An unidentified gene regulates body size in toy poodles in addition to GHR, STC2, SMAD2, IGF1R

Rui Nakata

Abstract

Body size is a complex trait that is determined by a combination of genes. In dogs, it has been determined that between two to six QTLs contain about 70% of the genes that regulate body size. So far the six genes GHR, STC2, SMAD2, IGF-1, IGF-1R, and HMGA2 have been discovered to determine body size in dogs. When focusing on the genotype combinations of toy poodles at these genes, there is evidence of some variation in body size within the same genotype combinations. Therefore, we hypothesized that another unidentified gene that contributes to body size determination is present in this breed. In this study, methods were taken to identify this gene of interest. Toy poodles were genotyped at five loci of interest (IGF-1R, GHR(1), GHR(2), SMAD2, STC2), and this data was used to create a multiple linear regression model that uses the genotypes at these loci to predict body size. Using this equation, each dog was categorized into one of two groups: dogs we predict that are affected by the unidentified gene and dogs that are not. Next, an association scan of the CanMap dataset was used to look for additional genes that contribute to tiny size (with a breed average height of 10 inches or less). A strong association peak was found near the C-terminus region of the LEMD3 gene ($p = 1.1 * 10^{-52}$). Further investigation of this gene using fine-mapping and a GWAS using the two newly categorized subgroups will help identify new genes of interest and also verify the significance of the LEMD3 gene.

Introduction

Because body size is a complex trait, there are a significant number of genes that contribute to body size in humans. In a GWAS study conducted on 129,000 individuals, about 180 loci have been identified to be associated with body size but this only accounts for about 12% in determining the actual height (Lango et al., 2010). Therefore, it is considered very difficult to account for every causative gene that relates to this phenotype and when investigating potential genes involved in body size, many samples will be needed to get statistically significant results (Hoopes et al., 2012). In contrast, dogs are great model organisms when researching body size because of the variety of physical traits they possess when comparing different breeds, and this is a result of selective breeding by humans that started about 100,000 years ago (Vila et al., 1997). Because breeders focus on selecting dogs by certain physical factors, this increases the diversity between breeds but reduce diversity within (Rimbault et al., 2013). Some purebred dogs have whither height measures that may range up to four fold when compared to other breeds (Sutter et al. 2008).

Chase et al suggested that the fact that these dogs adapted so quickly is a result of a few genetic loci that control physical features. They found multiple quantitative trait loci associated with body size by examining the genotypes of Portugese Water Dogs. Another study that investigated 60,968 SNPs identified 51 regions of the genome that are associated with body structure such as average breed body size and bone shape by generating a high-density map of genetic variation. In this study, it was concluded that only two to six QTLs explain more than 70% of the variance in morphological traits in dogs (Bokyo et al., 2010).

The 6 different genomic association signals that they found to be significant correspond to the locations of specific genes. The first signal corresponds to the gene IGF1, which codes for the insulin like growth factor 1. This was also shown to be a causative gene for big body size in dogs (Sutter et al., 2007). The second signal corresponds to the gene HMGA2, which was known to correlate with body size in humans and mice. The third signal corresponds to the gene SMAD2 which is also known to correlate with body size in dogs. The fourth signal corresponds to the gene STC2, which is known to be a growth inhibitor in mice. The last two signals were on the X chromosome (Bokyo et al., 2010). Jones et al also found six significant quantitative trait loci associated with regulation of body size by using SNP markers (Jones et al., 2008). By conducting a GWAS study using the CanMap project data, Hoopes et al found a nonsynonymous SNP at the gene that codes for the IGF1 receptor protein to also regulate dog body size. Considering these past studies and incorporating more fine-mapping, Rimbault et al concluded that polymorphisms at six loci (GHR, HMGA2, STC2, SMAD2, IGF1, IGF1R) are sufficient to explain about half of the variance in dog size (Rimbault et al., 2013).

These six genes operate in different ways to regulate body size. GHR is a gene that encodes for the growth hormone receptor. The growth hormone receptor protein is a transmembrane protein that associates with the JAK2 tyrosine kinase in humans. Binding of growth hormone to its receptor causes dimerization of the GHR which phosphorylates JAK2 and starts the JAK2 signal transduction (Postel-Vinay et al., 1995). In dogs, the growth hormone and growth hormone receptor have been shown to have important roles in regulation of cell proliferation. The lack of GHR activation decreased the body size of mice by 50-60% when they were 3 weeks old, which indicates that GHR is an important factor in determining postnatal longitudinal growth (Efstratiadis, 1998).

STC2 is a gene that encodes for the stanniocalcin 2 protein. Stanniocalcin 2 is a homodimeric glycoproteins that is upregulated by stress and regulates bone development in mice. Mice that express STC2 show signs of intra-uterine growth restriction and post-natal growth retardation (Johnston et al., 2010). It is independent of the GH/IGF1 pathway and causes developmental delay in intramembranous bone (Gagliardi et al., 2005).

SMAD2 is a gene that encodes for SMAD family member 2. SMAD proteins are known as signal transducers and transcriptional modulators. They also function as a transforming growth factor-beta (TGF-beta) signaling transducer, which regulates cell proliferation and differentiation (Lin, 2013). Kim et al concluded that SMAD2 polymorphisms were significantly associated with bone marrow density and bone metabolism in a study done on women due to its role in regulating the development of osteoclasts and osteoblasts (Kim et al., 2011).

HMGA2 is a gene that encodes for the high mobility group AT-hook 2 protein which belongs to the non-histone chromosomal high mobility group protein family. HMGA2 in humans is known to be associated with prenatal growth and postnatal stature. A SNP in this gene is also associated with head circumference, which is another indicator that this gene is linked to growth (Horikoshi et al., 2013). Weedon et al also observed the HMGA2 gene, and found that mutations in this gene alter body size in both mice and humans including children (Weedon et al., 2007).

IGF-1 is a gene that encodes for the insulin like growth factor 1. This protein is a circulating hormone that is produced in the liver and can act in endocrine, autocrine, and paracrine pathways. IGF-1 also acts as a downstream target of growth hormone. In mice, overexpression of IGF-1 in serum was associated with an increase in body mass and skeletal size (Elis et al., 2010). Sutter et al. focused on a QTL that was predicted to affect body size on chromosome 15 located using a GWAS. In the IGF-1 gene, they found a SNP haplotype present in small dog breeds and absent in large breeds (Sutter et al., 2007).

IGF1R is a gene that encodes for insulin like growth factor 1 receptor. It binds to the insulin like growth factor protein that acts as a signaling molecule. In humans, under-expression of IGF1R inhibited cell proliferation by causing apoptosis and cell cycle arrest (Dong et al., 2013). In addition, the IGF-1 receptor was proven to be essential for proper growth in humans in a study that focused on a short statured family with postnatal growth failure (Netchine et al., 2009). Although most small dogs have the same haplotype for the insulin like growth factor gene, the IGF1 receptor gene contributes additional factors for the variation within those small dogs (Hoopes et al., 2012).

Rimbault et al looked at six autosomal loci that were associated with body size and found highly associated variants in each locus. Because the ancestor of all dogs is the gray wolf, smaller dogs are thought to have evolved at a later time compared to larger dogs. Therefore, to determine the ancestral allele at each SNP, Rimbault et al genotyped various gray wolves, red wolves and coyotes. When the genotypes for multiple dogs with various sizes were observed, larger dogs always possessed more ancestral alleles whereas smaller dogs always possessed more derived alleles. Two nonsynonymous SNPs in GHR (ancestral allele G / derived allele A , Ancestral allele C / derived allele T), one SNP 20 kb downstream from STC2 (ancestral allele T / derived allele A), one SNP in the 5' UTR of HMGA2 (ancestral allele G, derived allele A), and a 9.9 kb deletion 24 kb downstream from SMAD2 were located. In combination with the SNP found by Hoopes et al in the IGF1R gene (ancestral allele G, derived allele A), we can focus on these variants to be the major factors in determining body size in toy poodles. From further investigation, Rimbault et al concluded that in all variants (IGF1, GHR(1), GHR(2), SMAD2, STC2, HMGA2, IGF1R), the standard body weights of dogs that are homozygous for the ancestral allele are generally higher than the ones that are heterozygous, and the heterozygous are generally heavier than the ones that are homozygous for the derived allele. Both the number of derived alleles and the allele frequency increased as the overall body size decreased in these dogs. Rimbault et al concluded that other variants do not exist in the exons in the genes studied, but there is a possibility that other markers that have regulatory effect are present (Rimbault et al., 2013).

Data from the Rimbault study show that the dogs that have a smaller body size have mostly derived alleles in the loci whereas the dogs that are larger usually have less derived alleles. There is evidence of some miniature poodles and toy poodles that have the same genotype combination at the 6 genes but differ in body size. This shows that there may be another unidentified gene separate from the six already defined genes that allows for these toy poodles to be smaller than the predicted size by genotype combination.

In this study, we focused on toy poodles and miniature poodles in an attempt to find a new gene that regulates body size. We also focused on the five genes (IGF1R, GHR(1), GHR(2), SMAD2, STC2) in this study and disregarded genes that were most likely conserved in all miniature and toy poodles. Because not much is known about the STC2 or SMAD2 pathways, we predicted that this mystery gene is regulating or interfering with these pathways. The mystery variant may also possibly in a transcription factor and thus change the way the protein functions.

To find this unidentified gene, we genotyped 43 dogs at the five genes of interest. Using this data, we formulated a method to predict and categorize if the unidentified gene affected each dog's body size or not. We also used the CanMap data (Boyko et al., 2010) and the subcategories determined by Hoopes et al ("tiny" and "control") to find a potential gene associated with SMAD2 or STC2.

Methods

Sample Collection

Samples were collected or obtained by breeders by swabbing the interior cheeks of the dogs. Genomic DNA were isolated out of these samples using the protocol from the QIAamp DNA Micro kit (modified from QIAgen). In addition to obtaining the DNA samples, each dog was measured. The resource for breed standards by the American Kennel Club was used to estimate the average height at the withers in inches.

Genotyping

These various toy poodles were genotyped at the four genes that are of importance: IGF1R, STC2, SMAD2, and GHR. For the four genes IGF1R, STC2, and GHR(1), GHR(2), PCR products were made and the alleles at the position of the SNP were examined after sequencing. For each gene, a specific primer was designed so that the amplified PCR products included the position of the SNP. The program Primer3 was used to design the primers. In order to sequence the genes at the specific location, about 10 ng of genomic DNA was used to make PCR products using AmpliTaq Gold DNA polymerase and buffer by Applied Biosystems. The PCR ran for 20 touchdown cycles with annealing temperatures between 61 °C and 51 °C and dropping 0.5 °C for each cycle. To ensure product is properly made, the PCR products were tested using gel electrophoresis on a 1.5% agarose with ethidium bromide gel. The primers were then removed from the PCR products by treating it with exonuclease I and shrimp alkaline phosphatase at 37 °C for 30 minutes. This was then inactivated by incubating at 80 °C for 15 minutes. These products were sequenced in both directions (except for STC2 which was only sequenced in the forward direction). For the sequencing, the same primers were used in the beginning step in addition to the BigDye 3.1 Applied Biosystems master mix. The amplified sequencing products were cleaned using ethanol precipitation and suspended in 15 µl of water and analyzed using the ABI3730 capillary sequencer.

The data obtained from the sequencer was assembled and analyzed using the Sequencher program. The position of the SNP was observed and the result (homozygous or heterozygous, ancestral or derived) was recorded for each dog sample.

To sequence the SMAD2 gene, we amplified the gene with 10 ng of genomic DNA to make PCR products using the same method described above. We also amplified the DNA of control dogs along with the toy poodle samples of interest. These products were run on an electrophoresis gel to determine the sizes of the products. The resulting bands were used to determine if the DNA was intact or if there was a deletion.

Combining the data previously obtained by Hoopes et al with the additional dogs that were genotyped, data for a total of 43 dogs were obtained.

Data Analysis

The collected data was used to compare the relationship between body size and amount of derived alleles. For each gene, the number of derived alleles (0 (AA), 1 (AD), or 2 (DD)) were recorded. Visualization of the combination of genotypes was done in Microsoft Excel. Construction of the multiple linear regression equation was conducted using the program IBM SPSS Statistics.

Categorization

The toy poodles that are categorized to possess the mystery gene were introduced into the plot comparing actual body size and the calculated body size. A boundary was set to categorize the dogs into the two groups again and this process was repeated until the R-squared value of the multiple regression equation was minimized.

UCSC Genome Browser

By using the genotype data from the CanMap Project, a previously conducted GWAS (Boyko et al., 2010), we looked for a potential candidate for the mystery gene. The dogs were categorized as either "tiny" or "control" as was done by Hoopes et al. The "tiny" breeds were defined as those that were smaller than 10 inches at the withers and included Brussels Griffons, Cairn Terriers, Chihuahuas, Havaneses, Norwich Terriers, Papillons, Pomeranians, Toy Poodles, and Yorkshire Terriers. All other breeds were put in the "control" group (Hoopes et al., 2012). The genome category was set to the dog (canis lupus familiaris) with the assembly set for May of 2005. By going through each of the genes that interact with the STC2 gene and the SMAD2 genes in humans, we predict to find the potential mystery gene in the toy poodles. If a SNP at a location in one of these genes appears to have a P- value of 8 or greater when comparing between the category of dogs with the mystery gene and without, the gene that possesses this SNP can be considered a possible candidate for the mystery gene.

Results

Allele data

To collect data for this study, we genotyped DNA from 43 miniature and toy poodles with body sizes ranging from 9 inches to 15.75 inches at the 5 genes of interest: SMAD2, IGF1R, STC2, GHR(1), and GHR(2). The heights at the withers were averaged for each genotype. These averages reflect the body size differences where dogs with more derived alleles are generally smaller than dogs with ancestral alleles.

	SMAD2	IGF1R	STC2	GHR(1)	GHR(2)
DD	9.8	10.7	11.1	10.9	NA
AD	11.2	11.6	10.6	14.5	9.4
AA	11.2	10.5	11.2	NA	11.1

Figure 1.
Average body size in inches categorized by the genotypes at each gene. DD indicates dogs homozygous for the derived allele, AD represents dogs that are heterozygous, and AA represents dogs homozygous for the ancestral allele. In general at each gene, dogs that possess more derived alleles have a smaller body size than dogs that possess no derived alleles.

To simplify and visualize the combination of genotypes of all dogs, the alleles at the five genes were color-coded. For most of the dogs, the allele combinations do not differ significantly from each other. The most common combination of alleles is SMAD2: AD, IGF1R: DD, STC2: AA, GHR(1): DD, GHR(2): AA with a frequency of 7 dogs. Their body sizes are 10.25, 10.5, 10.5, 11, 11.5, 11.5, and 13.5 inches. The second most common combination is SMAD2: AD, IGF1R: DD, STC2: AD, GHR(1): DD, GHR(2): AA with a frequency of 5 dogs. The third most common combination is SMAD2: AD, IGF1R: AD, STC2: AD, GHR(1): DD, GHR(2): AA with a frequency of 4 dogs. Out of all 243 combinations possible, we only observed 18 combinations. This results from the fact that only two out of three combinations of alleles were observed in three of the genes. There were no dogs at the STC2 gene that were homozygous derived and GHR(1) and GHR(2) are mostly fixed at homozygous derived and homozygous ancestral respectively.

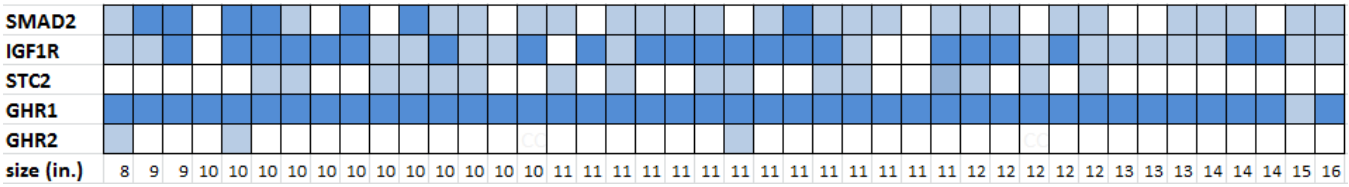

Figure 2.
Combinations of the alleles observed at each gene for all dogs. The homozygous derived alleles (DD) are colored blue and the heterozygous alleles (AD) are colored light blue. The dogs were sorted by increasing size and size was rounded to the nearest whole number. Although the size of the dogs varies almost two fold, the combinations of the genes are all very similar to one another.

Linear Multiple Regression

From the collected genotype data, each genotype was assigned a number in an effort to quantify the data numerically. The homozygous derived alleles were collectively assigned the number 2 to represent the two derived alleles, the heterozygous alleles were assigned the number 1, and the homozygous ancestral alleles were assigned the number 0. This was done for all five genes in each dog, and the numbers were summed per dog.
Because each gene contributes differently from one another to determine body size, a statistics program (SPSS) was used to predict a model that would correct for these differences. SPSS was used to create a linear regression model that accounts for the contribution difference of each of the five genes. The data for the number of derived alleles for each gene was used for this analysis. Because we hypothesized that the mystery gene only affects dogs with a smaller body size making them smaller than predicted from the genes that have already been determined to affect body size, we assumed that dogs taller than 11.5 inches do not possess the mystery gene. We only used data from these dogs to accurately predict a linear regression model therefore minimizing the probability of the mystery gene affecting the equation. However, because all dogs chosen in this way had the same alleles for GHR(2), this gene was disregarded in the equation. This linear regression equation was used to predict the body size of the dogs from the genotypes (number of derived alleles).

These variables predicted body size, $F(4, 7) = 1.612$, $p = 0.272$, $R^2 = .480$ but was not statistically significant.

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	16.897	2.737		6.173	.000
	SMAD2	-.466	.912	-.162	-.510	.626
	IGF1R	-.750	.785	-.287	-.956	.371
	STC2	-1.793	1.012	-.519	-1.772	.120
	GHR(1)	-1.181	1.405	-.254	-.840	.428

Table 1.
The SPSS program was used to predict the multiple linear regression equation for dogs that were categorized to not have the mystery gene (body size of 11.5 inches or higher). The equation was calculated to be **Body size = 16.897 - 0.466(SMAD2) - 0.750(IGF1R) - 1.793(STC2) - 1.181(GHR).** These coefficients represent the degree of contribution of each derived allele at each specific gene. GHR(2) was not considered in the calculation because all dogs bigger than 11.5 inches were homozygous for the ancestral allele. This equation was used to predict the body size of all dogs. The R squared value for this equation was 0.48.

Using the linear regression equation and the number of derived alleles for each gene, we obtained a value for the predicted body size for each dog. This predicted value and the measured value were plotted and a line of best fit was constructed. Because this equation for the line of best fit was constructed using the dogs that were bigger than 11.5 inches, dogs that are not affected by the mystery gene will fit closely to the line.

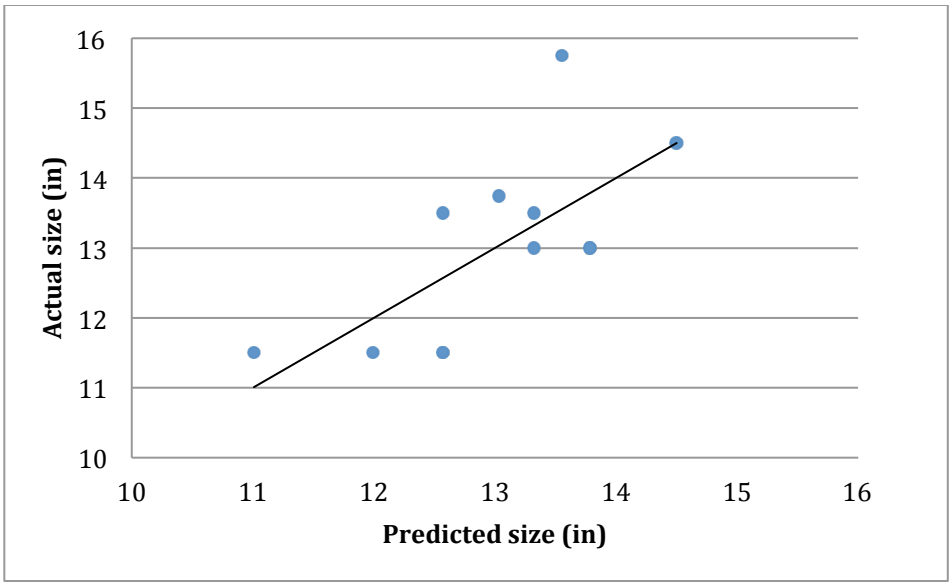

Figure 3.
A linear model was generated using dogs that were categorized to not possess the mystery gene. The linear regression equation and the genotype data were used to calculate the predicted size in inches. The actual body size of the dogs that were 11.5 inches or bigger are represented on the y-axis and the predicted body size of these dogs are represented on the x-axis. A line of best fit was calculated to be $y=x-0.0008$ ($R^2=0.4795$).

This linear model was compared with all dogs of all sizes. Some of the dogs from the additional 33 dogs that were introduced (smaller than 11.5 inches) digressed greatly from the linear model. Because the dogs that fit the linear model well were determined as dogs with perfect equality between the predicted size and the measured size, the dogs that digressed from the linear model were considered to be affected by the unidentified gene. This digression was quantified by calculating the residuals for each dog.

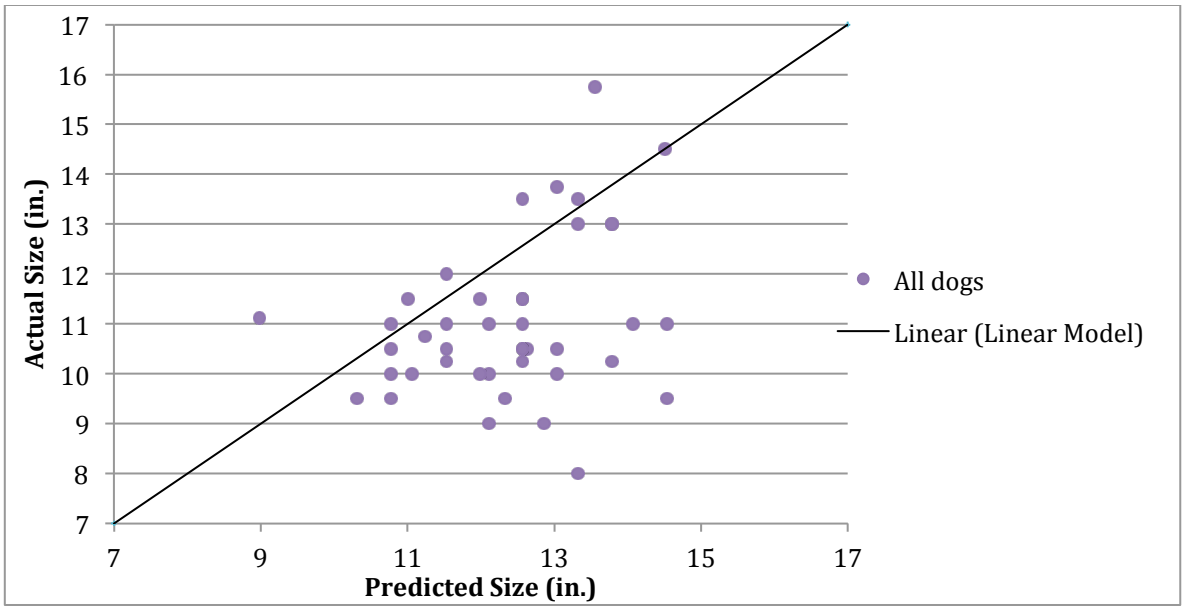

Figure 4.
A scatter plot of all dogs representing the measured body size and the fitted body size in relation to the linear model calculated by the data from the dogs taller than 11.5 inches. The linear model indicates perfect equality between the predicted size and the measured size. The dogs with the actual size digressed more from the linear model and are simultaneously far below the line are more likely to be affected by the mystery gene of interest.

Categorization

The residuals were calculated and sorted in order from lowest to highest. Because the mystery gene affects the actual size by making the dogs smaller than the size is predicted using only the genotypes from the already determined genes, the dogs with the lowest residual values have the highest probability of possessing the mystery gene.

<-3		<-2		<-1		<0		>0	
number	residual	number	residual	number	residual	number	residual	number	residual
306	-5.3182	225	-2.8352	246	-1.9912	245	-0.8092	329	0.0008
210	-5.0342	235	-2.5342	229	-1.5682	338	-0.7842	253	0.1818
204	-3.8522	313	-2.3182	212	-1.2752	340	-0.7842	244	0.2248
261	-3.5342	217	-2.1247	312	-1.2752	310	-0.7752	251	0.4748
262	-3.5342	232	-2.1022	231	-1.1022	247	-0.5252	221	0.4918
218	-3.1022	311	-2.0682	241	-1.0682	226	-0.4912	255	0.7158
205	-3.0682	314	-2.0682	315	-1.0682	264	-0.4912	328	0.9318
252	-3.0682			263	-1.0592	341	-0.3182	228	2.1428
211	-3.0342			254	-1.0252	316	-0.2752	336	2.1988

Table 2.
Dogs separated by residual data. Residuals were calculated by taking the vertical difference (in inches) between the linear fit model and the actual size (linear model – actual size).

From the regression data, dogs that were more than 3 inches below the linear model and thus predicted to be affected by the mystery gene were 306, 210, 204, 261, 262, 218, 205, 252, and 211. When a more lenient cutoff was chosen (more than 2 inches below), the dogs were 306, 210, 204, 261, 262, 218, 205, 252, 211, 225, 235, 313, 217, 232, 311, and 314.

Including GHR(2) in the equation

To include the GHR(2) gene in the linear regression equation which was excluded in the previous model, the dogs that had residuals bigger than -2 were used to construct another model. This group of dogs have genotype variety at the GHR(2) gene. The new equation was
Body size = 17.738–0.878(SMAD2)-0.354(IGF1R)–1.501(STC2)–2.006(GHR(1))–0.768(GHR(2))

These variables predicted body size in a statistically significant manner, $F(5, 20) = 6.27$, $p = 0.001$, $R^2 = .611$.

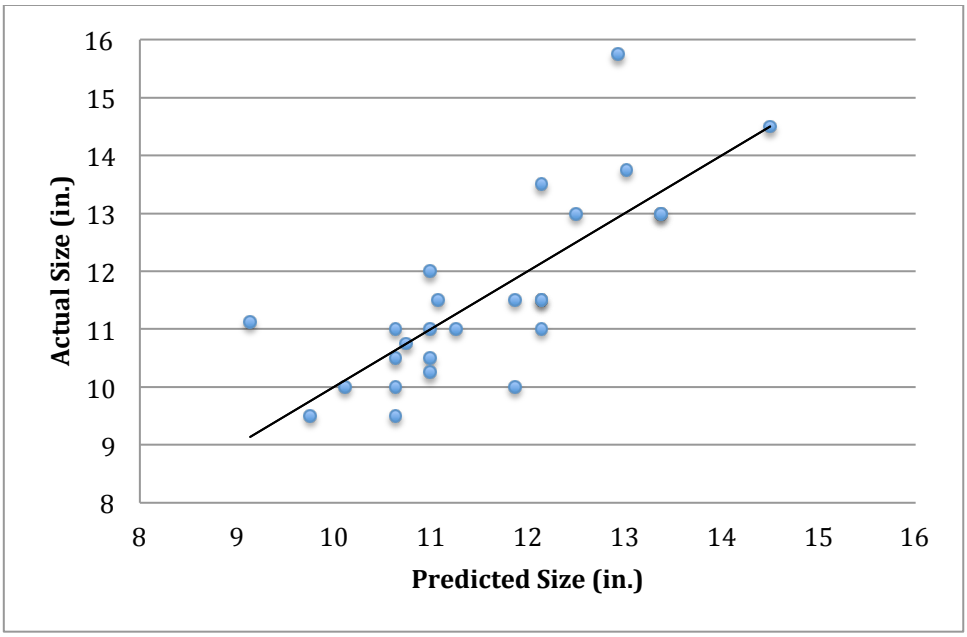

Figure 5.
A linear model was generated using dogs that had residuals bigger than -2. The linear regression equation and the genotype data were used to calculate the predicted size in inches. The actual body size of the dogs that were 11.5 inches or bigger are represented on the y-axis and the predicted body size of these dogs are represented on the x-axis. A line of best fit was calculated to be y=1.0001- 0.0009 (R^2=0.61053).

We repeated the process using this equation and plotted all dogs with the new linear model.

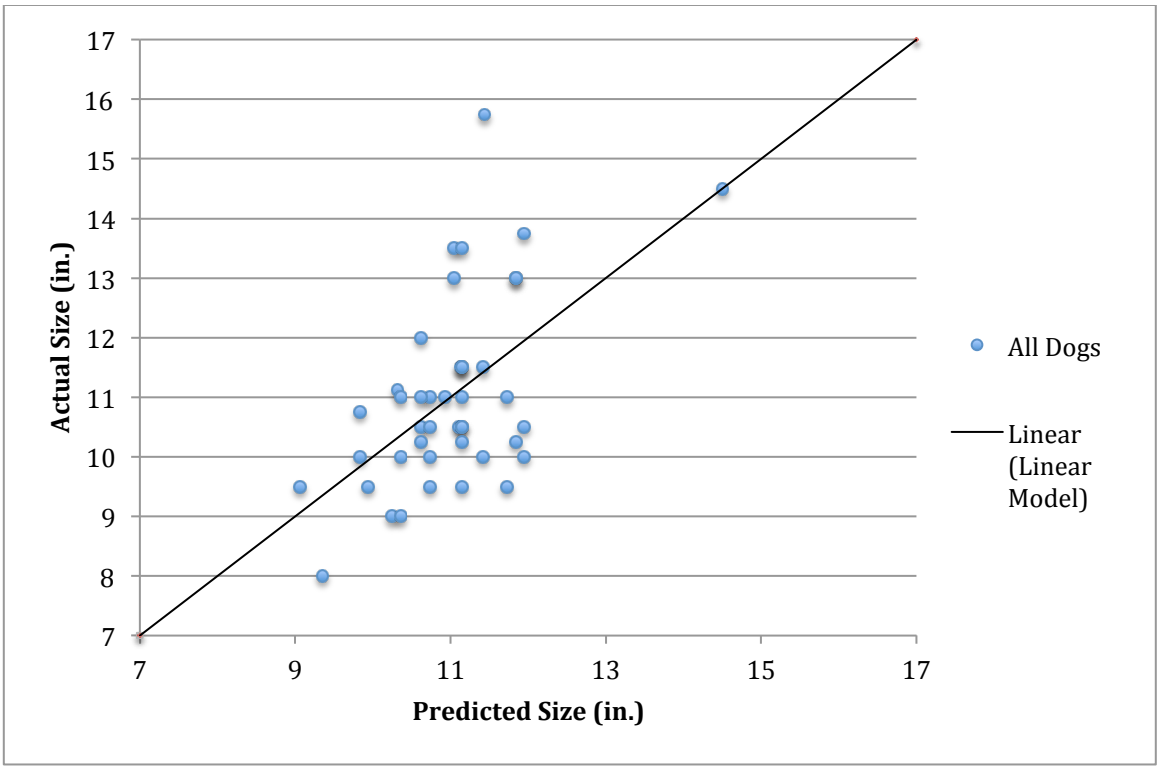

Figure 6.
A second scatter plot of all dogs representing the measured body size and the fitted body size in relation to the linear model. This linear model was calculated by the data from the dogs with residual values bigger than -2 (Table 2).

With this linear model, there are fewer dogs that digress significantly from the model. This can be explained by the increase in amount of dogs used to get the linear regression equation. The first linear regression equation was calculated using data from 10 dogs whereas data from 24 dogs were used to calculate this equation.

<-3		<-2		<-1		<0		>0	
number	residuals	number	residuals	number	residuals	number	residuals	number	residuals
210	-4.226	262	-2.726	313	-1.89	212	-0.743	329	0
306	-3.726	205	-2.64	246	-1.871	241	-0.64	226	0.001
261	-3.122	204	-2.616	252	-1.848	315	-0.64	247	0.007
211	-3.018	235	-2.518	311	-1.64	310	-0.639	244	0.361
		218	-2.262	314	-1.64	254	-0.493	221	0.422
				217	-1.5055	338	-0.372	341	0.506
				225	-1.433	340	-0.372	255	0.732
				232	-1.262	264	-0.371	253	1.006
				229	-1.14	231	-0.262	251	1.007
				312	-1.139	245	-0.261	328	1.36
						316	-0.139	228	1.987
						263	-0.115	336	2.817

Figure 7.
Dogs separated by residual data. Residuals were calculated by taking the vertical difference (in inches) between the linear fit model and the actual size (linear model – actual size). There are significantly less dogs with smaller residuals with this new model.

In the future, this new residual data can be used to create two subgroups of dogs that are affected by the unidentified gene of interest and dogs that are not. These two subgroups can be used to conduct a genome wide association scan and fine-mapping to locate loci that are associated with body size determination.

Genome Association Scan

To locate specific genes that are potential candidates for the mystery gene, the previously determined categorization criteria by Hoopes et al. of "tiny" dogs (<10 in.) and "control" dogs was used to scan the dogs from the CanMap project in which 915 dogs were genotyped at 60,968 SNP markers (Boyko et al., 2010). Loci around genes that code for various proteins that interact with STC2 or SMAD2 were observed with a 1 million bp window. A strong association peak was observed near one gene that encodes for the LEM domain containing protein 3 (LEMD3) with a $-\log_{10}$ p-value of 51.9531 (p-value=1.1 x $10^{-52)}$. Other genes that encode for retinoblastoma, p21, and Cyclin-D had $-\log_{10}$ p-values higher than 8. The values were 15.3716, 10.3643, and 9.02388 respectively.

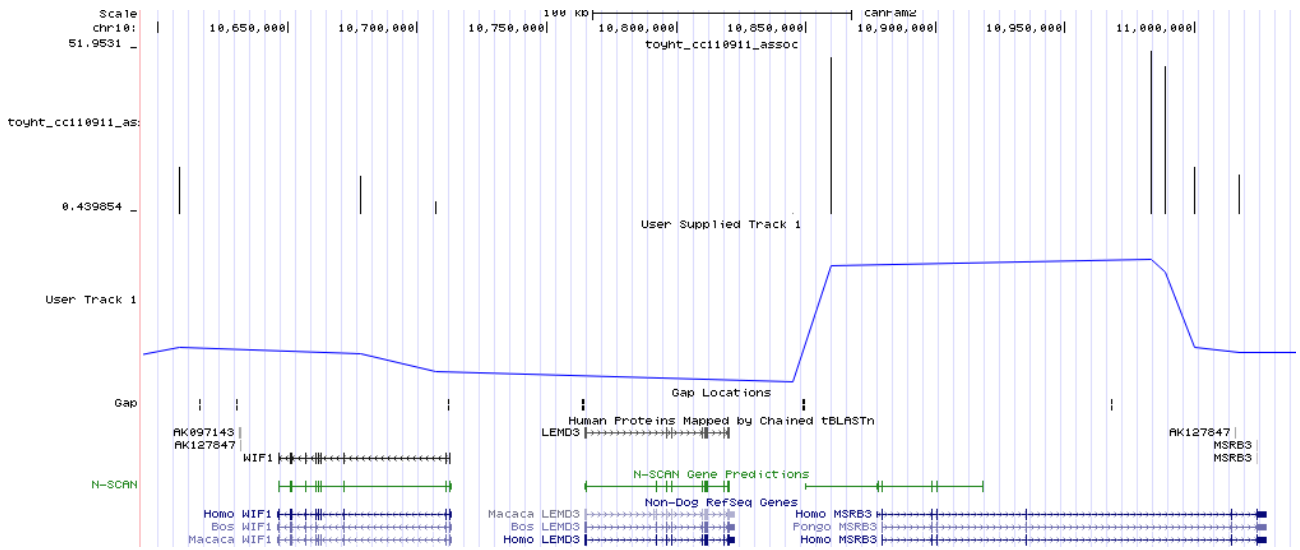

Figure 4.
A genome scan of a locus predicted to associate with toy poodle body size. The loci observed was near the gene that encodes for LEM domain containing protein 3 (LEMD3) which is also known as MAN1. Near the C-terminus of the gene, there is a signal with a $-\log_{10}$ p-value of 51.9531 which indicates a highly associated SNP.

Discussion

In this study, we aimed to find another gene that affects body size within toy poodles. We first used the genotype data at the five genes already identified to determine the allele combinations for each dog. When looking at Figure 2, GHR(1) and GHR(2) did not vary much between all dogs which indicates that they do not have a great affect in the variation in body size in toy poodles and miniature poodles. SMAD2 shows a trend with most of the homozygous derived alleles in dogs that are 10 inches or smaller (except for one dog that was 11 inches) which indicates that the SMAD2 genotype greatly affects body size variation. When looking at the dogs with the 7 common genotypes, their heights ranged from 10.25 inches to 13.5 inches. This is a clear indication that these five genes are not enough to perfectly determine body size in these dogs. Therefore, toy poodles and miniature poodles make great subjects in searching for an unidentified gene that controls for body size.

When the first linear model was made using only the dogs that were taller than 11.5 inches (10 dogs), there were more dogs with a smaller residual number indicating a grouping with more dogs that were categorized to have the mystery gene (Table 2). When the second linear model was made by using the residual data from Table 2, there were less dogs in the mystery gene category because more dogs were used to make the linear model. Although this second recategorization was important because it allowed for us to include the dogs that were heterozygous for the GHR2, the linear model shifted more towards the center of the collective dog data (Figure 6) weakening the residual numbers for some dogs. Including the GHR(2) gene in the linear regression model seems to create a trade-off in this way. For future studies, we can genotype more dogs in an effort to find more dogs that are heterozygous at the GHR(2) gene in addition to being fairly tall in height (preferably taller than 11.5 inches) to avoid this trade-off. This will allow us to be sure that the dogs that define the linear model are not affected by the unidentified gene but can still consider the GHR(2) gene when creating the linear regression model.

When the CanMap data was used to compare the "toy" dogs and the "control" dogs by doing an association scan, a signal with a high $-\log_{10}$ p value of 51.9531 was observed near 10,850,000 bp and another signal with a similar height around 11,000,000 bp. The first signal is near the C-terminus region of the gene LEMD3 and the second signal is within the MSRB3 gene. The MSRB3 gene encodes for the protein methionine sulfoxide reductase B3 (also known as DFNB74). This protein catalyzes the reduction of methionine sulfoxide to methionine (Kim et al, 2014), and is not known to affect growth or interact with any of the genes that are known to determine body size. However, the second signal near the C-terminus of the LEMD3 gene may correlate with determining body size. LEMD3 is an integral protein in the inner nuclear membrane of the nuclear envelope and has an RNA recognition motif on the carboxylic region to which all of the R-Smad proteins (except SMAD4) can bind. Pan et al found that the interaction between MAN1 and SMAD represses transcription activity of the R-Smads. This in turn represses signaling by the TGF-beta cytokines. In detail, overexpression of this protein inhibits R-Smad phosphorylation, nuclear translocation, and prevents transcriptional activation of TGF-beta, bone morphogenic protein 2, and activin-responsive promoters. Therefore, LEMD3 is a SMAD protein regulator (Pan et al, 2005), and is a promising candidate gene for body size regulation. This SNP was found by using the CanMap data for various breeds by comparing the "tiny" and the "control" group but we did not run a GWAS comparing the toy poodles in the categories that we proposed. Therefore, observing the

SNP at this locus using the categories determined by the residuals will be the next step in this project. This will not only verify if the LEMD3 gene is important in determining body size but will also let us look for other loci that seem to correlate with body size determination.

Literature Cited

Boyko AR, Quignon P, Li L, Schoenebeck JJ, Degenhardt JD, Lohmueller KE, Zhao K, Brisbin A, Parker HG, VonHoldt BM, Cargill M, Auton A, Reynolds A, Ostrander EA. 2010. A simple genetic architecture underlies morphological variation in dogs. PLoS Biol 8: e1000451.

Chase K, Carrier DR, Adler FR, Jarvik T, Ostrander EA, Lorentzen TD, Lark KG. 2002. Genetic basis for systems of skeletal quantitative traits: Principal component analysis of the canid skeleton. Proc Natl Acad Sci USA. 99(15):9930-5.

Dong Z, Yao M, Wang L, Yan X, Gu X, Shi Y, Yao N, Qiu L, Wu W, Yao D. 2013. Abnormal expression of insulin-like growth factor-I receptor in hepatoma tissue and its inhibition to promote apoptosis of tumor cells. Tumour Biol. 34(6):3397-405.

Efstratiadis, A. 1998. Genetics of mouse growth. Int J Dev Biol. 42(7):955-76.

Elis S, Courtland HW, Wu Y, Rosen CJ, Sun H, Jepsen KJ, Majeska RJ, Yakar S. 2010. Elevated serum levels of IGF-1 are sufficient to establish normal body size and skeletal properties even in the absence of tissue IGF-1. J Bone Miner Res. 25(6):1257-66.

Hellemans J, Preobrazhenska O, Willaert A, Debeer P, Verdonk PC, Costa T, Janssens K, Menten B, Van Roy N, Vermeulen SJ, Savarirayan R, Van Hul W, Vanhoenacker F, Huylebroeck D, De Paepe A, Naeyaert JM, Vandesompele J, Speleman F, Verschueren K, Coucke PJ, Mortier GR. 2004. Loss-of-function mutations in LEMD3 result in osteopoikilosis, Buschke-Ollendorff syndrome and melorheostosis. Nat Genet. 36(11):1213-8.

Horikoshi M1, Yaghootkar H, Mook-Kanamori DO, Sovio U, Taal HR, Hennig BJ, Bradfield JP, St Pourcain B, Evans DM, Charoen P, Kaakinen M, Cousminer DL, Lehtimäki T, Kreiner-Møller E, Warrington NM, Bustamante M, Feenstra B, Berry DJ, Thiering E, et al. 2013. New loci associated with birth weight identify genetic links between intrauterine growth and adult height and metabolism. Nat Genet. 45(1):76-82.

Johnston J, Ramos-Valdes Y, Stanton LA, Ladhani S, Beier F, Dimattia GE. 2010. Human stanniocalcin-1 or -2 expressed in mice reduces bone size and severely inhibits cranial intramembranous bone growth. 19(6):1017-39.

Jones P, Chase K, Martin A, Davern P, Ostrander EA, Lark KG. 2008. Single-nucleotide-

polymorphism-based association mapping of dog stereotypes. Genetics:179:2:1033-1044.

Kim BJ, Hwang JY, Han BG, Lee JY, Lee JY, Park EK, Lee SH, Chung YE, Kim GS, Kim SY, Koh JM. 2011. Association of SMAD2 polymorphisms with bone mineral density in postmenopausal Korean women. Osteoporos Int. 22(8):2273-82.

Kim JY, Kim Y, Kwak GH, Oh SY, Kim HY. 2014. Over-expression of methionine sulfoxide reductase A in the endoplasmic reticulum increases resistance to oxidative and ER stresses. Acta Biochim Biophys Sin. 46(5): 415-9.

Hoopes BC, Rimbault M, Liebers D, Ostrander EA, Sutter NB. 2012. The insulin-like growth factor 1 receptor (IGF1R) contributes to reduced size in dogs. Mamm Genome. 23(11-12): 780-790.

Lango AH, Estrada K, Lettre G, et al. 2010. Hundreds of variants influence height and cluster in genomic loci and biologicalpathways. Nature 467:832-38.

Law AY, Wong CK. 2013. Stanniocalcin-1 and -2 promote angiogenic sprouting in HUVECs via VEGF/VEGFR2 and angiopoietin signaling pathways. Mol Cell Endocrinol. 374(1-2):73-81.

Law AY, Wong CK. 2010. Stanniocalcin-2 promotes epithelial-mesenchymal transition and invasiveness in hypoxic human ovarian cancer cells. Exp Cell Res. 316(20):3425-34.

Lin CS, Chen MF, Wang YS, Chuang TF, Chiang YL, Chu RM. 2013. IL-6 restores dendritic cell maturation inhibited by tumor-derived TGF-β through interfering Smad 2/3 nuclear translocation. Cytokine. 62(3):352-359.

Pan D, Estevez-Salmeron LD, Stoschein SL, Zhu X, He J, Zhou S, Luo K. 2005. The integral inner nuclear membrane protein MAN1 physically interacts with the R-Smad proteins to repress signaling by the transforming growth factor- beta superfamily of cytokines. J Biol Chem. 280(16):15992-6001.

Postel-Vinay MC, Finidori J. 1995. Growth hormone receptor: structure and signal transduction. Eur J Endocrinol. 133(6):654-9.

Riggins GJ, Thiagalingam S, Rozenblum E, Weinstein CL, Kern SE, Hamilton SR, Willson JK, Markowitz SD, Kinzler KW, Vogelstein B. 1996. Mad-related genes in the human. 13(3):347-9.

Rimbault M, Beale HC, Schoenebeck JJ, Hoopes BC, Allen JJ, Glynn PK, Wayne RK, Sutter NB, Ostrander EA. 2013. Derived variants at six genes explain nearly half of size reduction in dog breeds. Genome Res. 23:1985-1995.

Sutter NB, Bustamante CD, Chase K, Gray MM, Zhao K, Zhu L, Padhukasahasram B, et

al. 2007. A single IGF1 allele is a major determinant of small size in dogs. Science: 316: 112-115.

Sutter NB, Mosher DS, Gray MM, Ostrander EA. 2008. Morphometrics within dog breeds are highly reproducible and dispute Rensch's rule. Mammalian Genome. 19: 713-723.

Tu B, Peng ZX, Fan QM, Du L, Yan W, Tang TT. 2014. Osteosarcoma cells promote the production of pro-tumor cytokines in mesenchymal stem cells by inhibiting their osteogenic differentiation through the TGF-β/Smad2/3 pathway. Exp Cell Res. 320(1):164-73.

Vila C, Savolainen P, Maldonado J E, Amorim I R, Rice J E, Honeycutt R L, Crandall K A, Lundeberg J, Wayne R K (1997) Multiple and Ancient Origins of the Domestic Dog Science 276:1687–1689.

Weedon MN, Guillaume L, Freathy RM, Lindgren CM, et al. 2007. A common variant of HMGA2 is associated with adult and childhood height in the general population. Nature Genet. 39: 1245-1250.