death caused by cancer
- Table 2 Mutation with top prevalence difference
- Table 3 Validation of the mutations

List of Figures

- Fig 1 Position of the nonsense mutation in GSDMC DNA sequence
- Fig 2 Position of the nonsense mutation in GSDMC transcript

Introduction

The overall aim of this study is to recognize disease-associated gene mutations among the next-generation sequencing (NGS) data from canine models. Specifically, this study aims to find homozygous nonsense mutations in dogs related to higher cancer risk, including the hypothesis generation and validation steps.

Canine is an excellent model for human cancer genetic research. Though the domestication of dog took place at around 10,000 years BP, but over 400 distinct breeds today were result from artificial selection during the last centuries.¹ As a result of deliberate breeding over the hundreds of years, many canine breeds have been established. Due to the selection pressure from human interest and the founder effects, the difference among individual genomes from one same breed is less than the difference between breeds.² In addition to the isolated breeding population structure, dog, as a large animal, is more similar to human in the degree of life span and physiology comparing to mouse.³ Spontaneous neoplasia was observed in dogs like in human and many genes related with canine cancer are orthologs of human, according to Online Mendelian Inheritance in Animals (OMIA).⁴ Comparing with mouse genome, canine genome has less differences with human genome, thus providing a more precise model.⁵ Besides, dogs share similar environment with owners, which allows them serve as animal model for further gene environment interaction research.

This study focused on the nonsense mutations to enhance the significance. The nonsense mutations, also known as stop-gained mutations, refer to point mutations resulting in an early presented stop codon in the transcribed mRNA sequences, which could finally provide premature proteins. Nonsense mutations could also cause shortened mRNA lifespan through activating nonsense-mediated mRNA decay (NMD), which leads to difficulty in producing truncated proteins.⁶ The nonsense mutation is one of the most severe mutation types and generate loss of function mutation, and nonsense mutations were proved to be related with multiple chronic diseases in human, such as Cystic fibrosis⁷, Beta thalassaemia⁸, and multiple cancers.⁹

The breed-specific homozygous nonsense mutations are investigated. If one homozygous nonsense mutation takes a much larger fraction of the population in breeds with high risks of cancer compared to the breeds with lower risks, the mutation is probably responsible for cancer development. Based on this logic, homozygous stop-gained mutations which have the largest fraction difference between the two groups could provide valuable cues for hypothesis development.

Background

Previous population studies showed that cancer was the most common canine causes of death, around four million dogs were affected per year.^{10,11} In the cohort study by J.M. Fleming et al., 74,556 dogs in North America were followed for the research of breed-related causes of death. And the death causes of 15,881 samples were collected by V. J. ADAMS et al., in a health survey of purebred dogs in the UK. Both studies reported proportion of cancer caused death to vary among different breeds. Generally, dogs from breeds with smaller adult body mass live longer than dogs from breeds with larger adult body mass, and dogs from smaller also have lower risk of death from cancer.^{12,13}

Several forms of naturally occurring tumors and cancer have been found to exist in both human and dogs, and dogs provide a powerful model for improving the understanding of cancer biology as well as developing cancer treatments. As reviewed by Pinho, S. S., et al., canine mammary tumors could serve as valuable tools for studying human breast cancer.¹⁴ The human and canine osteosarcoma are similar, while the incidence in dogs are approximately 10 times to that in people, which addressed the use of canine osteosarcoma to study human osteosarcoma.^{15,16}

Genes associated with canine cancer also could help the cancer research in people. HYAL1, HYAL2, HYAL3, HYAL4, SPAM1 and HYALP1 has been found to be associated with mast cell tumors in golden retrievers, and the variant genes and genetic networks are important in both dogs and human.¹⁷ A GWAS study for canine osteosarcoma study has identified prominent cancer genes CDKN2A/B, AKT2, and BCL2.¹⁸ In standard poodles, mutations in coat color control genes KITLG and MC1R are related with squamous cell carcinoma.¹⁹

Research Design

Whole genome sequencing

446 samples were included in this study. 17 of them were lab-recruited dogs from 12 breeds. WGS data of the other 429 samples were downloaded from Sequence Read Archive. Afterwards, the raw sequencing data was cleaned and aligned to the reference genome (ROS_Cfam_1.0), annotation software was used to find all potential single-nucleotide mutations. The cleaning and alignment process was done by Zicheng following standard guidelines. BWA package was used for mapping sequences to a reference genome and GATK

was used for realignment and annotation. To focus on sequencing alleles, in this study calling of genotypes was skipped and sites with major allele frequency smaller than 0.9 was not included. Genotypes will be studied in further studies. In addition, to ensure the alleles data solid and compelling, sites with combined allele depth no larger than 10 was not included in this study.

Hypothesis generation

The 17 lab-recruited samples were used for hypothesis generation. The 17 dogs were classified into two groups based on their breed-level cancer-specific mortality rate data from the cohort study by J.M. Fleming et al. and the cohort study by V. J. ADAMS et al. Because according to previous study, nearly 30% of dogs die from cancer in average, samples with breed-level cancer-specific mortality rate higher than 30% were labeled as “High cancer risk”, while the samples with mortality rate lower than or equal to 30% was labeled as “Low cancer risk”. Programs are made to count the number of each homozygote nonsense mutations carriers for the two groups separately. Then the counts are divided by the group size to estimate the prevalence of mutations. After that, the difference of the prevalence is calculated by prevalence in the higher-risk group minus prevalence in the lower-risk group. All identified premature mutations were ranked based on the difference of prevalence of mutation between the two groups. Top 5 ranked stop-gained mutations were used for the following validation process.

Validation of the association between the cancer risk and mutations

The 429 samples downloaded from Sequence Read Archive were used for validation. 645 samples were available from Sequence Read Archive, but some samples were excluded because they have no breed-level ratio of death from cancer, in which case we can't evaluate the cancer risk of the sample. Fisher's exact test was used and the null hypothesis is that ratio of cancer death in two groups are not different. To address the multiple comparisons fallacy, Bonferroni correction is applied. Because five sites were tested, the alpha-value needs to shrink from 0.05 to 0.01.

Result

Watson, Colby, Anna, Louis, RB, and 6LR are the 6 samples with a higher risk of cancer, while the other 11 samples are in the lower risk group. (Table 1) The top five candidate stop-gain mutation sites with the largest mutation prevalence difference between the high cancer risk group and the low cancer risk group were identified from our discovery dataset. (Table 2) It can be

hypothesized that these five mutations are related to a higher risk of cancer in dogs, and the association between cancer risk and each of the four mutations is evaluated in the next step.

Table 1 Ratio of death caused by cancer

Group	Breed	Ratio of death caused by cancer*	Sample ID**
High cancer risk	GOLDEN RETRIEVER	0.4782	Watson, Colby
	FRENCH BULLDOG	0.3803	Anna, Louis, RB
	LABRADOR RETRIEVER	0.3367	6LR
Low cancer risk	CHESAPEAKE BAY RETRIEVER	0.2849	Graham
	GERMAN SHEPHERD DOG	0.2770	GS, Scotia
	AMERICAN ESKIMO DOG	0.2376	4AE
	AUSTRALIAN SHEPHERD	0.2365	5AS
	ENGLISH BULLDOG	0.1992	1Y, 2O
	DACHSHUND	0.0891	3D
	PEKINGESE	0.0795	DW
	CHIHUAHUA	0.0750	2C
MINIATURE PINSCHER	0.0357	1MP	

Note:

* Published breed-level data from the study by J.M. Fleming et al. and the study by V. J. ADAMS et al.

**Samples firstly sequenced in this study

Table 2 Mutation with top prevalence difference

Rank	Prevalence Difference	Chr	Position	Gene	Mutation
1	0.7142	13	27257932	GSDMC	G>T
2	0.6667	13	59889202	LOC119874465	C>A
3	0.6667	18	54475267	LGALS12	G>A
4	0.6071	36	4708056	DAPL1	C>T
5	0.6000	12	2378563	LOC102156885	C>T

429 samples were collected from online resources. Then samples with depth no larger than 10 or potentially heterozygous were removed. And the samples passed the filter were divided into two groups based on their breed-level risk of cancer – samples belonging to breed with a ratio of death from cancer lower than 0.30 were categorized in the low-risk group, while those with a ratio equal or larger than 0.30 were categorized in the high-risk group.

Fisher exact test (Table 3) was then employed to validate the effect of the mutations on cancer risk. For mutation in gene GSDMC, the null hypothesis is that cancer risk is the same samples with or without the homozygous nonsense mutation. With the two-tailed p-value smaller than 0.01 ($p = 0.0008$),

the null hypothesis is rejected and the homozygous nonsense mutation in GSDMC is significantly associated with cancer risk. The other four mutations were tested as well. (Table 3) With the p-values greater than 0.01, null hypotheses failed to be rejected. Therefore, none of them are significantly associated with higher cancer risk.

Table 3 Validation of the mutations

Gene		Number of samples		p-value†
		Alt	Ref	
GSDMC	High risk	69	67	0.0008*
	Low risk	41	94	
LOC119874465	High risk	18	98	0.1485
	Low risk	20	189	
LGALS12	High risk	26	88	0.3643
	Low risk	31	143	
DAPL1	High risk	29	55	0.0658
	Low risk	35	120	
LOC102156885	High risk	17	83	0.6311
	Low risk	33	136	

Note:

† P-value calculated from the Fisher exact test

* Statically significant

Discussion

Position of the nonsense mutation in GSDMC

GSDMC is a gene in canine Chromosome 13 and the nonsense mutation NC_051817.1:g.27257932G>T is located in the fourth exon of GSDMC (Fig 1). The mutation locates in the 601st base out of 1524 nucleotides in the protein-coding region. Because there were no existing mRNA expression data based on the reference genome ROS_Cfam_1.0, the mutation was located in an older form of reference genome CanFam3.1 to study its effect on mRNA and protein expression. And the mutation causes the protein premature at the 201st residue out of 507 residues (Fig 2).

Based on the Ensembl database, only one transcript was reported. Till now, no alternative splicing of this gene has been reported. This stop gain mutation would cause 40 residues lost in the C-terminus of the pore forming domain and loss of the whole PUB domain (Fig 2).

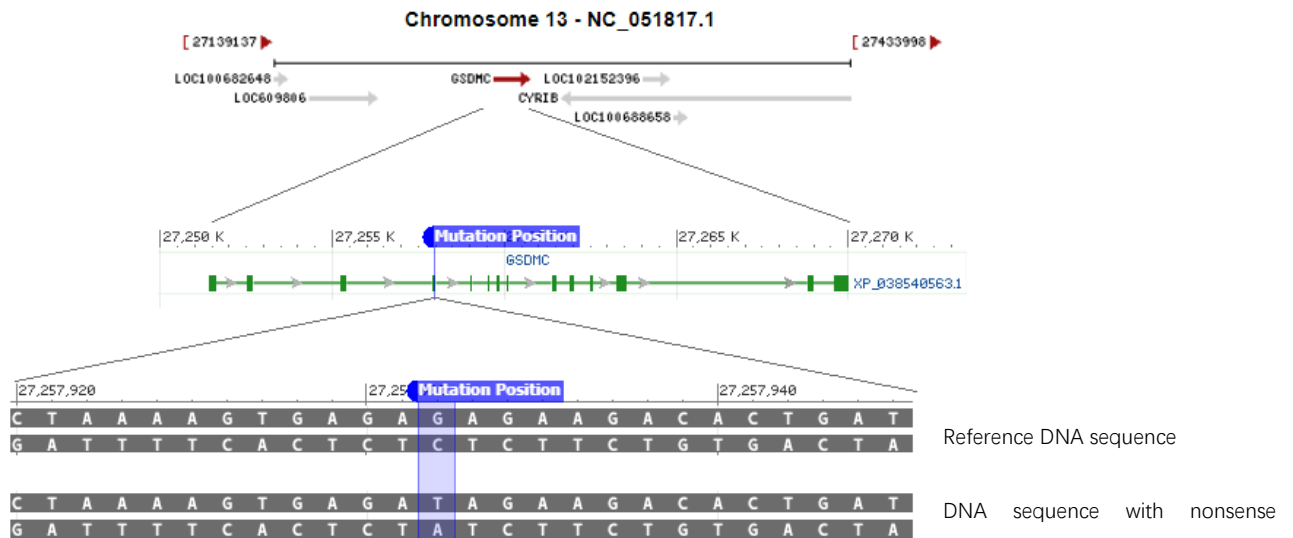


Fig 1. Position of the nonsense mutation in GSDMC DNA sequence

Note: Based on a screenshot from gene viewer, NCBI

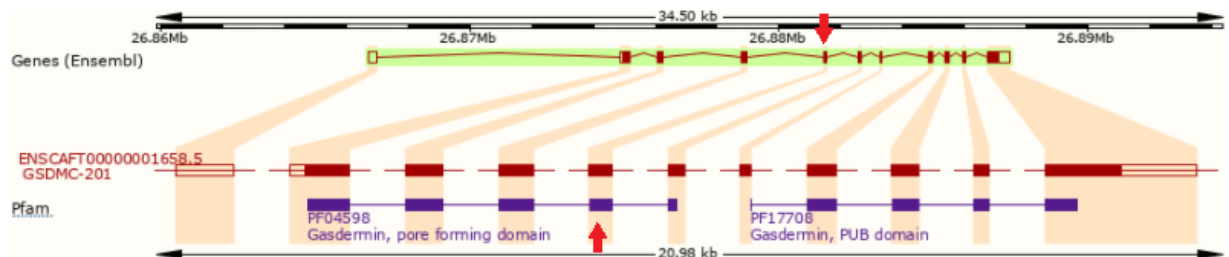


Fig 2. Position of the nonsense mutation in GSDMC transcript

Note: Based on a screenshot from Ensembl, red arrows indicate the mutation position

The potential function of GSDMC

GSDMC gene expresses Gasdermin C, a protein belonging to the Gasdermin family. The gasdermin-N domains of human GSDMC were proved to induce extensive pyroptosis in human 293T cells, while the full-length GSDMC protein shows little cytotoxicity.²⁰ In a recent study, the caspase-8-mediated cleavage of GSDMC induced pyroptosis, reducing the repression of tumor growth.²¹ Though cleavage of GSDMC was proved by several studies to suppress tumorigenesis, the effect of GSDMC expression is controversial: high-level expression of GSDMC was reported in colorectal and lung cancers, while in gastric and esophageal cancer expression of GSDMC were reported to suppress tumor progression.^{22,23,24} In addition, the function of the C

terminal domain in GSDMC has not been studied before. In this study, we find nonsense mutation NC_051817.1:g.27257932G>T in GSDMC is associated with a higher risk of cancer in canines, which suggests that dogs could serve as good animal models for GSDMC function research in the future.

Strength of the study

This study first reported the association between the identified GSDMC stop-gain mutation and increased chances to die from cancer in dogs, providing supports for that the normal GSDMC function is essential for preventing tumor development. This study also hints that, dogs could serve as great model for the *in vivo* function study on GSDMC.

The study was based on high-depth whole genome sequencing data, ensuring the sequencing accuracy. In addition, the filter based on the sequencing depth and allele proportion of alleles helped recognize true mutations. Tradition studies started with disease-associated haplotype blocks and then search for biologically meaningful mutations. Different from the tradition studies, this study focused on the stop-gain mutations directly, enhancing the biological significance.

Limitation of the study

This study didn't consider the linkage disequilibrium. Because many dog breeds were developed from a few dogs, the founder event leads to genetic diversity within breeds. Previous study has also proved the high levels of linkage disequilibrium in dogs.²⁵ This study didn't investigate mutations other than stop-gain mutation. Mutations associated with the nonsense mutation could cause the higher cancer risk. As a result, linkage group could be studied in the future to decide if other mutations nearby have correlation with the mutation in GSDMC could lead to increased ratio of death from cancer.

As discussed previously, dogs from breeds with larger body weight are more frequent to die from cancer compared with dogs with lighter body weight, while dogs from breeds with longer lifespan were more easily to die from tumor.^{26,27} The nonsense mutation could be related with body size or average age of the breeds. Relationship between breed body weight and the mutation, as well as the relationship between breed average age and the mutation were not analyzed in this study. And average age and body size should be controlled in further investigation.

The cancer risk was estimated by the ratio of death caused by neoplasm to all-cause death. And the ratio was breed-level data summarized from previous studies, but not based on the actual individual cause of death, which could

weaken the accuracy of the study. When analyzing the whole genome sequencing data, sample size of some breeds was small, causing many extreme values in the breed-level alternative allele prevalence in the validation step. Such extreme values could reduce the precision of the study.

Conclusion

Based on high-depth whole genome sequencing data, this study found the association between a nonsense mutation in GSDMC and increased chances to die from cancer in dogs, providing supports for that the normal GSDMC function is essential for preventing progression of cancer. As discussed previously, this study lacked the investigation on potential linkage disequilibrium, confounders, and individual-level cancer risk data. In addition to the limitation, further research could also focus on the pathological function of GSDMC using dogs as animal models to study human cancer.

Reference

- ¹ Galibert, F., Quignon, P., Hitte, C., & André, C. (2011). Toward understanding dog evolutionary and domestication history. *Comptes rendus biologies*, 334(3), 190-196.
- ² Alvarez, C. E. (2014). Naturally occurring cancers in dogs: insights for translational genetics and medicine. *ILAR journal*, 55(1), 16-45.
- ³ Alvarez, C. E. (2014). Naturally occurring cancers in dogs: insights for translational genetics and medicine. *ILAR journal*, 55(1), 16-45.
- ⁴ The University of Sydney - OMIA - Online Mendelian inheritance in Animals. OMIA. (n.d.). Retrieved April 10, 2022, from <https://www.omia.org/home/>
- ⁵ Lindblad-Toh, K., Wade, C. M., Mikkelsen, T. S., Karlsson, E. K., Jaffe, D. B., Kamal, M., ... & Lander, E. S. (2005). Genome sequence, comparative analysis and haplotype structure of the domestic dog. *Nature*, 438(7069), 803-819.
- ⁶ Culbertson, M. R. (1999). RNA surveillance: unforeseen consequences for gene expression, inherited genetic disorders and cancer. *Trends in Genetics*, 15(2), 74-80.

-
- ⁷ Rolfini, R., & Cabrini, G. (1993). Nonsense mutation R1162X of the cystic fibrosis transmembrane conductance regulator gene does not reduce messenger RNA expression in nasal epithelial tissue. *The Journal of clinical investigation*, 92(6), 2683-2687.
- ⁸ Chang, J. C., & Kan, Y. W. (1979). beta 0 thalassemia, a nonsense mutation in man. *Proceedings of the National Academy of Sciences*, 76(6), 2886-2889.
- ⁹ Frischmeyer, P. A., & Dietz, H. C. (1999). Nonsense-mediated mRNA decay in health and disease. *Human molecular genetics*, 8(10), 1893-1900.
- ¹⁰ Adams, V. J., Evans, K. M., Sampson, J., & Wood, J. L. N. (2010). Methods and mortality results of a health survey of purebred dogs in the UK. *Journal of Small Animal Practice*, 51(10), 512-524.
- ¹¹ Dorn, C. R., Taylor, D. O. N., Schneider, R., Hibbard, H. H., & Klauber, M. R. (1968). Survey of animal neoplasms in Alameda and Contra Costa Counties, California. II. Cancer morbidity in dogs and cats from Alameda County. *Journal of the National Cancer Institute*, 40(2), 307-318.
- ¹² Li, Y., Deeb, B., Pendergrass, W., & Wolf, N. (1996). Cellular proliferative capacity and life span in small and large dogs. *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences*, 51(6), B403-B408.
- ¹³ Greer, K. A., Canterbury, S. C., & Murphy, K. E. (2007). Statistical analysis regarding the effects of height and weight on life span of the domestic dog. *Research in veterinary science*, 82(2), 208-214.
- ¹⁴ Pinho, S. S., Carvalho, S., Cabral, J., Reis, C. A., & Gärtner, F. (2012). Canine tumors: a spontaneous animal model of human carcinogenesis. *Translational Research*, 159(3), 165-172.
- ¹⁵ Fenger, J. M., London, C. A., & Kisseberth, W. C. (2014). Canine osteosarcoma: a naturally occurring disease to inform pediatric oncology. *ILAR journal*, 55(1), 69-85.
- ¹⁶ Pinho, S. S., Carvalho, S., Cabral, J., Reis, C. A., & Gärtner, F. (2012). Canine tumors: a spontaneous animal model of human carcinogenesis. *Translational Research*, 159(3), 165-172.
- ¹⁷ Arendt, M. L., Melin, M., Tonomura, N., Koltookian, M., Courtoy-Cahen, C., Flindall, N., ... & Lindblad-Toh, K. (2015). Genome-wide association study of golden retrievers identifies germ-line risk factors predisposing to mast cell tumours. *PLoS genetics*, 11(11), e1005647.

-
- ¹⁸ Fenger, J. M., London, C. A., & Kisseberth, W. C. (2014). Canine osteosarcoma: a naturally occurring disease to inform pediatric oncology. *ILAR journal*, 55(1), 69-85.
- ¹⁹ Karyadi, D. M., Karlins, E., Decker, B., vonHoldt, B. M., Carpintero-Ramirez, G., Parker, H. G., ... & Ostrander, E. A. (2013). A copy number variant at the KITLG locus likely confers risk for canine squamous cell carcinoma of the digit. *PLoS genetics*, 9(3), e1003409.
- ²⁰ Ding, Jingjin, et al. "Pore-forming activity and structural autoinhibition of the gasdermin family." *Nature* 535.7610 (2016): 111-116.
- ²¹ Zhang, Jia-yuan, et al. "The metabolite α -KG induces GSDMC-dependent pyroptosis through death receptor 6-activated caspase-8." *Cell research* 31.9 (2021): 980-997.
- ²² Miguchi, Masashi, et al. "Gasdermin C is upregulated by inactivation of transforming growth factor β receptor type II in the presence of mutated Apc, promoting colorectal cancer proliferation." *PloS one* 11.11 (2016): e0166422.
- ²³ Wei, Jie, et al. "Overexpression of GSDMC is a prognostic factor for predicting a poor outcome in lung adenocarcinoma." *Molecular medicine reports* 21.1 (2020): 360-370.
- ²⁴ Saeki, Norihisa, et al. "Distinctive expression and function of four GSDM family genes (GSDMA-D) in normal and malignant upper gastrointestinal epithelium." *Genes, Chromosomes and Cancer* 48.3 (2009): 261-271.
- ²⁵ Ostrander, E. A., & Wayne, R. K. (2005). The canine genome. *Genome research*, 15(12), 1706-1716.
- ²⁶ Li, Y., Deeb, B., Pendergrass, W., & Wolf, N. (1996). Cellular proliferative capacity and life span in small and large dogs. *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences*, 51(6), B403-B408.
- ²⁷ Greer, K. A., Canterbury, S. C., & Murphy, K. E. (2007). Statistical analysis regarding the effects of height and weight on life span of the domestic dog. *Research in veterinary science*, 82(2), 208-214.