Available online at www.sciencedirect.com
ScienceDirect
Genomics 90 (2007) 354–363
GENOMICS
www.elsevier.com/locate/ygeno

Heritability and segregation analysis of osteosarcoma in the Scottish deerhound

Jeffrey C. Phillips [a,*], Betty Stephenson [b], Marlene Hauck [c], John Dillberger [d]

[a] Department of Small Animal Clinical Sciences, University of Tennessee, Knoxville, TN 37996-4544, USA
[b] Greystone Pet Hospital, Bowling Green, KY 42104, USA
[c] North Carolina State University, Raleigh, NC 27606, USA
[d] Preclinical Development Services, Nashville, TN 47448, USA

Received 12 December 2006; accepted 1 May 2007
Available online 12 July 2007

Abstract

Osteosarcoma is the most common malignant bone tumor in dogs and, like its human orthologue, is characterized by aggressive local behavior and high metastatic rates. The Scottish deerhound is a breed of dog with a >15% incidence of osteosarcoma and represents an excellent spontaneously occurring large-animal model of the human disease. We modeled the transmission of the osteosarcoma phenotype in a population of over 1000 related deerhounds ascertained as part of a prospective health study. Variance component analysis, segregation analysis, and linear modeling were performed to evaluate heritability, to infer the presumptive transmission model, and to identify covariate effects for this phenotype within the breed, respectively. Based on variance component analysis, heritability (h^2) was estimated to be 0.69. Six transmission models were analyzed by segregation analysis; based on Akaike's information criteria, the most parsimonious model was the Mendelian major gene model with dominant expression. Linear modeling identified gender and genotype as significant predictors of disease outcome. Importantly, duration of gonadal hormone exposure, weight, and height at maturity were not significant predictors of outcome. Inheritance of the putative high-risk allele was thus associated with >75% risk of disease occurrence compared to the <5% baseline risk. These results support the hypothesis that a major gene with a dominant effect explains most of the osteosarcoma phenotype within the Scottish deerhound.

Keywords: Cancer; Osteosarcoma; Canine; Segregation; Heritability

Osteosarcoma (osteogenic sarcoma; OSA) is the most common (>85%) malignant bone tumor of dogs and represents an excellent model for osteogenic sarcoma in humans [1]. In the United States, approximately 8000 canine and 400 human cases of osteosarcoma are diagnosed each year [2,3]. In the canine population, large and giant breeds of dogs are at increased risk for the development of OSA; together these breeds account for the majority (>80%) of reported cases [4].

The Scottish deerhound is a giant-breed dog first recognized by the American Kennel Club (AKC) in 1935 (Fig. 1A). The overall breed-specific incidence of OSA in the deerhound has been estimated to be greater than 150 cases per 1000 dogs (compared to 7 cases per 100,000 dogs in the general dog population) [4,5]. Clinical signs seen in affected dogs include lameness, swelling in an extremity, and in some cases pathologic fracture. A presumptive diagnosis of OSA can be made through a combination of signalment, clinical presentation, and radiographic findings (Fig. 1B). Pathologic tissue evaluation, however, is needed to confirm the diagnosis. Given the high incidence of OSA within a relatively small, closely related population, it is likely that a major gene for the disease exists within this breed.

OSA in dogs and humans has many similarities, including the predilection for metaphyseal regions of long bones, high-grade malignancy, a high rate of metastasis, and the lung as the most common site of metastasis [1]. Because of the successful combination of adjuvant chemotherapy with local therapies, the survival rates in both dogs and humans have significantly improved. Despite these advances, no preventative and very few prognostic strategies have been implemented because the molecular events driving tumorigenesis in OSA remain unknown in both species. Identification of risk factors that correlate with

* Corresponding author. Fax: +1 865 974 5554.
E-mail address: jeff_phillips@utk.edu (J.C. Phillips).

0888-7543/$ - see front matter © 2007 Elsevier Inc. All rights reserved.
doi:10.1016/j.ygeno.2007.05.001

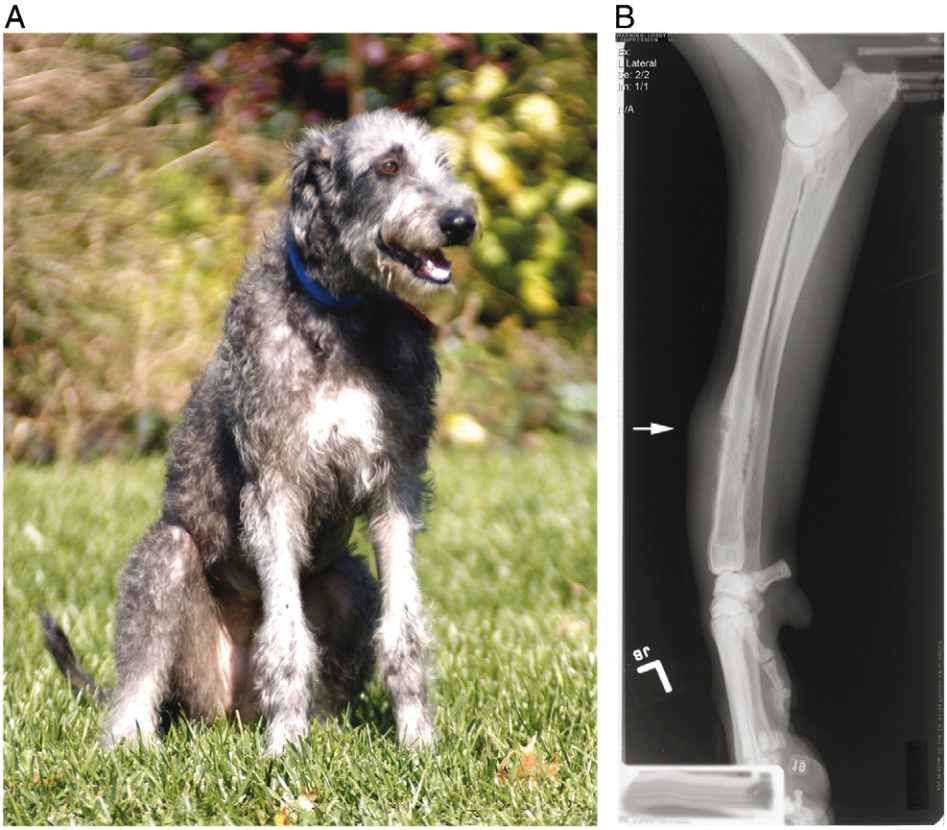

Fig. 1. (A) Photograph of an unaffected juvenile female Scottish deerhound. (B) Lateral radiograph of the distal forelimb from an adult deerhound exhibiting the classic findings seen in osteosarcoma. Radiographic features include bony lysis, periosteal reaction, and a cortical defect from a previous bone biopsy. Circumferential soft tissue swelling is also present at the tumor site (arrow).

tumorigenesis, tumor progression, or response to therapy would greatly improve our understanding of this disease.

Numerous molecular studies have been undertaken in recent years to elucidate the etiology of human and canine OSA [6–10]. It was evident from the earliest studies of DNA ploidy that the karyotypes in human OSA are usually complex, particularly in high-grade tumors. Loss-of-heterozygosity screening has identified several chromosomal regions associated with high frequencies of allelic loss. Comparative genomic hybridization has also been used to identify regions of genomic amplification in human-origin tumors. In contrast, much less information exists regarding the cytogenetic, genetic, or molecular changes underlying the development of canine OSA. The few studies to date have used immunohistochemistry to identify genes with altered expression in canine OSA, including TP53, RB1, MDM2, PTGS2, and PTEN. The prognostic significance of these findings has not been shown [8–10].

Several veterinary reports [3,4,11,12] have examined the incidence of canine osteosarcoma in an attempt to identify risk factors for disease. Two of these reports [4,11] have suggested adult body size to be the strongest correlate of risk for osteosarcoma development. As noted previously, certain breeds [4] also have a well-known increased risk for osteosarcoma. Actual breed-specific incidence, however, has been reported only for the Rottweiler at 12.6% [12]. Studies have also suggested gender to be a risk factor; although contradictory information exists in the literature [3,4,12,13]. A recent report [12] has evaluated the role of gonadal hormone exposure in the development of osteosarcoma. Results from this study suggested that overall lifetime risk for bone sarcoma formation was inversely related to total gonadal hormone exposure. No studies to date have evaluated the combined effects of age, gender, height, weight, and familial correlates on the risk of bone tumor formation.

The objectives of this study were to estimate the heritability of osteosarcoma by variance component analysis, to use complex segregation analysis to identify the presumptive mode of inheritance, and to use a generalized linear model to identify predictors of risk for osteosarcoma development within purebred Scottish deerhounds. Further knowledge of the genetic factors underlying canine OSA could provide important information regarding mechanisms of tumorigenesis, prognostic biomarkers, and potential targets for novel gene therapy (for both canine and human OSA).

Results

Descriptive analysis

There were 1057 dogs identified through five affected probands. Based on AKC registrations [23], this population represents ~21% of all registered deerhounds during the years

Table 1A
Descriptive statistics for 1057 dogs in 314 families analyzed in the segregation analysis of osteosarcoma (mean±standard deviation)

Total population	Males (n=479)	Females (n=578)	All
Total individuals			1057
Affected	74	137	
Unaffected	382	412	
% Affected	16%	25%	
Missing	23	29	
Probands	4	1	
Sibships			314
Sibship size			1–11 (3.0±1.9)
Inbreeding loops			187
Inbreeding coefficient			0.032
Age (years)	0.5–14.9[a] (6.6±2.8)	0.5–14.9[a] (7.4±3.0)	
Age of onset (years)	2–13.9[a] (7.0±2.0)	2.5–13.0[a] (8.1±2.5)	

[a] Significant difference noted between gender-specific distributions (p<0.001).

Table 1B
Subgroup statistics for 274 dogs analyzed in covariate analysis of weight, height, and length of time intact (mean±standard deviation)

Subgroup	Males (n=140)	Females (n=134)
Total individuals		
Affected	22	36
Unaffected	118	98
Missing	0	0
Age (years)	1.2–11.5[a] (7.5±2.5)	3.7–13.9[a] (9.0±2.1)
Age of onset (years)	6.2–7.9[a] (7.0±1.0)	6.5–9.9[a] (8.4±1.3)
Height (in.)		
Affected	31.5–34.5[a] (32.0±0.7)	29.0–32.5[a] (30.5±1.0)
Unaffected	30.0–35.0[a] (32.7±1.1)	28.5–33.5[a] (30.8±1.0)
Weight (lb)		
Affected	95.0–118.0[a] (109±8.2)	75.0–105.0[a] (92.0±7.6)
Unaffected	82.0–125.0[a] (106±9.5)	65.0–100.0[a] (90.0±8.4)
Length of time intact (years)		
Affected	1.5–13.9 (6.4±3.2)	0.5–13.0 (7.2±3.5)
Unaffected	0.5–14.9 (6.6±2.8)	0.5–14.9 (7.4±3.0)

[a] Significant difference noted between gender-specific distributions (p<0.001).

1975–2003. A single pedigree was constructed to include all 1057 dogs: 479 males and 578 females (Table 1A). The pedigree contained 314 sibships of average size 3 (SD=1.9) and range 1–11. A total of 187 inbreeding loops were identified, with an average inbreeding coefficient of 0.032 estimated using the kinship matrices option of MENDEL 6.01. Due to the graphical complexity of this pedigree, only a small subset of the pedigree is shown in Fig. 2.

Of these dogs, 211 were confirmed affected, 794 were unaffected at the time of exam, and 52 had unknown health status or were lost to follow-up. The absolute percentage of affected dogs among all dogs with known health status was 21% (211/1005). Affected dogs ranged in age from 2 to 13.9 years with a mean age of onset of 7.7 years. Affected dogs were also significantly older than unaffected dogs (7.7 vs 7.0 years, p=0.003); suggesting age to be an important predictor of health status.

Gender-specific differences were present in this population. Overall, male dogs were significantly younger than female dogs (6.6 vs 7.4 years, p<0.001). Age of onset also varied significantly between the genders, with affected male dogs significantly younger than affected female dogs (7.0 vs 8.1 years, p=0.001). The female-specific incidence of osteosarcoma was higher than the male-specific incidence (25% vs 16%). Information on height, weight, and length of time sexually intact was available for 274 dogs (Table 1B). Based on gender and affected status this subgroup was not significantly different from the larger population (data not shown). Males in this subgroup were significantly taller than females (33 vs 31 in., p<0.0001) and weighed more than the females (107 vs 91 lb, p<0.0001) at maturity. However, there was no significant difference in height or weight between affected and unaffected individuals of either male or female gender in this subgroup (p>0.4 for each comparison, respectively). Finally, length of time sexually intact was not significantly different between affected and unaffected dogs or between genders. Together, these results suggest that age and gender are significant predictors of health status in this population, while height, weight, and length of time sexually intact are not.

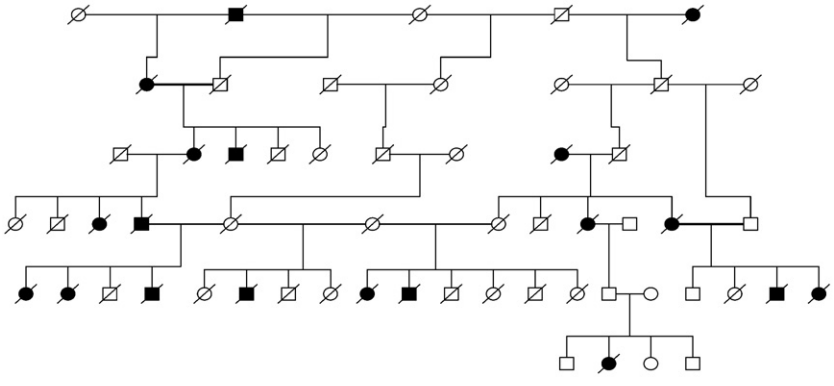

Fig. 2. Subset pedigree of 57 Scottish deerhounds with known affected status, drawn using the Cyrillic pedigree software (Cyrillic 2.1.3; Cherwell Software). Standard pedigree symbols are used. Affected individuals are represented by filled symbols. Multiple inbreeding (bold lines) and mating loops are present. Deceased dogs have diagonal lines through their symbols.

Heritability

The results of variance component analysis are listed in Table 2. Models with and without different variance components were compared to the most general model (Model 4). The hypothesis of no genetic variance was rejected in favor of models including a genetic component (Models 1 and 2 vs Models 3 and 4). Addition of covariates for gender significantly improved the fit of the variance model (Model 1 vs Model 2). Addition of a covariate for age, however, did not significantly improve the variance model (data not shown). Inclusion of a dominance (nonadditive) variance component (Model 3 vs Model 4) significantly improved the variance model ($p=0.008$, $\chi^2=7.0$, $df=1$); together, these nonadditive components accounted for 30% of the total phenotypic variance. The final model thus included both additive and dominance variances along with a covariate for gender.

Segregation analysis

Parameter estimates and likelihoods from the segregation analysis using the software jPAP [15] are presented in Table 3. This software does not include an option for the incorporation of nongenetic covariates; however, it does allow for the categorical variables such as gender. The software also allows for the addition of within-genotype correlations; the presence of which implies a mixed (major gene + polygenic component) model. Therefore, a total of seven models were analyzed, each with gender-specific disease penetrance and transmission probabilities and with correction for ascertainment bias. The fully parameterized model (Model 7) also included within-genotype correlations modeled as variance components. Addition of these within-genotype components to the general model, however, failed to improve the fit significantly (Model 7 vs Model 6, $p=0.17$); these components were thus not included in any submodels (Models 1–6).

As shown in Table 3, the No major locus and Environmental models were rejected against the general model ($p<0.0001$). Based on the likelihood ratio test (LRT), there was no significant difference between the Mendelian recessive ($p=0.39$), dominant ($p=0.90$), or codominant ($p=0.76$) model and the fully parameterized model (Model 7). Not surprisingly, both general models provided penetrance estimates and transmission probabilities that were very similar to the Mendelian dominant model (Table 3). Finally, based on Akaike's Information Criteria (AIC), the Mendelian dominant model proved to be the most parsimonious. Inheritance of a putative high-risk allele ($q_A=0.128$) predicted gender-specific risks (i.e., penetrance) for osteosarcoma development of 72 and 99% for males and females, respectively. The estimated baseline (i.e., sporadic) risk for osteosarcoma development among individuals without the high-risk allele was ~3–4%.

Linear modeling

Results of linear modeling of the osteosarcoma phenotype with correction for ascertainment bias are summarized in Table 4. Inclusion of a major gene effect (two or three ousiotypes) provided a significantly better fit over the No major gene model (Models 2–4 vs 1). To refine the Mendelian transmission model, and determine the effects of covariates on phenotype, three autosomal Mendelian models

Table 2
Variance component analysis and heritability estimates with modeling of covariate effects for gender (mean±standard deviation)

	Model 1 Environmental	Model 2 Environmental w/ covariates	Model 3 Environmental + additive w/ covariates	Model 4 Environmental + additive + dominance w/ covariates
Parameter means				
GRAND	0.2100±0.0128	0.2059±0.0128	0.2025±0.0260	0.2033±0.0264
Female gender		0.0436±0.0128	0.0389±0.0116	0.0368±0.0114
Male gender		−0.0436±0.0128	−0.0389±0.0116	−0.0368±0.0114
Variance components				
Additive		NA	0.1203±0.0138	0.1230±0.0155
Dominance		NA	NA	0.0529±0.0226
Environmental	0.1659±0.0074	0.1640±0.0073	0.0541±0.0075	0.0022±0.0223
Total	0.1659	0.1640	0.1744	0.1782
Heritability	NA	NA	0.69 [a]	0.99 [b]
Summary statistics				
ln(L)	400.3	406.0	491.6	495.1
Parameters	2	3	4	5
df	3	2	1	–
χ^2	189.6	178.2	7.0	–
p value	<0.001	<0.001	0.0082	–

Parameters are estimated using the variance component option of MENDEL [15] software and maximum likelihood methods. p values compare the most general model (Model 4) to submodels and were obtained by using a likelihood ratio test. The null hypothesis of no heritability is represented in Models 1 and 2. ln(L), natural log-likelihood; df, degrees of freedom.

[a] Model 3 heritability value corresponds to the narrow-sense heritability ($h^2=\sigma_a/\sigma_T$).

[b] Model 4 heritability value corresponds to the broad-sense heritability ($H^2=\sigma_a+\sigma_d/\sigma_T$).

Table 3
Segregation analysis of osteosarcoma in 1057 Scottish deerhounds

Parameter	Model 1 No major locus	Model 2 Environmental	Model 3 Mendelian recessive	Model 4 Mendelian dominant	Model 5 Mendelian codominant	Model 6 General Model I	Model 7 General Model II[a]
q_A[b]	[1.0]	0.098	0.499	0.128	0.131	0.131	0.093
q_B		0.902 (0.009)	0.501 (0.03)	0.872 (0.02)	0.869 (0.02)	0.869 (0.02)	0.907 (0.02)
ψ_{AA}[c]			0.249	0.016	0.015	0.015	0.009
ψ_{AB}[c]			0.500	0.224	0.212	0.212	0.168
ψ_{BB}[c]			0.251	0.760	0.773	0.773	0.823
$\tau_{AA\ male}$		0.903 (0.009)	[0]	[0]	[0]	0.059 (0.36)	0.270 (0.42)
$\tau_{AB\ male}$		$=\tau_{AA}$	[0.5]	[0.5]	[0.5]	0.514 (0.02)	0.505 (0.02)
$\tau_{BB\ male}$		$=\tau_{AA}$	[1.0]	[1.0]	[1.0]	0.995 (0.02)	1.0[d]
$\tau_{AA\ female}$		0.903 (0.009)	[0]	[0]	[0]	0.246 (0.65)	0.018 (0.50)
$\tau_{AB\ female}$		$=\tau_{AA}$	[0.5]	[0.5]	[0.5]	0.481 (0.02)	0.501 (0.02)
$\tau_{BB\ female}$		$=\tau_{AA}$	[1.0]	[1.0]	[1.0]	1.0[d]	1.0[d]
$\beta_{AA\ male}$	0.184 (0.02)	1.0[d]	0.750 (0.08)	0.718 (0.08)	1.0[d]	1.0[d]	0.658 (0.06)
$\beta_{AB\ male}$	$=\beta_{AA\ male}$	0.998 (0.01)	0.0[d]	$=\beta_{AA}$	0.669 (0.09)	0.694 (0.09)	0.691 (0.06)
$\beta_{BB\ male}$	$=\beta_{AA\ male}$	0.0[d]	$=\beta_{AB}$	0.035 (0.02)	0.034 (0.01)	0.028 (0.02)	0.109 (0.04)
$\beta_{AA\ female}$	0.267 (0.02)	0.999 (0.01)	0.944 (0.05)	0.995 (0.05)	1.0[d]	1.0[d]	0.947 (0.02)
$\beta_{AB\ female}$	$=\beta_{AA\ female}$	0.999 (0.01)	0.012(0.01)	$=\beta_{AA}$	0.980 (0.06)	0.986 (0.06)	0.999 (0.02)
$\beta_{BB\ female}$	$=\beta_{AA\ female}$	0.101 (0.03)	$=\beta_{AB}$	0.042 (0.02)	0.039 (0.02)	0.041 (0.02)	0.134 (0.04)
Heritability	NA	NA	[0]	[0]	[0]	[0]	1.0[d]
Statistics							
Parameters	2	7	5	5	7	13	14
df	12	7	9	9	7	1	–
$-2\ \ln(L)$	753.72	753.73	644.50	638.77	637.75	636.90	634.99[e]
AIC	777.72	767.73	654.50	648.77	651.75	662.90	662.99
p value	<0.0001	<0.0001	0.39	0.93	0.91	0.17	Reference

Parameters were estimated using jPAP [13] software and maximum likelihood methods. Submodels are compared to the most general model (Model 7) using a standard likelihood ratio test. Models are compared to each other using AIC. Numbers in brackets [ ] are fixed in this analysis. Standard errors are listed in parentheses ( ). q_A, frequency of putative high-risk allele A. $\psi_{i/j}$, genotype probability for individual with type *i/j* assuming Hardy–Weinberg equilibrium. $\tau_{i/j}$, transmission probability for individual with type *i/j*. $\beta_{i/j}$, gender-specific penetrance for individual with type *i/j*. The parameter Heritability refers to the within-genotype variance of the genetic model. *df*, degrees of freedom. $-2\ \ln(L)$, minus twice the log-likelihood. AIC, Akaike's information criteria. *p* value, comparing general model with submodels assuming χ^2 distribution.

[a] General Model II includes an estimated within-genotype parameter modeled as variance components.
[b] Frequency of high-risk allele (q_A) calculated as $1-q_B$.
[c] Values are calculated from allele frequencies assuming Hardy–Weinberg equilibrium.
[d] Boundary maximum verified.
[e] Reference value used to compare submodels.

were compared (dominant, recessive, and genotypic). The Mendelian recessive model appeared to be a poor fit with high standard errors and was rejected against the most general model ($p<0.0001$). Based on AIC, the Mendelian model with dominant transmission appeared to be the most parsimonious model. The frequency of the high-risk allele in this model is $q_A=0.118$.

Covariate effects of gender, age, genotype, height and weight at maturity, and length of gonadal hormone exposure were also estimated by linear modeling. Only gender and genotype significantly improved the fit of the data in this population and were therefore included in all analyses. Assuming the Mendelian dominant model of transmission, female gender increased the relative risk (or odds ratio) of affectation by 2.8 in comparison to males. Under the dominant model, the penetrance (probability of affectation) for individuals carrying the high-risk allele (genotypes ψ_{AA} and ψ_{AB}) was 77 and 90%, for males and females, respectively. The baseline probabilities of being affected for individuals without the high-risk allele were 1.7 and 4.5% for males and females, respectively.

Discussion

Several reports [3,4,11] have suggested that height, weight, and gender are significant modifiers of risk for the development and/or the biological behavior of osteosarcoma in large-breed dogs. More recently, a nonrandomized retrospective cohort study in the Rottweiler breed [12] suggested that duration of gonadal hormone exposure was predictive for both the development of bone cancer and its clinical behavior. Other covariates such as height, weight, age, or gender were not found to be predictive of either tumor development or behavior within the Rottweiler breed. In this deerhound population, covariates for height, weight, age, gender, and length of gonadal hormone exposure were examined. While the descriptive analysis identified age and gender as important covariates, gender was the only covariate that proved to be predictive for bone tumor formation in multivariate analysis (segregation analysis and linear modeling). These apparent differences in risk factors may be population specific (i.e., breed or subpopulation within breed) or secondary to other confounding factors. This population of deerhounds has an average inbreeding coefficient

Table 4
Linear modeling of osteosarcoma in 1057 Scottish deerhounds

Parameter	Model 1 No major gene (H_0)	Model 2 Recessive	Model 3 Dominant	Model 4 Genotypic
q_A	0.120 (0.03)	0.834 [a]	0.118 (0.02)	0.117 (0.02)
ψ_{AA}	–	0.696	0.014	0.014
ψ_{AB}	–	0.276	0.206	0.206
ψ_{BB}	–	0.028	0.780	0.780
β_o	−1.375 (0.30)	−18.59 (330.1)	−0.925 (0.29)	−0.091 (0.61)
β_{AA}	–	−19.31 (330.1)	2.638 (0.23)	1.633 (1.50)
β_{AB}	–	19.31 (330.1)	$=\beta_{AA}$	1.833 (0.91)
β_{BB}	–	$=\beta_{AB}$	−2.638 (0.23)	−3.466 (0.67)
Covariate				
Male	−0.266 (0.12)	−0.598 (0.2)	−0.521 (0.20)	−0.521 (0.20)
Female	0.266 (0.12)	0.598 (0.2)	0.521 (0.20)	0.521 (0.20)
Statistics				
Parameters	3	4	4	6
df	3	2	2	–
$-2\ln(L)$	1018.52	895.07	875.86	875.85
χ^2	142.67	19.22	0.01	–
AIC	1024.52	903.07	883.86	887.85
p value	<0.0001	<0.0001	0.995	Reference

Parameter estimates were calculated using the penetrance option of MENDEL [15] software and maximum likelihood methods. *p* values were obtained by comparing submodels to the most general model (Model 4) using a standard likelihood ratio test and assuming a χ^2 distribution. Models are further compared to each other using Akaike's information criteria (AIC). Standard errors are listed in parentheses. q_A, frequency of putative high-risk allele A. $\psi_{i/j}$, genotype probability for individual with type *i/j* assuming Hardy–Weinberg equilibrium. $\beta_{i/j}$, genotype offset for individual with type *i/j*. *df*, degrees of freedom. $-2\ln(L)$, minus twice the log-likelihood.

[a] Standard error could not be calculated.

of 0.032, which is likely higher than that found in larger and more diverse populations such as the Rottweiler. Increased levels of inbreeding serve to reduce genetic variability and thus increase major gene effects (a process referred to as trait "Mendelization"), while decreasing the significance (i.e., variability) of polygenic effects such as height, weight, and duration of gonadal hormone exposure. Also, this population of dogs is enriched for dogs affected with osteosarcoma; this enrichment can make it even more difficult to quantify polygenic risk factors [24].

It is also possible that the covariate effects of height and weight are confounded by gender-specific differences in disease incidence. Notably, the incidence of osteosarcoma in female dogs is ~1.6 times higher than the incidence in male dogs. Possible origins for this gender predilection include sampling bias, or alternatively, one could postulate that a portion of male deerhounds die of unrelated causes (i.e., nonosteosarcoma) at a younger age, on average, than female deerhounds. In reference to sampling bias, ascertainment was performed through disease phenotype and covariates were not involved in the sampling process; as such, they should not require correction. Furthermore, the population distribution with respect to these covariates (age, gender, height, and weight) is not significantly different from that previously reported [5]. Support for the alternate "early death" hypothesis is found in the observation that males appear to be more susceptible to heart-disease-related deaths, which occur more frequently (27% vs 16%) and at a younger age in males than in females (7.6 years in males vs 8.9 years in females) [5]. These "early" deaths serve to eliminate potentially affected male carriers from the population prior to the onset of their cancer; while they are phenotypically unaffected. The result is a significant age differential between unaffected males and females (6.6 vs 7.0 years, $p=0.0009$) and a skewing of gender-specific disease incidences as described here and transmission/penetrance estimates as described below.

The heritability of the osteosarcoma phenotype within the deerhound breed was evaluated by variance component analysis. The results demonstrated the presence of a significant additive and dominance variance component. The existence of a large dominance variance suggests the presence of single (dominant) high-risk allele within this population. On a theoretical level, there are two problems with the variance component analysis. First, the likelihood ratio test used to compare models assumed differences to be asymptotically distributed as a χ^2. The presence of a lower bound of 0, inherent in the model, changes the distribution of the likelihood ratio test to an even mixture of a point mass at zero and a χ^2 with 1 degree of freedom [25]. Thus the real *p* value for comparing models with and without a dominance component is likely different from the estimated value of 0.008. Second, the variance component model used herein assumed the absence of inbreeding. The presence of inbreeding results in a more complex dominance relationship that cannot be modeled with currently available software. Furthermore, inbreeding and the resultant complex population structure also necessitated population subdivision prior to statistical analysis. Due to this subdivision we also cannot rule out the possibility of more than one dominantly acting gene (contributing to the dominance variance) in this population; although given a relatively small and inbred population, a single locus is most likely. In summary, while variance component analysis identified a significant additive effect in this population ($h^2=0.69$), the true significance of a nonadditive (dominance) effect is unknown.

Transmission of the putative high-risk allele was studied using both segregation analysis and linear modeling. Results from each approach were consistent with the presence of a dominant major gene with gender-specific transmission and penetrance parameters. Specifically, in segregation analysis comparison of parameter estimates in the Mendelian dominant model to those of the unrestricted (general) model agrees surprisingly well (Table 3), providing strong support to the dominant transmission model. High standard errors, however, were noted for the transmission parameters associated with the affected homozygote in the general model. These likely reflect the presence of relatively few affected homozygotes, which are required to differentiate accurately the three genotypes. Mendelian models avoid this difficulty by requiring the presence of only two discernible genotypes. In linear modeling, likelihood ratio tests provide strong support for the genetic models (Table 4), with the dominant model being the most parsimonious. Large standard errors were also noted in the genotype offsets for the general model, again reflecting the presence of few affected homozygotes. Finally, the presence of gender-specific transmission and penetrance parameters suggest

a mixed model of inheritance rather than a true autosomal dominant. However, one could argue that it is quite common to identify autosomal dominant traits that exhibit clear sex predilections [26–28]. As discussed previously, a possible origin for these gender-specific parameters is the selective "early" death of unaffected male dogs. This early death hypothesis could be stringently tested by applying an age-dependent penetrance function in segregation analysis. This function would serve to correct for the death of young unaffected dogs. However, age-dependent penetrance functions are not an option in either the jPAP or the MENDEL software package.

Segregation analysis and linear modeling also allowed for allele frequency and penetrance estimation. The methodology used for this estimation differs between the software packages jPAP and MENDEL, each having different theoretical methods and specific constraints [15,16]. Regardless, parameter estimates in both segregation analysis and linear modeling are fairly consistent; attesting to the robustness of this dataset. Differences in ousiotype-specific penetrance (β_{AA}, β_{AB}, β_{BB}) with each approach were within parameter measurement errors. Allele frequency (q_A) estimates from segregation analysis and linear modeling ranged from 0.12 to 0.13 (±0.02), while penetrance (β_{ij}) estimates ranged from 72 to 77% and 90 to 99% for males and females, respectively.

The allele frequency values estimated here are likely higher than those found in the general deerhound population, directly reflecting the ascertainment process used herein. The families selected for analysis were identified using a sequential process that greatly enriches the number of affected individuals [29]. Ascertainment correction was then applied by conditioning on the proband's phenotype. This is a common procedure used in pedigree analysis [30]. Strictly speaking, however, conditioning on a proband's phenotype results in a correct formulation of the likelihood only in cases of single ascertainment [14]. Single ascertainment is rarely an appropriate model in animal genetics (and to some extent in humans) given the finite population sizes and presence of inbreeding. There are more robust approximate methods for ascertainment correction; however, these methods have not been incorporated into any software applicable to the population described here [30].

When considering the apparent "intractability" of the ascertainment problem, it is important to realize that its largest impact in this case is on allele frequencies; penetrance and transmission parameters are likely more consistent both within this population and for the breed as a whole. These estimates were based on 314 sequentially collected nuclear families. Based on AKC registrations, this collection represents ~21% of all dogs registered between the years 1975 and 2003. Given the small population size of the Scottish deerhound and the ascertainment process, it is probable that the majority of affected dogs (i.e., transmission events) were observed, which argues for more reliable segregation ratios, penetrance estimates, and transmission parameters.

Elston et al. [31] developed three criteria that must be satisfied before accepting a major genetic locus model. (1) The data must show that a mixture of two distributions fits the data significantly better than one distribution. (2) Statistical comparison of the general and the Mendelian models must be nonsignificant. (3) Comparison of the general and environmental models must be significant. Variance components analysis (Table 2) demonstrated the first criterion by showing that a mixture of two distributions (Model 3) described the phenotypic variance in this population significantly better than one distribution (Model 2). The second and third criteria were fulfilled by segregation analysis; the likelihood ratio test between the general model and the Mendelian models is nonsignificant and the environmental model was statistically rejected against the general model.

By fulfilling the Elston criteria, this population of dogs represents only the second large-animal model for an inherited cancer syndrome. The first documented heritable cancer syndrome, renal cystadenocarcinoma (RCND), is seen almost exclusively within the German shepherd dog [32]. The RCND phenotype appears to segregate in an autosomal dominant fashion with high penetrance and is caused by mutations within the canine homologue of the human Birt-Hogg-Dube gene [33]. Other likely heritable cancer syndromes seen in large animals include osteosarcoma in the Rottweiler [12], malignant histiocytosis in the Bernese mountain dog [34], and lymphoma in a variety of breeds [35]. While both osteosarcoma in the Rottweiler and lymphoma in certain breeds represent excellent spontaneous models for the homologous disease in humans, no studies to date have documented the mode of inheritance or genetic basis for any of these heritable malignancies.

The osteosarcoma phenotype segregating within this deerhound population shows strong similarity to the human counterpart of primary bone osteosarcoma. In addition, the trait exhibits a well-defined Mendelian inheritance, prescribed risk factors, and a rapid clinical course among affected dogs. This deerhound population thus represents an excellent (spontaneously occurring) large-animal model of human osteosarcoma. Significant public interest has focused on the use of such companion animal models for the characterization of novel therapeutics [36] and the delineation of genetic mechanisms underlying cancer. Dogs are uniquely suited to such studies because they share a common environment and have shorter life expectancies than their human companions. The shorter life expectancy would of course allow for more rapid endpoint determination for any interventional study. Genetic analysis of this population of deerhounds, through linkage and/or association studies, will thus be extremely useful for studying the molecular basis of this malignancy in both dogs and humans.

Materials and methods

Sample population

Dogs used in this study were all privately owned and AKC registered Scottish deerhounds ranging from 6 months to 15 years of age. Pedigrees were constructed around 5 distantly related (more than three generations separated) affected dogs ascertained as part of a prospective cohort design deerhound health study [5]. Health information on all first-degree relatives of these 5 probands was ascertained. A sequential-sampling method [14] was then used to

extend each proband-centered pedigree to include ~200 dogs. Resultant pedigrees contained approximately four generations of AKC-registered ancestors of the original 5 probands. Through this combination of fixed and sequential sampling, information on a total of 1057 dogs alive between the years 1975 and 2003 was collected. Based on covariates (age, gender, height, and weight), this population did not differ significantly from a previously described randomly collected group of deerhounds [5]. Using AKC pedigree information, this entire collection of 1057 deerhounds was grouped into one large nondisjoint pedigree.

Pedigree information was obtained from either owners or breeders and confirmed using the AKC pedigree certificates. Current and historic health information on all dogs was collected from owners, breeders, and veterinarians. Affected status was determined by the presence of characteristic radiographic findings of aggressive bone disease and/or histologic examination. Pedigree information and medical histories were entered into a pedigree management program (Cyrillic 2.1.3; Cherwell Software). For construction of comprehensive pedigrees, additional dogs who, in some cases, lacked complementary health information were included.

Health information collected included a variety of qualitative and quantitative traits (i.e., covariates). Qualitative traits included OSA status, presence/absence of other diseases including cancers, and gender. OSA status had three possible outcomes: affected, unaffected, or unknown. Quantitative traits included age at exam, age of onset, height, weight, and length of time sexually intact (as a measure of cumulative sex hormone exposure). Age at exam was considered to be the age at last known contact or age at death where applicable. Age of onset documented the age at which definitive diagnosis of osteosarcoma was made in affected individuals. Time was measured in years, height in inches at the shoulder, and weight in pounds. Other information collected (where applicable) included cause of death, tumor location, and treatment option.

Statistical analyses

Population descriptors, estimates of heritability, segregation analysis, and linear modeling were obtained from analysis of the deerhound population in which osteosarcoma segregates. Osteosarcoma was considered a dichotomous trait in which individuals were classified as "unaffected" (at the time of exam) or "affected" or, in some cases, of "unknown status." Disease phenotypes of individuals with traits listed as "unknown" were ignored for each analysis. Furthermore, individuals with missing covariates did not contribute to covariate analysis. To adjust for the nonrandom nature of proband selection, ascertainment correction was applied by conditioning the sample likelihood on proband phenotype as described in each software package [15,16].

Descriptive analysis

The program PEDINFO was used to determine descriptive characteristics of the population. This program is available as a component of the genetic analysis package SAGE 5.2.0 [17]. Inbreeding coefficients were estimated using the Kinship Matrices option of the genetic software MENDEL 6.01 [16].

Heritability

Heritability (h^2) in the narrow sense represents the proportion of total (normally distributed) phenotypic variation that is due to genetics, or $h^2=\sigma^2_a/(\sigma^2_a+\sigma^2_e)$, where σ^2_a is the additive (polygenic) variance and σ^2_e is the variance due to individual-specific environmental effects [18]. The dominant (nonadditive) effects of a single allele can be accommodated by adding a dominance variance (σ^2_d) to the above model. Estimation of heritability and the individual components was performed by the method of variance components analysis.

Extending this method to large complex pedigrees and dichotomous (binomially distributed) phenotypes, as described herein, can be done by assuming that each individual belongs to a specific affection status if an underlying genetically determined risk (i.e., liability) exceeds a certain threshold. This latent "liability" is assumed to have an underlying multivariate normal distribution. In addition to the variance components, covariate effects were included in the model. Both variance components and covariate effects were estimated, simultaneously, by maximum likelihood techniques. Binary covariates were constrained to sum to zero; quantitative covariates had no such restraint. Calculations were performed using an enhanced version of the variance component program FISHER [19] as implemented in the heritability functionality of MENDEL 6.01 [16].

Segregation analysis

Complex segregation analysis of the osteosarcoma trait was carried out using the computer program jPAP version 1.4.0 beta [15]. A complete description of this method using the PAP package has been described [20]. This program is currently the only available program for the segregation analysis of discrete (qualitative) traits that allows the presence of inbreeding loops. Computational time, however, increases dramatically as the complexity of the pedigree increases. The complexity of this pedigree necessitated its division into the five originally ascertained pedigrees for both segregation analysis and linear modeling. This grouping was not a genetic or statistical requirement; rather a computational requirement, due to the known limitations of currently available software [15–17]. The effects of this necessary pedigree division were expected to make the detection of a major locus more difficult due to the loss of genetic information contained within the disrupted relationships. The magnitude of this effect cannot be estimated but was assumed to be minor.

In segregation analysis, five single-locus transmission models were evaluated, including Environmental, Mendelian dominant, Mendelian recessive, Mendelian codominant, and Major locus arbitrary transmission (general model). For completeness, a No major gene (sporadic) model was also included. The five transmission models are postulated by allowing the presence of three observable types with three corresponding transmission parameters. The observable genotypes (or more properly "ousiotypes") at this single major locus were considered to be products of two alleles, A and B. Allele A is considered the disease-associated allele, while allele B is considered the wild-type allele. Six gender-specific transmission probabilities ($\hat{o}_{AA}$, $\hat{o}_{AB}$, $\hat{o}_{BB}$), which represent the probability that an individual of a given genotype will transmit the A allele to his/her offspring, were assumed. Allele frequencies (q_A and q_B) and six gender-specific penetrances (β_{AA}, β_{AB}, and β_{BB} for females and males individually) were estimated. The jPAP software, as a component of its genetic model, allows for variations within the normal densities of each genotype (i.e., a mixed model). These additional within-genotype parameters were modeled as variance components and were estimated for the general model.

Each model of transmission stated above was created by placing appropriate restrictions on the ousiotypes and/or the transmission probabilities. The Environmental model is the most restricted model and assumes no transmission of major gene effect and the presence of three normally distributed ousiotypes. This model is generated by equating the allele frequency with a common transmission probability ($\tau_{AA}=\tau_{AB}=\tau_{BB}$). The Mendelian models assume the presence of a major locus, multiple (two or three) ousiotypes, and transmission probabilities fixed at their Mendelian expectations ($\tau_{AA}=0.0$, $\tau_{AB}=0.5$, $\tau_{BB}=1.0$). Dominant, recessive, and codominant expressions in the Mendelian model are created by placing appropriate restrictions on the penetrances. Codominant expression allows for three different genotypes that are not necessarily additive. The unrestricted single-locus model (general model) includes three normally distributed ousiotypes, assumed to occur in Hardy–Weinberg proportions, and transmission probabilities estimated with the restriction of generational equilibrium of trait distribution. This restriction of generational equilibrium is, of course, an assumption of the Mendelian and Environmental models and corresponds to

$$p = p^2\tau_1 + 2pq\tau_2 + q^2\tau_3.$$

A more general model with arbitrary transmission parameters failed to converge in these data. Finally, in the sporadic model, disease penetrance is not influenced by ousiotype and thus all individuals have the same inherent gender-specific risk for disease development. Maximum likelihood procedures were

used to estimate parameters, standard errors, and log-likelihoods for each model.

Generalized linear modeling

While the program package jPAP does include a within-genotype parameter, it does not allow for the direct inclusion of individual covariates that modify traits. Therefore, to determine the effects of individual covariates on phenotype, a generalized linear model [21] was used. This model is implemented in the Penetrance Estimation option of the computer software program MENDEL 6.01 [16]. A generalized linear model allows for the estimation of affection probability. Assumptions include the presence of a set of predictors that modify affection, a binomial distribution of the dependent variable (affected status), and Mendelian transmission of genotype. Under a binomial distribution, the basic formula for this regressive model is

$$\mu_{AA} = e^{y}/(1 + e^{y}),$$

where μ_{AA} is the affection probability of an individual with genotype AA and y is a linear equation composed of an intercept (GRAND) and parameters for each independent covariate. Maximum likelihood procedures are used to estimate the intercept and coefficients for each covariate. Tests for genetic control can be implemented through the addition of two or more covariates to represent genotypes (AA, AB, BB), with appropriate constraints on their boundaries to reflect recessive, dominant, or codominant (genotypic) transmission. Comparison of the models with and without covariates allows for hypothesis testing.

Hypothesis testing

The Student t test was used to compare population means of descriptive data. The LRT was used to test each submodel, in heritability analysis, segregation analysis, and logistic regression, against the more general models. This ratio is computed as minus twice the natural log-likelihood of the general model subtracted from that for a restricted submodel, $-2\ \ln(L_1 - L_0)$. This difference is typically asymptotically distributed as a χ^2 with degrees of freedom equal to the difference in the number of independent parameters estimated in the two models. Another method to compare models uses AIC, defined as AIC $= -2\ \ln(L) + 2$(number of parameters estimated). The most parsimonious model has the minimum AIC value [22].

Acknowledgments

Special thanks go to Kenneth Lange (UCLA) for his help with the MENDEL software and his tireless advice at several stages throughout this project. Also, thanks to Arnold Saxton (University of Tennessee) for his thorough review of the manuscript, to Sandy Hasstedt (University of Utah) for her help with the jPAP software, and finally to Matthew Breen (NCSU) for his advice early in the planning phase of this project.

References

[1] S.J. Withrow, Osteosarcoma, Vet. Q. 20 (1998) S19–S21.
[2] U.S. Cancer Statistics Working Group. United States Cancer Statistics: 1999–2002 incidence and mortality Web-based report version. Atlanta: Department of Health and Human Services, Centers for Disease Control and Prevention, and National Cancer Institute; 2005.
[3] W. Misdorp, A.A. Hart, Some prognostic and epidemiologic factors in canine osteosarcoma, J. Natl. Cancer Inst. 62 (1979) 537–545.
[4] G. Ru, G. Terracini, L. Glickman, Host related risk factors for canine osteosarcoma, Vet. J. 156 (1998) 31–39.
[5] J. Dillberger, Deerhound health survey: Part 1. Lifespan and causes of death, Claymore (1996; Mar/Apr) 36–44.
[6] G. Letson, C. Muro-Cacho, Genetic and molecular abnormalities in tumors of the bone and soft tissues, Cancer Control 8 (2001) 239–251.
[7] B. Ragland, W. Bell, R. Lopez, G. Siegal, Cytogenetics and molecular biology of osteosarcoma, Lab. Invest. 82 (2002) 365–373.
[8] S. Mendoza, T. Konishi, W. Dernell, S. Withrow, C. Miller, Status of the p53, Rb, and MDM2 genes in canine osteosarcoma, Anticancer Res. 18 (1998) 4449–4453.
[9] R.A. Levine, T. Forest, C. Smith, Tumor suppressor PTEN is mutated in canine osteosarcoma cell lines and tumors, Vet. Pathol. 39 (2002) 372–378.
[10] M.N. Mullins, S.E. Lana, W.S. Dernell, G.K. Ogilvie, S.J. Withrow, E.J. Ehrhart, Cyclooxygenase-2 expression in canine appendicular osteosarcomas, J. Vet. Intern. Med. 18 (2004) 859–865.
[11] R.A. Tjalma, Canine bone sarcoma: estimation of relative risk as a function of body size, J. Natl. Cancer Inst. 36 (1966) 1137–1150.
[12] D.M. Cooley, B.C. Beranek, D.L. Schlittler, N.W. Glickman, L.T. Glickman, D.J. Waters, Endogenous gonadal hormone exposure and bone sarcoma risk, Cancer Epidemiol. Biomarkers Prev. 11 (2002) 1434–1440.
[13] G.J. Spodnick, J. Berg, W.M. Rand, et al., Prognosis for dogs with appendicular osteosarcoma treated by amputation alone: 162 cases (1978–1988), J. Am. Vet. Med. Assoc. 200 (1992) 995–999.
[14] C. Cannings, E. Thompson, Ascertainment in the sequential sampling of pedigrees, Clin. Genet. 12 (1977) 208–212.
[15] S.J. Hasstedt, jPAP: document-driven software for genetic analysis, Genet. Epidemiol. 29 (2005) 255.
[16] K. Lange, R. Cantor, S. Horvath, et al., MENDEL version 4.0: a complete package for the exact genetic analysis of discrete traits in pedigree and population data sets, Am. J. Hum. Genet. 69 (2001) A1886 (Suppl.).
[17] SAGE. Statistical Analysis for Genetic Epidemiology. Computer program package available from Statistical Solutions Ltd., Cork, Ireland, release 5.2.0, 2006.
[18] P. Le Roy, J.M. Elsen, Simple test statistics for major gene detection: a numerical comparison, Theor. Appl. Genet. 83 (1992) 635–644.
[19] K. Lange, M. Boehnke, D.E. Weeks, Programs for pedigree analysis: MENDEL, FISHER, and dGENE, Genet. Epidemiol. 5 (1988) 471–472.
[20] S.H. Juo, T.H. Beaty, J. Xu, V.L. Prenger, J. Coresh, P.O. Kwiterovich Jr., Segregation analysis of two-locus models regulating apolipoprotein-A1 levels, Genet. Epidemiol. 15 (1998) 73–86.
[21] G.E. Bonney, Regressive logistic models for familial disease and other binary traits, Biometrics 42 (1986) 611–625.
[22] H. Akaike, A new look at the statistical model identification, IEEE Trans. Autom. Control 19 (1974) 716–723.
[23] American Kennel Club Registration Statistics, 1975–2003, AKC, New York, 1975–2003.
[24] J. Terwilliger, K. Weiss, Confounding, ascertainment bias, and the blind quest for a genetic 'fountain of youth, Ann. Med. 35 (2003) 532–544.
[25] S.G. Self, K.Y. Liang, Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions, J. Am. Stat. Assoc. 82 (1987) 605–610.
[26] J.D. Dubosson, D. Klein, J. Pettavel, C.H. Rey, Familial cancer syndrome studies in 4 generations of a family, Schweiz. Med. Wochenschr. 107 (1977) 875–881.
[27] H.T. Lynch, R.M. Fusaro, J. Pester, J.F. Lynch, Familial atypical multiple mole melanoma (FAMMM) syndrome: genetic heterogeneity and malignant melanoma, Br. J. Cancer 42 (1980) 58–70.
[28] D.E. Cornings, B.G. Cornings, Evidence for an X-linked modifier gene affecting the expression of Tourette syndrome and its relevance to the increased frequency of speech, cognitive, and behavioral disorders in males, Proc. Natl. Acad. Sci. USA 83 (1986) 2551–2555.
[29] C. Cannings, E. Thompson, Sampling schemes and ascertainment. In CF Sing, M Skolnick, The Genetic Analysis of Common Diseases: Applications to Predictive Factors in Coronary Heart Disease. Liss, New York, 1979sd.
[30] E. Ginsburg, I. Malkin, R. Elston, Sampling correction in pedigree analysis, Stat. Appl. Genet. Mol. Biol. 2 (2003) Article 2.
[31] R.C. Elston, K. Namboodiri, C. Glueck, R. Fallat, R. Tsang, V. Leuba, Study of the genetic transmission of hypercholesterolemia and hypertriglyceridemia in a 195 member kindred, Ann. Hum. Genet. 39 (1975) 67–87.

[32] T. Jonasdottir, C. Mellersh, L. Moe, et al., Genetic mapping of a naturally occurring hereditary renal cancer syndrome in dogs, Genetics 97 (2000) 4132–4137.
[33] F. Lingaas, K. Cornstock, E. Kirkness, et al., A mutation in the canine BHD gene is associated with hereditary multifocal renal cystadenocarcinoma and nodular dermatofibrosis in the German shepherd dog, Hum. Mol. Genet. 12 (2003) 3043–3053.
[34] G.A. Padgett, B.R. Madewell, E.T. Keller, et al., Inheritance of histiocytosis in Bernese mountain dogs, J. Small Anim. Pract. 36 (1995) 93–98.
[35] J.F. Modiano, M. Breen, R.C. Burnett, et al., Distinct B and T cell lymphoproliferative disease prevalence among dog breeds indicates heritable risk, Cancer Res. 65 (2005) 5654–5661.
[36] G.S. Mack, Cancer researchers usher in dog days of medicine, Nat. Med. 11 (2005) 1018.