pISSN 1598-298X / eISSN 2384-0749
J Vet Clin 36(4) : 190-195 (2019)
http://dx.doi.org/10.17555/jvc.2019.08.36.4.190

Journal of Veterinary Clinics
www.ksvc.or.kr

Identification of the Marker Genes Related With Chronic Mitral Valve Disease in Dogs

Byung-Gook Yoon, Dong-Soo Lee*, Kyoung-Won Seo and Kun-Ho Song[1]

College of Veterinary Medicine, Chungnam National University, Daejeon 34134, Korea
**College of Korean Medicine, Gachon University, Seongnam 13120, Korea*

(Received: June 18, 2019 / Accepted: August 11, 2019)

Abstract : We aimed to identify genomic variations as well as the marker genes related with chronic mitral valve disease (CMVD) in *Canis lupus familiaris* using whole genome resequencing, which provides valuable resources for further study. Two ten-year old female *Canis lupus familiaris* English cocker spaniels were used for this study, one control and one who had been diagnosed as CMVD. For the whole genome resequencing, muscles from the left ventricular wall were collected from each dog. With the HiSeq DNA Shotgun library and HiSeq™ 2000 platform, whole genome resequencing was performed. From the results, we identified 5 million and 6 million variants in gene expression in the control and CMVD-diagnosed subject, respectively. We then selected the top 1,000 genes from the SNP, INS, and DEL mutation and 675 genes among them were overlapped for every mutation between the control and CMVD-diagnosed patient. Interestingly, in both groups, the intron variant (91.16 and 91.18%) and upstream variant (3.10 and 3.08%) are most highly related. Among the overlapped 675 genes, gene ontology for intracellular signal transduction is highly counted in INS, and DEL, and SNPs (35, 33, 31, respectively). In this study, we found that the COL and CDH gene families could be key molecules in identifying the difference in gene expression between control and CMVD-diagnosed dogs. We believe further studies will prove the importance of variants in key molecule expression and that these data will serve as a valuable foundation stone the study of canine CMVD.

Key words **:** English cocker spaniel, CMVD, whole genome resequencing, dog.

Introduction

Chronic mitral valve disease (CMVD) is the most common cause of canine heart disease and is commonly found in small-sized dogs including Cocker spaniels, Cavalier King Charles Spaniels, Poodles, Yorkshire Terriers and Chihuahuas (4,18). CMVD is a chronic degenerative disease causing thickening valve edges resulting in valve insufficiency (19). The chronic patients with CMVD are usually recognized by exhibiting symptoms such as coughing, intolerance to exercise, rapid breathing and dyspnea. The end-stage disease results in significant mitral regurgitation that can lead to left-side congestive heart failure (5,12).

To date, while structural and cellular changes in canine CMVD are well known, molecular mechanism and genomic changes are far less studied. Lu *et al.* (15) reported on a genomics study of canine CMVD with the Affymetrix Canine Gene 1.0 array. They showed that evaluation of global expression patterns provides a molecular map of mitral valve disease and, yields insight into the pathophysiologic aspects of CMVD. Since that study, there has been significant development of the tools of transcriptome study and bioinformatics. Recently, other group introduced gene network and canonical pathway analysis in canine CMVD with microarray (15). They identified differentially expressed genes related with inflammation, cell movement, development, and extracellular matrix organization in CMVD transcriptome.

Recently, whole genome resequencing has been introduced, which is one of the next generation sequencing (NGS) techniques, and it has the powerful ability to detect a broad range of variants to identify single nucleotide variants (SNVs), Indels, structural variants (SVs), copy number variants (CNVs), and repeat expansions all with a single assay. Despite the rapid drop in the cost associated with NGS, microarrays are still more economical and yield higher throughput, providing significant advantages when working with a large number of samples. However, arrays suffer from fundamental 'design bias' - they only return results from those regions for which probes have been designed. Consequently, arrays are only as good as the databases from which they are designed (9,21).

In this study, we aimed to identify genomic variations as well as the marker genes related with CMVD in *Canis lupus familiaris* using whole genome resequencing and to provide valuable resources for further study, especially to get a less biased view of the genome with a simple, streamlined assay compared to microarray.

Materials and Methods

Animals and tissue collection

Two ten-year old female *Canis lupus familiaris* English cocker spaniels were used for this study, one control and one

[1]Corresponding author.
E-mail : songkh@cnu.ac.kr

who had been diagnosed as CMVD. This patient diagnosed as CMVD with a cough, dyspnea, exercise intolerance and by using echocardiography. For the whole genome sequencing, muscles from the left ventricular free wall were collected from dog after death.

HiSeq DNA Shotgun library preparation

We submitted the tissues for library preparation and whole genome sequencing at the Macrogene Co. (Seoul, South Korea). DNA each sequenced sample is prepared using TruSeq DNA sample Prep Kit v2 Support (Illumina, SD, USA) according to manufacturer's instruction. Briefly, one microgram of genomic DNA was fragmented by covaries. The fragmented DNA was repaired by converting overhangs into blunt ends. After completion of the repair, we then selected the appropriate library products ranged by 400-500 base pairs. A single 'A' nucleotide was added to the 3′-end of repaired fragments followed by ligating Illumina adapters to the fragments. The products were then amplified using polymerase chain reaction, and the final products were then validated using the Agilent Bioanalyzer. We then sequenced using the HiSeq™ 2000 platform (Illumina).

Alignment and variant calling

After sequencing, Fast QC (v0.11.5) (1) was performed to check the read quality. Trimmomatic (v0.36) (3) was used to remove adapter sequences and low-quality reads to reduce biases in analysis. The filtered data from *Canis lupus familiaris* (CanFam3.1) genome were then mapped using BWA (v0.7.17) (14) with mem algorithm to NCBI reference sequence from NCBI database. After read mapping, duplicated reads were removed with Picard (v2.17.2) (20). Genome coverage and mapping ratio of mapped reads on the reference genome were calculated and variant calling was performed using SAMtools (v.1.6) (13) and BCFtools (v.1.6) (17). In this step, SNPs and short Indels candidates with Phred score over 30 (base call accuracy of 99.9%) were captured using the information of the mapped reads. These variants were then classified by chromosomes or scaffolds, and the information of the location was marked. Captured variants were annotated with SnpEff (v.3.1) (7) to predict the effects of genetic variants. In this step, annotation type and putative impact and amino acid change information were predicted. For further information, CNVs and SV were analyzed. CNVs were calculated with control-FREEC (v.11.0) (2) with the parameter 'ploidy 2', and SV was detected with BreakDancer (v.1.4.5) (6).

Table 1. Summary of whole genome resequencing in this study

	Control	Patient
Total read base (bp)	111,275,210,222	139,688,997,096
Total reads	736,921,922	925,092,696
GC (%)	41.39	42.68
Q20 (%)	90.34	87.7
Q30 (%)	80.82	76.66
Filtered		
Ref. length	292,715,236	292,715,236
Mapped sites	2,384,572,615 (99.66%)	2,386,545,234 (99.74%)
Total read	631,370,582	665,845,232
Mapped reads	592,065,462 (93.77%)	614,864,532 (92.34%)
Mapped bases	76,372,976,181	75,493,168,121
Mean depth	31.92	31.55
Variants	5,467,736	5,996,075

Results

Whole genome resequencing

In the result, the gross mapping yields were 111.28 and 139.69 Gb for the control and patient samples, and the contents of CG were 41.39% and 42.68% in the control and patients (Table 1). The filtered read sequences were then aligned to the reference genome sequence. The overall read mapping ratio of fully-called bases for the control and patient samples were 93.77% and 92.34%, respectively (Table 1). We then then analyzed the variants from filtered reads. The results showed that the total numbers of variants was 5,467,736 in the control and 5,996,075 in the patient (Table 1).

Type of genomic mutations

Using called variants, we classified the variants into three different types, including SNPs, insertions, and deletions.

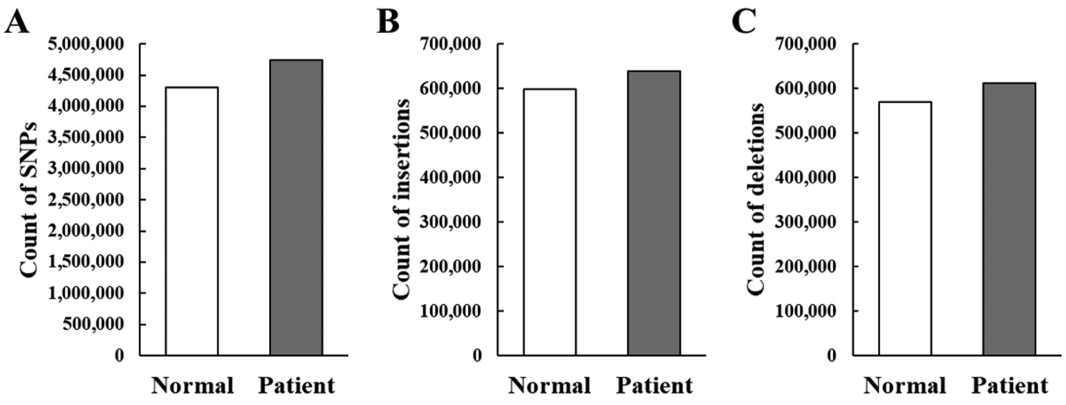

Fig 1. Structural variants of normal (control) and patient samples. The bar graphs represented the number of SNPs (A), insertion (B), and deletions (C) in control (white) and patient (gray). Compared to the counts of insertions and deletions, the counts of SNPs showed more than 4 million in both groups.

SNPs are the most common variant type in both control and patient which were 4,299,475 and 4,745,755, respectively (Fig 1A). Among variants, insertions/deletions were 597,826/570,435 and 638276/612044 in the control and patient, respectively (Fig 1B, 1C).

SNP analysis and annotations

Based on the SNPs data, we analyzed the amount of single nucleotide changes and calculated the transition and transversion ratio. The results showed that the most frequent base changes was G to A, followed by C to T (650,016), A to G (621,506) and T to C (620,959) in both the control and patient samples (Fig 2A, 2B). Because the quality of SNP data can be characterized by the ratio of transition (Ts) to transversion (Tv), we additionally calculated transition (Ts)/transversion (Tv) ratio from each sample. We identified 2,816,563 transitions and 1,482,912 transversions in the control sample (Ts/Tv ratio of 1.9), and 3,133,941 Ts and 1,611,814 Tv were identified in the patient (Ts/Tv ratio of 1.94) (Fig 2C, 2D). Furthermore, the SNPs were function-

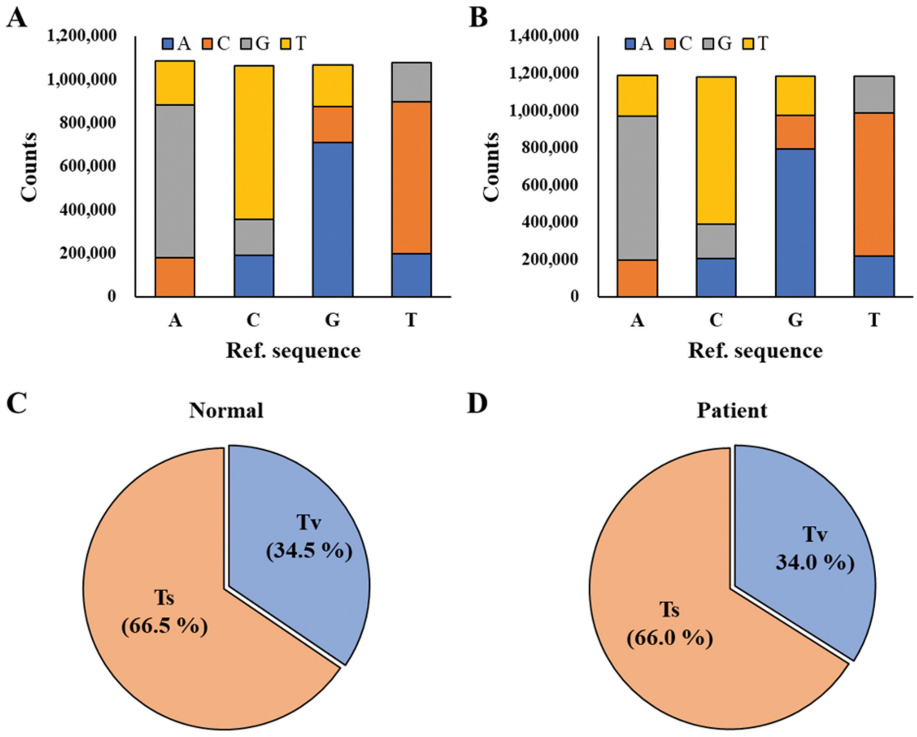

Fig 2. Structural variants of normal (control) and patient samples. In the normal (control) dog (A) and patient (B), bar graph described how many counts were shown for SNPs in each nucleotide. Based on SNPs, transition and transversion count were calculated in the normal (control) dog (C) and patient (D). Ts: transition, Tv: transversion.

Table 2. Comparison of sequence ontology in this study

Type of annotation	Control		Patient	
	Count	Ratio (%)	Count	Ratio (%)
Intron variant	1,817,840	91.16	2,029,633	91.18
Upstream gene variant	61,769	3.10	68,649	3.08
3′ UTR variant	33,850	1.70	37,796	1.70
Downstream gene variant	21,303	1.07	23,682	1.06
Synonymous variant	18,276	0.92	20,200	0.91
Missense variant	16,077	0.81	17,970	0.81
5′ UTR variant	12,096	0.61	13,258	0.60
Intragenic variant	4,451	0.22	5,267	0.24
Splice region variant and intron variant	3,225	0.16	3,537	0.16
5′ UTR premature start codon gain variant	1,640	0.08	1,849	0.08
Frameshift variant	750	0.04	875	0.04
Splice region variant and synonymous variant	370	0.02	408	0.02
Missense variant and splice region variant	358	0.02	398	0.02
Frameshift variant and splice region variant	249	0.01	261	0.01
Splice region variant	234	0.01	260	0.01

ally classified using SnpEff annotations and provided at Table 2. Most SNPs were in intron variants which totaled 1,817,840 (91.16% of total) in the control and 2,029,633 (91.2% of total) in the patients. We also found 37,796 SNPs in the 3′ UTRs and 17,970 SNPs in 5′ UTRs (Table 2).

Functional annotation

We then selected top 1,000 genes from SNP, INS, and DEL which were only detectable in patients and counted the number of genes in each type of mutations. One hundred eighty-eight genes had SNP only, 103 genes had INS only, 92 genes had only DEL. Additionally, 63 genes existed with SNP and INS, 159 genes with INS and DEL, and 73 genes with SNP and DEL. All three types of mutations existed in 675 genes (Fig 3). Furthermore, we functionally annotated these genes using DAVID. Based on the data, we found that the majority of variants were included in functional categories including intracellular signal transduction, morpholoho-mophilic cell adhesion via plasma membrane adhesion molecules, proteinaceous extracellular matrix, and high voltage-gated calcium channel activity (Table 3, Fig 3C).

Gene network

We carried out STRING which is a tool for verifying the network between genes. As shown in Fig 3, we found that COL genes and CDH genes were connected in a group of homophilic cell adhesion via plasma membrane adhesion molecules and proteinaceous extracellular matrix, respectively (Fig 3A, 3B).

Discussion

This study was carried out using the TruSeq DNA Sample Prep Kit v2 Support. Whole-genome resequencing of control and CMVD-diagnosed Cocker spaniels was carried out to evaluate the genome-wide variants including SNPs, insertion, and deletion. Previous whole genome resequencing studies for canines obtained higher mean read coverages and

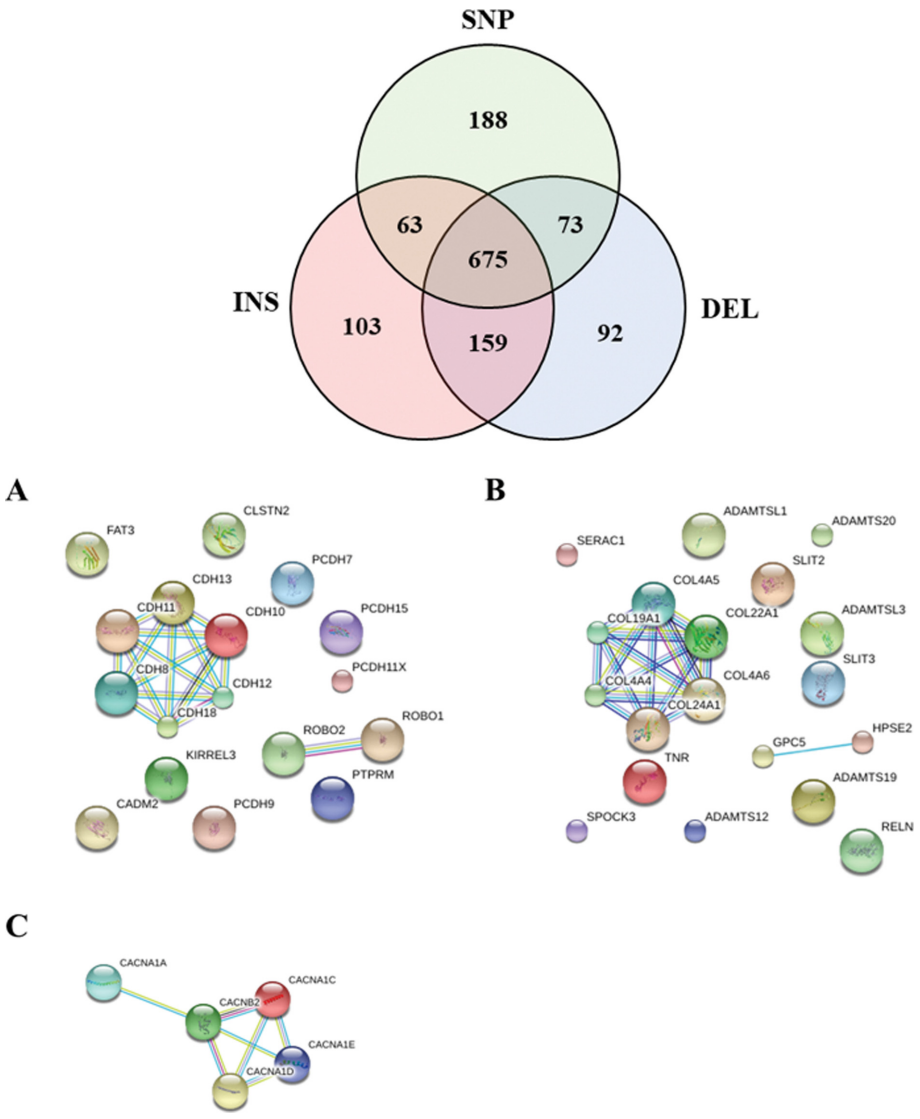

Fig 3. Structural variants of control and patient samples**.** In the list of overlapped top 1,000 genes for INS, DEL, SNP, the highly ranked gene ontology (GO) showed genes linked by each other. A. Homophilic cell adhesion via plasma membrane adhesion molecules, B. Proteinaceous extracellular matrix, C. Voltage-gated calcium channel activity.

Table 3. Comparison of sequence ontology. Selected top 1,000 genes from SNP, INS, and DEL, which were only detectable in patients, were overlapped to identify genes which could be the candidates for the key molecules related with CMVD-diagnosis

Term	Del			Ins			SNP		
	Count	%	p-Value	Count	%	p-Value	Count	%	p-Value
Intracellular signal transduction	35	3.65	0.000	33	3.44	0.000	31	3.29	0.000
Cell morphogenesis	13	1.36	0.000	11	1.15	0.000	10	1.06	0.001
Homophilic cell adhesion via plasma membrane adhes ion molecules	17	1.77	0.000	18	1.88	0.000	18	1.91	0.000
Dendrite morphogenesis	9	0.94	0.000	8	0.83	0.001	-	-	-
Synapse assembly	8	0.84	0.00	7	0.73	0.001	8	0.85	0.000
Axon guidance	13	1.36	0.001	18	1.88	0.000	12	1.27	0.002
Receptor localization to synapse	4	0.42	0.001	-	-	-	-	-	-
Long-term synaptic potentiation	7	0.73	0.001	7	0.73	0.001	-	-	-
Regulation of synaptic transmission, glutamatergic	6	0.63	0.002	-	-	-	-	-	-
Regulation of axonogenesis	6	0.63	0.002	-	-	-	-	-	-
Regulation of Rho protein signal transduction	-	-	-	14	1.46	0.000	11	1.17	0.003
Signal transduction	-	-	-	36	3.75	0.000	-	-	-
Peptidyl-serine phosphorylation	-	-	-	15	1.56	0.001	-	-	-
Adult locomotory behavior	-	-	-	-	-	-	9	0.95	0.001
Positive regulation of synapse assembly	-	-	-	-	-	-	9	0.95	0.002
Neuron development	-	-	-	-	-	-	7	0.74	0.002
Neuromuscular process controlling balance							8	0.85	0.003

identified a number of variants between 6.1 and 7.4 million (8,10,11). In this study, we detected 5.5 and 6 million variants in the control and patient. Furthermore, SNPs are the most abundant type of variant for each dog with the individual variant calling approach.

Genomic variants including SNPs, insertions, and deletions can be annotated using SnpEff based on their locations such as intron, untranslated region, 5′ UTR or 3′ UTR, upstream, downstream, splice site, or intergenic regions. It also predicts coding effects such as synonymous or non-synonymous amino acid replacement, start codon gains or losses, stop codon gains or losses, or frame shifts. In the present study, most SNPs were in intronic regions in both the control and patient followed by downstream/3′ UTR and upstream/5′ UTR. These variants may be responsible for the regulations of gene expression.

From the data, we overlapped top 1,000 genes for INS, DEL and SNP. The highly ranked gene ontology (GO) showed genes linked to each other. Compared to the control, variants of gene expression in the patient are related with intracellular signal transduction, homophillic cell adhesion via plasma membrane adhesion molecules, and cell morphogenesis in the sequence ontology. We further investigated to identify the specific genes which showed the significant difference between the control and patient. As shown in Fig 3, the variants in collagen (COL), cadherin (CDH) and the calcium voltage-gated channel subunit (CACN) genes are highly linked with other genes in each gene family. Previous studies have reported this about the COL and CDH genes (16). They showed the comparative transcriptomic profiling and gene expression using microarray and identified some genes for tissue structure such as collagen genes COL1A1, 2 & 3, 12A1 and 14A1. However, in this study, we found new candidate genes in the COL family such as COL19A1, COL4A5, COL22A1, COL4A6, COL24A1, COL4A4 for degeneration in tissue structure of CMVD-diagnosed dogs. Also, the previous study showed the intensity of CAH5 was downregulated in the CMVD patient. However, we understand that many genes of the CDH family are engaged in CMVD such as CDH8, CDH11, CDH13, CDH10, CDH12 and CDH18.

In conclusion, we conducted whole genome resequencing to compare a control dog subject with one diagnosed with CMVD. Although this study only included a single dog for each group, it study provides valuable information for both diagnostics and pathogenesis of CMVD. Based on our data, further studies need to verify the important genes associated with CMVD. Therefore, further study is necessary with whole genome resequencing for large numbers of dogs with CMVD.

Conflict of Interest

The authors declare no conflict of interests.

References

1. Babraham Bioinformatics. Available at: https://www.bioinformatics.babraham.ac.uk/projects/fastqc/.
2. Boeva V, Popova T, Bleakley K, Chiche P, Cappo J, Schleiermacher G, Janoueix-Lerosey I, Delattre O, Barillot E. Control-FREEC: a tool for assessing copy number and allelic content using next-generation sequencing data. Bioinformatics 2012; 28: 423-425.
3. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 2014; 30: 2114-2120.
4. Borgarelli M, Buchanan JW. Historical review, epidemiology and natural history of degenerative mitral valve disease. J Vet Cardiol 2012; 14: 93-101.
5. Borgarelli M, Häggström J. Canine degenerative myxomatous

mitral valve disease: natural history, clinical presentation and therapy. Vet Clin North Am Small Anim Pract 2010; 40: 651-663.
6. Chen K, Wallis JW, McLellan MD, Larson DE, Kalicki JM, Pohl CS, McGrath SD, Wendl MC, Zhang Q, Locke DP, Shi X, Fulton RS, Ley TJ, Wilson RK, Ding L, Mardis ER. BreakDancer: an algorithm for high-resolution mapping of genomic structural variation. Nat Methods 2009; 6: 677-681.
7. Cingolani P, Platts A, Wang le L, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden DM. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. Fly (Austin) 2012; 6: 80-92.
8. Drögemüller M, Jagannathan V, Becker D, Drögemüller C, Schelling C, Plassais J, Kaerle C, Dufaure de Citres C, Thomas A, Müller EJ, Welle MM, Roosje P, Leeb T. A mutation in the FAM83G gene in dogs with hereditary footpad hyperkeratosis (HFH). PLoS Genet 2014; 10: e1004370.
9. Edeki C. Comparative study of microarray and next generation sequencing technologies. IJCSMC 2012; 1: 15-20.
10. Gilliam D, O'Brien DP, Coates JR, Johnson GS, Johnson GC, Mhlanga-Mutangadura T, Hansen L, Taylor JF, Schnabel RD. A Homozygous KCNJ10 Mutation in Jack Russell Terriers and Related Breeds with Spinocerebellar Ataxia with Myokymia, Seizures, or Both. J Vet Intern Med 2014; 28: 871-877.
11. Guo J, Johnson GS, Brown HA, Provencher ML, da Costa RC, Mhlanga-Mutangadura T, Taylor JF, Schnabel RD, O'Brien DP, Katz ML. A CLN8 nonsense mutation in the whole genome sequence of a mixed breed dog with neuronal ceroid lipofuscinosis and Australian Shepherd ancestry. Mol Genet Metab 2014; 112: 302-309.
12. Häggström J, Höglund K, Borgarelli M. An update on treatment and prognostic indicators in canine myxomatous mitral valve disease. J Small Anim Pract 2009; 50: 25-33.
13. Li H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics 2011; 27: 2987-2993.
14. Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics 2010; 26: 589-595.
15. Lu CC, Liu MM, Culshaw G, Clinton M, Argyle DJ, Corcoran BM. Gene network and canonical pathway analysis in canine myxomatous mitral valve disease: a microarray study. Vet J 2015; 204: 23-31.
16. Markby GR, Summers KM, MacRae VE, Corcoran BM. Comparative transcriptomic profiling and gene expression for myxomatous mitral valve disease in the dog and human. Vet Sci 2017; 4: E 34.
17. Multiallelic calling model in bcftools (-m). Available at : http://samtools.github.io/bcftools/call-m.pdf.
18. Parker HG, Kilroy-Glynn P. Myxomatous mitral valve disease in dogs: Does size matter? J Vet Cardiol 2012; 14: 19-29.
19. Petric AD. Myxomatous mitral valve disease in dogs - an update and perspectives. Mac Vet Rev 2015; 38: 13-20.
20. Picard. Available at : http://broadinstitute.github.io/picard.
21. Roh SW, Abell GC, Kim KH, Nam YD, Bae JW. Comparing microarrays and next-generation sequencing technologies for microbial ecology research. Trends Biotechnol 2010; 28: 291-299.