INNSPUB

International Journal of Biosciences | IJB |
ISSN: 2220-6655 (Print), 2222-5234 (Online)
http://www.innspub.net
Vol. 11, No. 1, p. 75-82, 2017

RESEARCH PAPER OPEN ACCESS

Genome-wide association study of cat mammary tumor using 63,000 SNP chip through PLINK data analysis toolset

Rashid Saif*[1], Leslie Lyons[2], Barbara Gandolfi[2], Maleeha Ahmad[3], Ramlah Ejaz[3], Muhammad Wasim[4]

[1]*Department of Biotechnology, Virtual University of Pakistan, Lahore, Pakistan*

[2]*Department of Veterinary Medicine and Surgery, University of Missouri-Columbia, USA*

[3]*Department of Biotechnology, Kinnaird College for Women, Lahore, Pakistan*

[4]*Institute of Biochemistry and Biotechnology, University of Veterinary and Animal Sciences, Lahore, Pakistan*

Key words: GWAS, LD plot, Pakistani siamese cat, 63K Chip, PLINK

http://dx.doi.org/10.12692/ijb/11.1.75-82 Article published on July 11, 2017

Abstract

Genome wide case-control association study using Illumina Infinium Feline 63K *I* Select DNA array was performed with7 cancer cases and 23 controls from the Siamese cat breed. The purpose of this study to identify the SNPs associated with mammary tumor in cats. PLINK data analysis toolset was used to analysis the SNP data obtained through microarray genotyping experiment. Zerovalue of Mendel error was observed, Similarly, deviations from the HWE was also detected which depicts the excessive in breeding within the sampled population. Allelic association test highlighted ten most associated SNPs through Manhattan plot. Linkage-disequilibrium plots were also drawn through these associated SNPs, which showed that, one of the SNP at locus ChrC1:202,770,816 on chromosome 8 exist in haplotype, which is part of *RFTN2* gene and belongs to raftlin protein family. This protein is involved in the activation of B type immune cells. Another SNP ChrE1:53,681,930 on chromosome 14 was found to be in linkage disequilibrium with 4 genes named *C7orf64, APPBP2, PPMID and BCAS3*, all of these genes are highly expressed in breast cancers. This study revealed, that outbreak of mammary tumor in cats may be associated be with aforementioned SNPs and simultaneously outbreak of this cancer is a micro-evolutionary process and cumulative effect of number of SNPs located on different chromosomes.

* **Corresponding Author:** Rashid Saif ✉ saif37@gmail.com

Introduction

Cancer is one of the most fatal ailments among domestic animals. Different types of cancers which are prevalent in *Feline catus* are Feline Vaccine Associated Sarcomas (FVAS), feline lymphoma and mammary gland adenocarcinoma (Chu *et al.*, 2001). Siamese cat breed is mostly affected by mammary tumor, which is very aggressive in nature, so its early diagnosis and detection is very much important(Shafiee *et al.*, 2013). Disease associated mutations may serve as potential tumor markers for every type of neoplasm. Alterations in DNA biomarkers might be helpful for cancer diagnosis and prognosis (Sherman and Multhoff, 2007; Bei *et al.*, 2010; Liu *et al.*, 2010). For this purpose, Genome Wide Association Studies (GWAS)is conducted with an aim to determine most associated SNP with mammary tumor adenocarcinoma in Siamese cats (Bush and Moore, 2012).

In this study, a large scale 63K SNP genotyping data was generated from seven cancer cases of mammary tumor in Siamese cat which are compared with normal controls to detect that which of the locus/SNP is highly associated with cancer. If the allele frequency is significantly higher in cancer, the odds ratio (OR) will also be higher than 1 and vice-versa, which indicates the strong association between locus/allele and disease (Holcomb Jr *et al.*, 2001).

PLINK is used as the data analysis toolset in this Genome-wide association study. Allelic association test displayed highly associated SNPs with this tumor type through Manhattan plot. Then, these SNPs are selected to make LD plots to analyze haplotypic associations. LD plots are visualized using Haploview, which is a supportive software that makes the visualization of associated haplotypes maps possible to understand the results in a better way (Purcell *et al.*, 2007; Magi and Morris, 2010; Klein *et al.*, 2005).

Materials and methods

Sample collection

A total of seven (n=7) tumorous tissues and (n=23) normal controls were collected during surgical excision of Siamese cats.

Proper diagnosis of the tumor was made by veterinary practitioner and core tissues from the tumorous mass were obtained for DNA extraction. Tumor excision and diagnosis was performed by adopting the proper procedures laid down by ethical committee of University of Veterinary and Animal Sciences, Lahore, Pakistan. DNA extraction was done using TIANGEN kit.

PLINK data analysis toolset and Haploview software

In this study, Illumina Infinium *Feline* 63k *i*s elect DNA chip was used which includes 62897 markers (Illumina, Inc., San Diego, CA), these run were made possible through generous funding from Lyons Feline Genetics lab at University of Missouri-Columbia, USA. Data analysis was conducted to predict the SNPs associated with mammary cancer in this breed. This case/control group study was performed by applying a number of statistical tests through PLINK data analysis toolset, while Haploview supportive software was used for the visualization of LD plots. The sample population after pruning was tested for allelic association, Hardy–Weinberg equilibrium, Mendel error, Cochran Mantel Haenszel, SNP disease cluster and Haplotype association.

Results and discussion

Various filters were applied in order to analyze this genotypic data; which includes filters of minor allele frequency (MAF=0.01), maximum SNP missing rate (GENO=0.1), and maximum individual missing rate (MIND=0.1). Before frequency and genotypic pruning, there were 62897 reported SNPs and the total genotyping rate was 0.906742. A total of 30 individual's data was analyzed including seven mammary tumor cases and 23 normal controls. Two cancer cases were removed with mind > 0.1. The total genotyping rate in the remaining individuals was observed to be 0.989092. A total of 555 SNPs failed missing ness test with gene> 0.1, similarly 5693 SNPs failed frequency test with MAF < 0.01. After frequency and genotypic pruning, there were 56703 SNPs were reported that contains five tumor cases and 23 normal controls.

Allelic association

The result of the allelic association test was demonstrated in the form of a Manhattan plot. The plot of highly associated SNPs was drawn taking chromosome number on x-axis and –log10 of *p*-values on y-axis. The ten significant SNPs includes chrB2.85962835 and chrB2.89690057 on chromosome 5, chrB3.8555453 and chrB3.153838905 on chromosome 6, chrB4.143996116 and chrB4.159170769 on chromosome 7, chrC1.199755762 and chrC1.202770816 on chromosome 8, and chrD2.4845925 on chromosome 11 and chrUn5.7601427 on chromosome 14. Genomic inflation factor based on median chi-squared was observed to be 2.13033, Mean chi-squared statistics is 2.43179 for 48500 tested SNPs (Fig. 1).

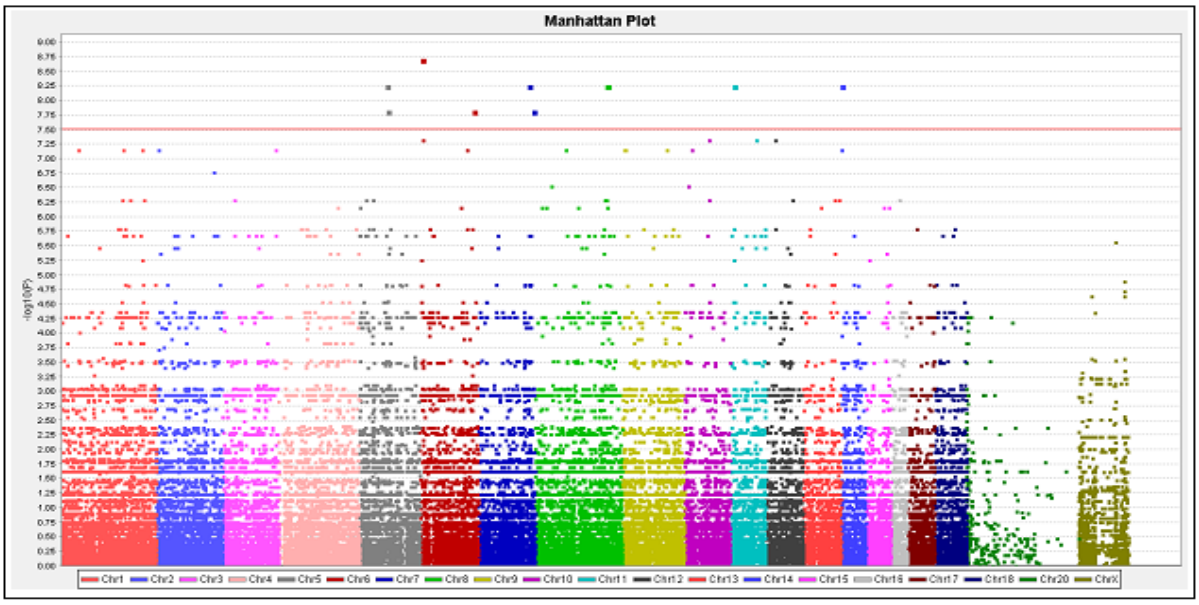

Fig. 1. Manhattan plot showing most associated SNPs.

Cochran mantel haenszel, SNP disease cluster

Cochran Mantel Haenszel SNP-disease cluster test determines whether the SNP cluster changes over time in a population. A graph was obtained with minor allele frequency on y-axis and chromosome number on x-axis (Fig. 2)The sample population was also tested for Mendel error which gives the summary of error rate in each SNP, individual and family(Liu *et al.*, 2013). In the current scenario, neither the affected offspring nor phenotypically discordant parents pair was found and overall zero Mendel error was observed.

Hardy-weinberg equilibrium

The Hardy Weinberg test is applicable to founders (parents) only. Total of 19 markers were excluded based on Hardy Weinberg Equilibrium (HWE), *p<= 0.001*. The sample does not follow Hardy Weinberg equilibrium as indicated by the lowest Chi-Square value=0.0 which mean that the difference in the observed and expected heterozygosity is 0% by chance, and 100% induced due to selective breeding.

These results account for the linkage disequilibrium among the SNPs in the population with low chances of recombination (Graffelman *et al.*, 2013).

Haplotype association and LD plots

The purpose of haplotype association test is to determine the most associated haplotype with this cancer. Most associated SNPs from allelic association test were taken as reference to obtain LD plots. ChrB2.89690057SNP was used as reference to draw the LD plot.

This SNP is located on the *"q"* arm of chromosome 5 as identified by UCSC and is one of the most associated SNP as per association analysis, but no haplotype was found in this region. Furthermore, the LD plots for SNPs chrB3.8555453 and

chrB3.153838905 located on chromosome 6 showed no haplotypic association, i.e. these SNPs do not exist in Linkage Disequilibrium. In order to find out the haplotype associated with disease, two SNPs chrB4.143996116 and chrB4.159170769 present on chromosome 7 were also used to draw LD plots but same results were observed. Similar results were obtained for chrC1.199755762 of chromosome 8 and chrUn5.7601427 of chromosome 14.

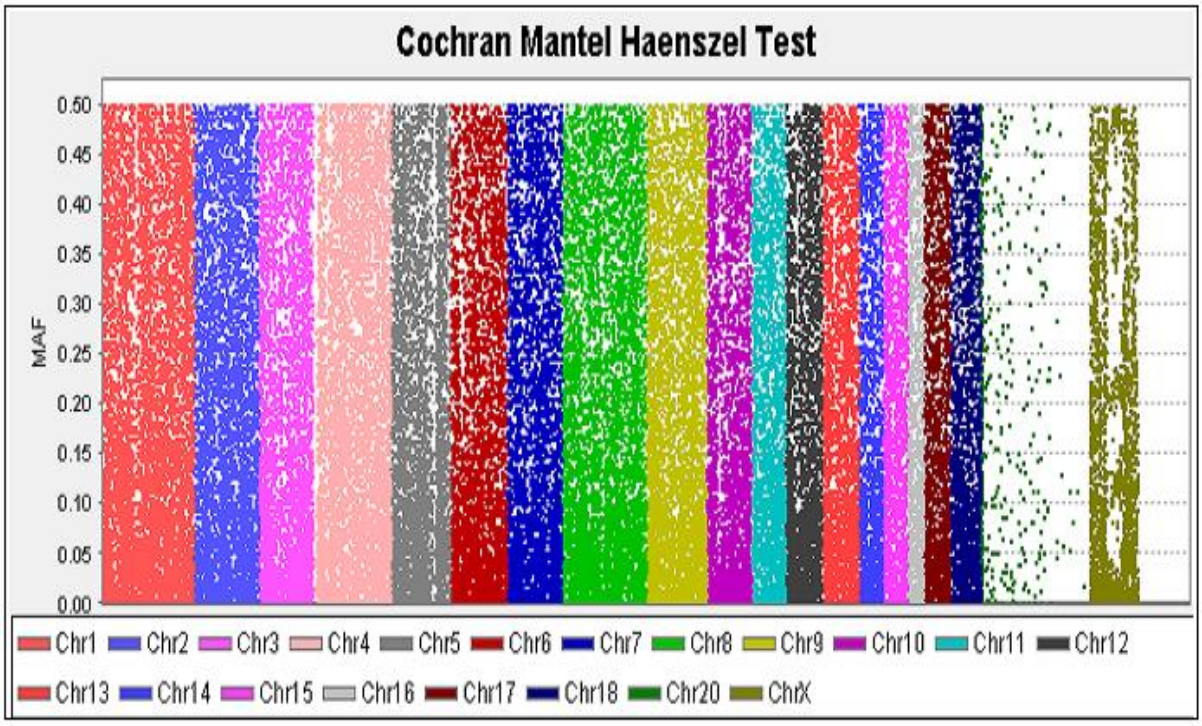

Fig. 2. Cochran Mantel Haenszel Analysis based on differences in allele frequencies between case/control studies.

LD map using SNP chrB2.85962835 located on chromosome 5 was obtained as reference and a haplotype block was found in the region containing two SNP s,i.e. chrB2.85917464 and chrB2.85931038 but no gene was found in the location of this haplotype (identified using ENSEMBL genome browser) (Fig. 3a).

Another LD plot obtained using SNP chrD2.4845925 located on *"p"* arm of chromosome 11 asa reference and associated with a single haplotype block that contains two SNPs, i.e. chrD2.4914239 and chrD2.4939794 (Fig. 3b).

When the locations of these SNPs were identified using an ENSEMBL genome browser, there was no gene identified in this region.

LD plot were also obtained by using SNP chrC1.202770816 as reference (Fig. 3c). This SNP is located on chromosome 8 and found associated with a single haplotype block of two SNPs, i.e. chrC1.202822232 and chrC1.202860978.

When the location of these SNPs was identified using the ENSEMBL genome browser, it was found that these SNPs were part of a gene RFTN2, i.e. raftlin family member 2 which is a protein coding gene(Fig. 3) and integral part of lipid rafts involved in regulating signal transduction in BCR. Its deficiency in T cells can reduce T-cell dependent antibody production and it's over expression and suppression can be associated with weak immune response (Saeki *et al.*, 2003; Barrett *et al.*, 2005; Clarke *et al.*, 2011) (Fig. 3).

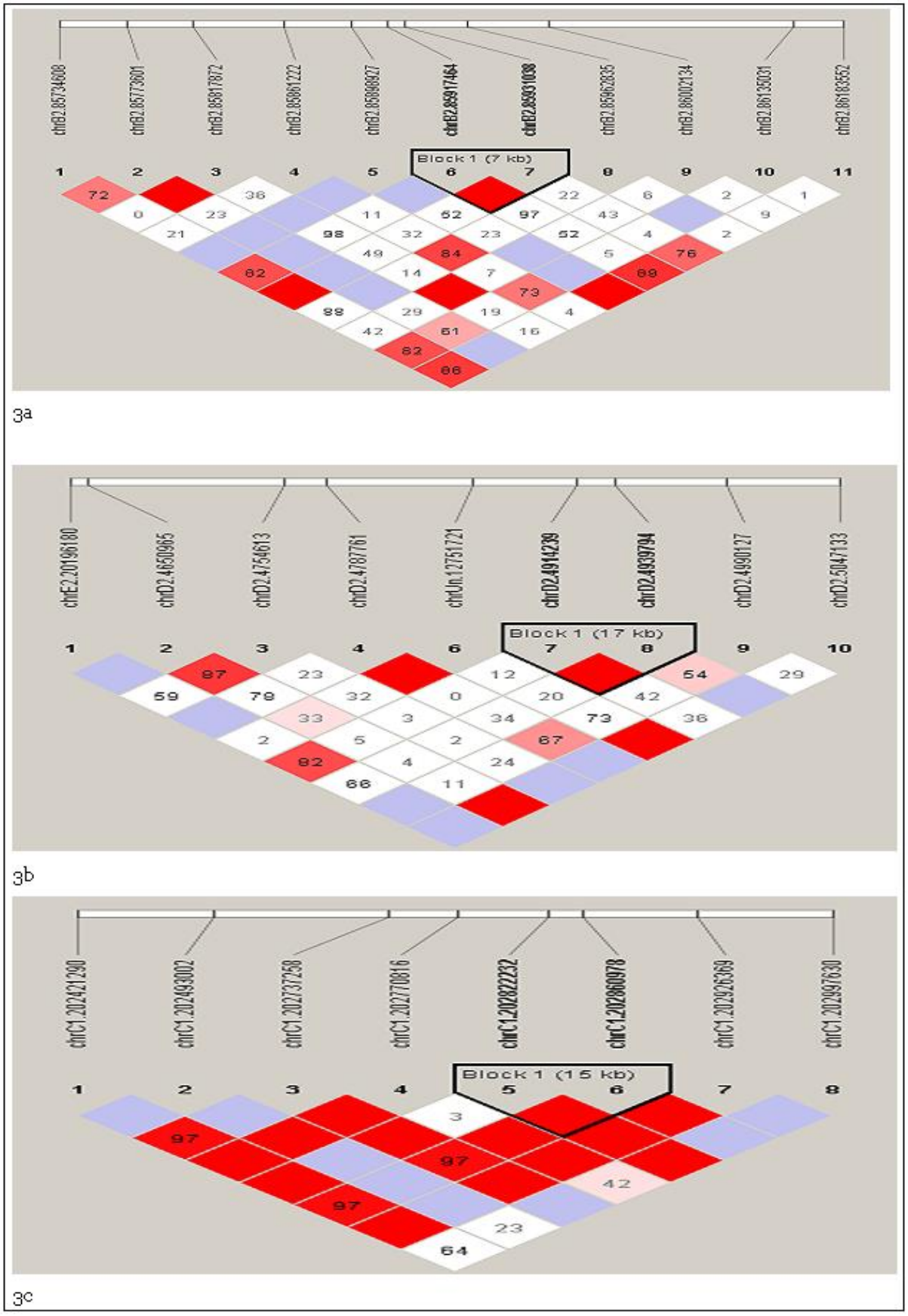

Fig. 3. LD Plot of chr B2.85962835 containing a haplotype block (a), LD plot of chrD2.4845925. Haplotype block was obtained (b), LD plot of chrC1.202770816. Haplotype block consisting of two SNPs was obtained (c).

Many genes such as ERBB2, BRCA1 and BCAS3 are associated with breast/mammary cancer, not only in humans but also in domestic animals i.e. cats. These genes and SNP located on chromosome number 14 in cats and selected as reference to obtain LD plot. SNP chrE1.53681930 located on the "*q*" arm of chromosome 14 identified by the UCSC genome browser, which showed a single large block (Fig. 4) and consists of twelve SNPs indicating that these SNPs are associated together and exist in the form of a haplotype (Fig. 4).

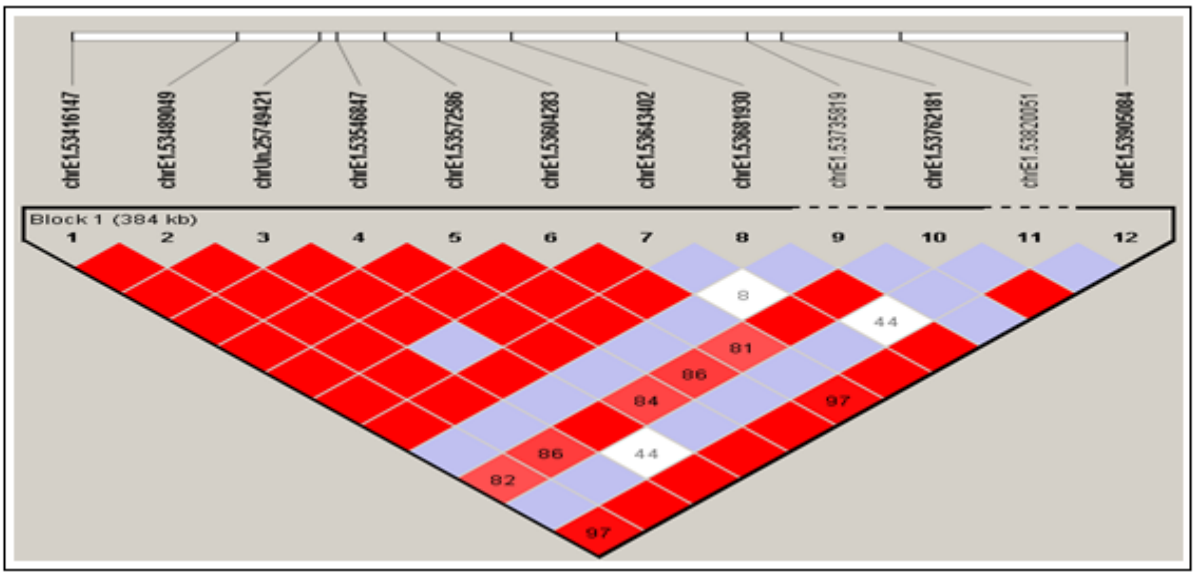

Fig. 4. LD plot of chrE1.53681930 haplotype block consisting of 12 SNPs.

When the location of these SNPs were identified using ENSEMBL genome browser, it was found that these SNPs are part of number of different genes which are C7orf64, APPBP2, PPMID and BCAS3 (Breast Carcinoma Amplified Sequence 3)(Fig. 4).These genes play their role e.g.C7orf64is an RNA binding protein. APPBP2 (amyloid protein binding protein 2) interacts with microtubules, which is involved in transport and processing of beta-amyloid precursor proteins and highly expressed in breast cancer. PPMID is a member of the protein phosphatases 2C family, its expression is induced in p53 dependent pathway. This gene is located in a chromosomal region known to be amplified in breast cancer. Its amplification is detected not only in breast cancer cell lines but also in primary breast tumors. BCAS 3 gene is associated with breast cancer as well (Singer-Hashler *et al*., 2012; Goldstein *et al*., 2013; Palomba *et al*., 2015).

Analysis of these LD plots showed that SNPs that are part of genes other than those associated with mammary tumor i.e BCAS3, BRCA1 and BRCA2 can also be associated with this mammary tumor in cats. It also shows that cancer/tumor is not caused by single gene located on a single chromosome but a accumulative effect of multiple SNPs or genes that may located on different chromosomes.

Acknowledgements

Authors are thankful to HEC-Pakistan and MU-USA for funding this project. We are also highly obliged to IBBt and Pet center UVAS-Pakistan.

References

Barrett JC, Fry B, Maller J, Daly MJ. 2005. Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics **21(2),** 263-265. http://dx.doi.org/10.1093/bioinformatics/bth457

Bei JX, Li Y, Jia, WH, Feng BJ, Zhou G, Chen LZ, Feng QS, Low HQ, Zhang H, He F. 2010. A genome-wide association study of nasopharyngeal carcinoma identifies three new susceptibility loci. Nature genetics **42(7),** 599-603. http://dx.doi.org/10.1038/ng.601

Bush WS, Moore JH. 2012. Genome-wide association studies. PLoS Comput Biol **8(12)**, e1002822.
http://dx.doi.org/10.1371/journal.pcbi.1002822

Chu R, Sun T, Yang H, Wang D, Liao K, Chuang T, LinC, Lee W. 2001. Heat shock proteins in canine transmissible venereal tumor. Veterinary Immunology and Immunopathology **82(2),** 9-21.
http://dx.doi.org/10.1016/s0165-2427 (01)00327-0

Clarke GM, Anderson CA, Pettersson FH, Cardon LR, Morris AP, Zondervan KT. 2011. Basic statistical analysis in genetic case-control studies. Nature protocols **6(2),** 121-133.
http://dx.doi.org/10.1038/nprot.2010.182

Goldstein O, Mezey JG, Schweitzer PA, Boyko AR, Gao C, Bustamante CD, Jordan JA, Aguirre GD, Acland GM. 2013. IQCB1 and PDE6B Mutations Cause Similar Early Onset Retinal Degenerations in Two Closely Related Terrier Dog BreedsRetinal Degenerations and Terrier Dog Breeds. Investigative ophthalmology & visual science **54(10),** 7005-7019.
http://dx.doi.org/10.1167/iovs.13-12915

Graffelman J, Sánchez M, Cook S, Moreno V. 2013. Statistical inference for Hardy-Weinberg proportions in the presence of missing genotype information. PLoS One **8(12),** e83316.
http://dx.doi.org/10.1371/journal.pone.0083316

Holcomb Jr WL, Chaiworapongsa T, Luke DA, Burgdorf KD. 2001. An odd measure of risk: use and misuse of the odds ratio. Obstetrics & Gynecology **98(4),** 685-688.
http://dx.doi.org/10.1097/00006250-200110000-00028

Klein RJ, Zeiss C, Chew EY, Tsai JY, Sackler RS, Haynes C, Henning AK, SanGiovanni JP, Mane SM, Mayne ST. 2005. Complement factor H polymorphism in age-related macular degeneration. Science **308(5720),** 385-389.
http://dx.doi.org/10.1126/science.1109557

Liu L, Zhang D, Liu H, Arendt C. 2013. Robust methods for population stratification in genome wide association studies. BMC bioinformatics **14(1),** 132.
http://dx.doi.org/10.1186/1471-2105-14-132

Liu X, Invernizzi P, Lu Y, Kosoy R, Lu Y, Bianchi I, Podda M., Xu C., Xie G., Macciardi F. 2010. Genome-wide meta-analyses identify three loci associated with primary biliary cirrhosis. Nature genetics **42(8),** 658-660.
http://dx.doi.org/10.1038/ng.627

Mägi R., Morris AP. 2010. GWAMA: software for genome-wide association meta-analysis. BMC bioinformatics **11(1),** 288.
http://dx.doi.org/10.1186/1471-2105-11-288

Palomba G, Loi A, Porcu E, Cossu A, Zara I, Budroni M, Dei M, Lai S, Mulas A, Olmeo N. 2015. Genome-wide association study of susceptibility loci for breast cancer in Sardinian population. BMC cancer **15(1),** 383.
http://dx.doi.org/10.1186/s12885-015-1392-9

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, De Bakker PI, Daly MJ. 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. The American Journal of Human Genetics **81(3),** 559-575.
http://dx.doi.org/10.1086/519795

Saeki K, Miura Y, Aki D, Kurosaki T, Yoshimura A. 2003. The B cell- specific major raft protein, Raftlin, is necessary for the integrity of lipid raft and BCR signal transduction. The EMBO Journal **22(12),** 3015-3026.
http://dx.doi.org/10.1093/emboj/cdg293

Shafiee R, Javanbakht J, Atyabi N, Bahrami A, Kheradmand D, Safaei R, Khadivar F, Hosseini E. 2013. RETRACTED ARTICLE: Comparative value of clinical, cytological, and histopathological features in feline mammary gland tumors; an experimental model for the study of human breast cancer. Diagnostic pathology **8(1),** 136.
http://dx.doi.org/10.1186/1746-1596-8-136

Sherman M, Multhoff G. 2007. Heat shock proteins in cancer. Annals of the New York Academy of Sciences **1113(1),** 192-201.
http://dx.doi.org/10.1196/annals.1391.030

Signer-Hasler H, Flury C, Haase B, Burger D, Simianer H, Leeb T, Rieder S. 2012. A genome-wide association study reveals loci influencing height and other conformation traits in horses. PLoS ONE**7(5),** e37282.
http://dx.doi.org/10.1371/journal.pone.0037282