INTERNATIONAL JOURNAL OF SCIENCE AND NATURE

www.scienceandnature.org

Review Article

CONCEPT OF GENOME WIDE ASSOCIATION STUDIES AND ITS PROGRESS IN LIVESTOCK

*Binoy Chandra Naha, Atherajendra Prasad, Lalrengpuii Sailo, Rajni Chaudhary & Om Prakash
Animal Genetics Division, ICAR- Indian Veterinary Research Institute, Izatnagar, Bareilly, U.P-243122
*Corresponding author email: binoychandranaha@gmail.com

ABSTRACT
Genome-wide association studies (GWAS) have a powerful strategy to detect the genetic contributors to complex traits in livestock. GWAS are based on discovery of new genetic variants that effect a phenotype of animal. For GWAS, we consider ethical considerations, study design, selection of phenotype of animal, power considerations, sample tracking, storage issues and genotyping product selection. During execution of GWAS, important steps involves DNA quantity and preparation, genotyping methods, quality control checks of genotype data, imputation, tests of association and replication of association signals. The field of animal genetics can help guide an investigator in making practical and methodological choices that will help to determine the overall quality of GWAS results. GWAS is helpful to be aware of these aspects to maximize the likelihood of success, mainly where there is an opportunity for implementing them prospectively in livestock.

KEYWORDS: GWAS (Genome-wide association studies), QTL, Livestock

INTRODUCTION

In 90's, QTL mapping was largely based on microsatellite markers (Lipkin *et al*., 1988). Quantitative trait loci (QTL) were the preferred method to detect genetic variation for economically important traits at the genomewide level. Reported QTLs were detected based on microsatellite markers with low resolution map and the confidence interval (CI) covers more than 20cm. So, it is very difficult to identify the important genes for economic traits of interest based on the information. The detection of causal mutations that underlying QTLs has been targeting in domestic animals. Till date, hundreds of QTLs have been identified in various independent studies (Sharma *et al*., 2015; Monolio *et al*., 2010). GWAS is a new technique for the identification of causal genes for economically important traits in animal (Pearson and Manolio, 2008; Johnson and O'Donnell, 2009). SNP along with the phenotype and pedigree information are utilized for gene mapping. GWAS has become feasible in humans as well as in livestock animals as a result of the development of large amount collections of SNPs and the development of cost-effective methods for large-scale SNP analysis (Bush and Moore, 2012; Altmuller *et al*., 2001). As compared to traditional QTL mapping strategies, GWAS covers the major advantages both in the power to detect causal variants with modest effects and indicating the narrower genomic regions that harbor causal variants (Risch and Merikangas, 1996; Greely, 2007). GWAS is a new ideal technique to discover the major genes for complex traits and is a main way to study the genetic mechanism of complex traits of livestock (Ikram *et al*., 2010).

PUBLICATIONS ON GWAS

The first successful GWAS was published in 2005 (Klein *et al*., 2005) and investigated patients with age-related macular degeneration. It is seen that two SNPs which had significantly altered gene frequency when comparing with healthy controls. As of 2011, hundreds or thousands of individuals are tested and over 1,200 human GWAS have examined above 200 diseases and traits (November *et al*., 2008).

PROGRESS OF GWAS IN DOMESTIC ANIMALS

Genomic sequences were available for many livestock species. Large numbers of SNPs were discovered as a result of by-product of sequencing. In domestic animal, GWAS has gained popularity in mapping QTL of economic importance traits. Different regions and genes are found to be associated with the same trait in different breeds of the same species. GWAS has proved to be a noval method to identify genes associated with various phenotypes. GWAS was first used in the analysis of human and animal disease and great progress was created. GWAS was extended to the field of livestock genetics and breeding when genomic sequences were available for many domestic species and large numbers of SNPs were discovered due to by-product of sequencing or in subsequent re-sequencing. There are several of commercial SNP chip available for cattle (50,000 SNPs; Illumina BovineSNP50 BeadChip), dogs (22,362 SNPs; Illumina CanineSNP20 BeadChip), sheep (56,000 SNPs), pigs (60,000 SNPs; Illumina PorcineSNP60 BeadChip), horses (54,602 SNPs; Illimina EquineSNP50 BeadChip) and chickens (60,000 SNPs; Illumina ChickenSNP60 BeadChip) reported by (Zhang *et al.*, 2012). The

application of GWAS to livestock has only occurred relatively recently and there have been a series of results reported, especially from the analysis of the genetic mechanisms of complex traits.

FIRST GWAS IN DIFFERENT ANIMALS

ANIMAL	YEAR	SNP CHIP	TRAIT	REFERENCES
Cattle	2009	50k	Meat quality	Zhang *et al*.(2012)
Buffalo	2013	90k	Milk yield	Hndorff *et al.(*2009)
Pig	2012	60k	Androstenone	Zhang *et al*.(2012)
Horse	2010	50k	Racing distance	Hill *et al*.(2010)
Sheep	2011	50k	Horn morphology	Johnston *et al*.(2011)
Dog	2010	50k	DM	Zhang *et al*.(2012)
Chicken	2007	3k	Fatness	Zhang *et al*.(2012)

ASSUMPTIONS IN THE ANALYSIS OF GWAS

Significant associations can be identified due to the SNPs are in linkage disequilibrium (LD) with the causative mutations for the complex traits of interest. The high density of SNP markers in the chip used in GWAS was sufficient to detect the LD between SNP markers and causative mutations (Barsh *et al*., 2012; Danesh and Pepys, 2009).

ADVANTAGES OF GWAS

1. Potential to discover new candidate genes not identified through other methodological processes (Haines *et al*., 2005; Ehret *et al.*, 2011).
2. Rules out specific genetic association (Loannidis *et al.*, 2009).
3. Identifies the mutations explaining few percent of phenotypic variant (Kathiresan *et al*., 2009).
4. Biological pathway of the trait does not have to be known (Mather *et al*., 2008).

STATISTICAL SIGNIFICANCE CRITERIA IN GWAS

To obtain statistical significance, a GWAS must include at least 100,000 markers and most of which are inherited SNPs or copy number variants (CNVs). The large numbers of statistical tests that are necessary mainly increase the likelihood of false positives (Schera *et al*., 1995; Sanna *et al*., 2011). Due to this risk, the accepted threshold for statistical significance in GWAS is $< 5*10^{-8.}$

INDICATIONS OF GWAS

Correct statistical methods are necessary to reduce the risk of multiple false positive results and avoid GWAS when statistical power is limited by small sample size (Folkersen *et al.*, 2010). Larger sample sizes are required to identifly an association when multiple genes are involved in a trait (Fareed and Afzal, 2013; Thomas *et al*., 2009).

LIMITATION OF GWAS

1. A large study of population is required which detects association, not causation (Purcell *et al*., 2007).
2. Identifies specific location not complete gene (Marchini and Howie, 2010).
3. Focus on common variants and many associated variants are not causal.
4. Unavailability of funding agencies (Howie *et al*., 2011; Sebastiani *et al*., 2011).
5. GWAS have many issues and limitations that can be taken care of through proper quality control and study setup (Visscher *et al*., 2012).
6. Lack of well-defined case and control groups and insufficient sample size for population stratification are common problems (Visscher *et al*., 2011).
7. GWAS can be problematic due to massive number of statistical tests performed presents an unprecedented potential for false-positive results (Bauer *et al*., 2011).
8. A high profile GWA study investigating individuals with very long life spans in order to detect SNPs associated with longevity has been mentioned (Dube *et al*., 2011).
9. GWAS have attracted more fundamental criticism due to their assumption that common genetic (Wellome trust case control consortium, 2007).
10. More recently, the fastly decreasing price of complete genome sequencing have also provided a realistic alternative to genotyping array-based GWAS (Scott and Hebbring., 2014).

CHALLENGES OF GWAS

Mixed models that handle population structure by accounting for the amount of phenotypic covariance (Clarke *et al*., 2011). Mixed models have been applied to GWAS, and can markedly reduce the number of false positive associations. Lack of statistical knowledge remains one of the major glitches in GWA projects (Hill *et al*., 2010).

APPLICATION OF GWAS

1. To associate between the variations in genotypes and phenotypes to identify the causal genetic mechanism.
2. To identify QTL underlying many common, complex disease.
3. To associate a trait with a region in the genome in order to map the clinically and economically important QTLs.

FUTURE OF GWAS

GWAS in domestic animals will focus on the identification of causative mutations for economically important traits. The findings will inevitably facilitate the understanding of the genetic architecture of complex traits in domestic animals and practical improving the breeding programmes. In future, further, understanding of the roles of epistasis (gene–gene interactions), gene–environment interactions, and copy number variants are anticipated to

provide additional insights into our understanding of complex human and animal disorders.

NEW ADVANCES IN GWAS

High density SNP chips – Up to 7 lakh SNPs. Whole Genome re-sequencing. Functional genomics – selection based on function of each gene.

RECENT SUCCESS IN ANIMAL GENETICS

GWAS of footrot in Texel sheep: This is the first study based on a genome-wide association approach that investigates the links between ovine footrot scores and molecular polymorphisms in Texel sheep using the ovine 50 K SNP array after quality control. Our aim was to identify molecular predictors of footrot resistance.

CONCLUSION

Genetics has come a long way since 1983 when the genetic markers were first used for the improvement of crops and livestock .With ever advancing technology and better knowledge of genetic mechanisms; we are surely a step closer to the understanding of complex traits. The challenges of GWAS include carefully choosing a homogeneous population for the study and to account for population stratification. The statistical models, if carefully chosen, can be useful to minimize the chances of false associations.

REFERENCES

Altmüller, J., Palmer, L.J., Fischer, G. & Scherb, H. (2001) Genome wide Scans of Complex Human Diseases: True Linkage Is Hard to Find. *Am. J. Hum. Genet.,* **69 (5):** 936–50.

Barsh, G.S., Copenhaver, G.P., Gibson, G & Williams, S. M. (2012) Guidelines for Genome Wide Association Studies. *PLoS Genet.,* **8 (7):** e1002812.

Bauer, R.C., Stylianou, I.M., Rader, D.J. (2011) Functional validation of new pathways in lipoprotein metabolism identified by human genetics. *Curr. Opin. Lipidol.,* **22 (2):** 123–8.

Bush, W.S. and Moore, J.H. (2012) Chapter 11: genome wide association studies. *PLoS Comput. Biol.,* **8 (12):** e1002822.

Clarke, G.M., Anderson, C.A., Pettersson, F.H., Cardon, L.R. (2011) Basic statistical analysis in genetic case control studies. *Nat Protoc.,* **6 (2):** 121– 33.

Danesh, J. and Pepys, M.B. (2009) Creactive protein and coronary disease: is there a causal link. *Circulation,* **120 (21):** 2036–9.

Dubé, J.B., Johansen, C.T. and Hegele, R.A. (2011) Sortilin: an unusual suspect in cholesterol metabolism: from GWAS identification to in vivo biochemical analyses, sortilin has been identified as a novel mediator of human lipoprotein metabolism. *BioEssays,* **33 (6):** 430–7.

Ehret, G.B., Munroe P.B., Rice K.M., Bochud, M. (2011) Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk. *Nature,* **478:** 103–9.

Fareed, M. and Afzal, M. (2013) Single nucleotide polymorphism in genomewide association of human population: A tool for broad spectrum service. *Egypt. J. Med. Hum. Genet.,* **14:** 123–134.

Folkersen, L., Hooft F., Chernogubova, E. and Agardh, H. E. (2010) Association of genetic risk variants with expression of proximal genes identifies novel susceptibility genes for cardiovascular disease. *Circ. Cardiovasc. Genet.,* **3 (4):** 365–73.

Greely, H.T. (2007) The uneasy ethical and legal underpinnings of large scale genomic biobanks. *Annu. Rev. Genomics. Hum. Genet.,* **8:** 343–64.

Haines, J.L., Hauser, M.A., Schmidt, S. and Scott, W.K. (2005) Complement Factor H Variant Increases the Risk of Age Related Macular Degeneration. *Science,* **308:** 419–421.

Hill, E.W., Gu, J., Eivers, S.S., Fonseca R.G., McGivney, B.A., Govindarajan, P., Orr, N., Katz, L.M., Machugh, D. E. (2010) A sequence polymorphism in MSTN predicts sprinting ability and racing stamina in thoroughbred horses. *PLoS One,* **5:** e8645.

Hindorff, L.A., Sethupathy, P., Junkins, H.A. and Ramos, E.M. (2009) Potential etiologic and functional implications of genome wide association loci for human diseases and traits. *Proc. Natl. Acad. Sci. U.S.A.,* **106(23):** 9362–7.

Howie, B., Marchini, J. & Stephens, M. (2011) Genotype imputation with thousands of genomes. *Genes/ Genomes/ Genetics.* **1(6):** 457–70.

Iadonato, S.P. and Katze, M.G. (2009) Genomics: Hepatitis C virus gets personal. *Nature,* **461:** 357–8.

Ikram M.K., Sim, X. and Xueling, S. (2010) Four novel Loci (19q13, 6q24, 12q24, and 5q14) influence the microcirculation in vivo. *PLoS Genet.,* **6 (10):** e1001184.

Ioannidis, J.P., Thomas, G. & Daly, M.J. (2009) Validating, augmenting and refining genome- wide association signals. *Nav. Rat. Genet.,* **10(5):** 318-29

Johnson, A.D. & O'Donnell, C. J. (2009) An Open Access Database of Genomewide Association Results. *BMC Med. Genet.,* **10:** 6.

Johnson, T., Gaunt, T.R., Newhouse, S.J & Padmanabhan S. (2011) Blood Pressure Loci Identified with a Gene Centric Array. *Am. J. Hum. Genet.,* **89 (6):** 688–700.

Kathiresan, S., Willer, C.J., Peloso, G.M and Demissie, S. (2009) Common variants at 30 loci contribute to polygenic dyslipidemia. *Nat. Genet.,* 41 (1): 56– 65.

Klein, R.J., Zeiss, C., Chew, E.Y. and Tsai, J.Y (2005) Complement Factor H Polymorphism in Age Related Macular Degeneration. *Science*, **308:** 385–9.

Ku, C.S., Loy, E.Y., Pawitan, Y. and Chia, K.S. (2010) The pursuit of genomewide association studies: where are we now. *J. Hum. Genet.,* **55 (4):** 195–206.

Lipkin E., Mosig M.O., Darvasi A., Ezra E., Shalom, A., Freedman A and Soller M. (1988) Quantitative trait locus mapping in dairy cattle by means of selective milk DNA pooling using dinucleotide microsatellite markers. *Genet.*, **149:** 1557-1567.

Maher, B. (2008) Personal genomes: The case of the missing heritability. *Nature*, **456:** 18–21.

Manolio, T.A., Guttmacher, A.E. and Manolio, T.A. (2010) Genome wide association studies and assessment of the risk of disease. *N. Engl. J. Med.,* **363(2):** 166–76.

Marchini, J. and Howie, B. (2010) Genotype imputation for genome wide association studies. *Nat. Rev.Genet.*,**11 (7):** 499–511.

Novembre, J., Johnson, T., Bryc, K. and Kutalik, Z. (2008) Genes mirror geography within Europe. *Nature*, **456**: 98–101.

Pearson, T. A. & Manolio, T. A. (2008) How to interpret a genome wide association study. *JAMA.,* **299(11)**: 1335–44.

Purcell, S., Neale, B., Toddbrown, K. and Thomas, L. (2007) PLINK: A Tool Set for Whole Genome Association and Population Based Linkage Analyses. *Am. J. Hum. Genet.,* **81 (3):** 559–75.

Risch, N. and Merikangas, K. (1996) The future of genetic studies of complex human diseases. *Science* **273** : 1516-1517.

Sanna, S., Li, B., Mulas, A., Sidore, C and Kang, H.M. (2011) Fine mapping of five loci associated with low-density lipoprotein cholesterol detects variants that double the explained heritability. *PLoS Genet.,* **7 (7):** e1002198.

Schena, M., Shalon, D., Davis, R.W and Brown, P.O. (1995) Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science*, **270:** 467–70.

Scott, J. & Hebbring (2014) The challenges, advantages and future of phenome-wide association studies. *Immunology,* **3(1):** 25.

Sebastiani P., Solovieff N., Puca A and Hartley S. W. (2011) Retraction. *Science*, **333:** 404.

Sharma, A., Leea, J.S., Dang C.G., Sudrajad, P., Kim, H.C., Yeon S.H. and Lee S.H. (2015) Stories and Challenges of Genome Wide Association Studies in Livestock—A Review. *Asian. Australas. J. Anim. Sci., 5(1): 27.*

Thomas D.L., Thio C.L., Martin M.P., Qi Y and Ge D. (2009) Genetic variation in IL28B and spontaneous clearance of hepatitis C virus. *Nature*, **461:** 798–801.

Visscher P.M., Brown M.A., McCarthy M.I and Yang, J. (2012) Five years of GWAS discovery. *Am. J. Hum. Genet.,* **90 (1):** 7–24.

Visscher, P.M., Goddard, M.E., Derks, E.M and Wray, N.R. (2011) Evidence based psychiatric genetics, AKA the false dichotomy between common and rare variant hypotheses. *Mol. Psychiatry.*, **17 (5):** 474–85.

Wellcome Trust Case Control Consortium (2007) Genome wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature*, **447:** 661–78.

Zhang, H., Wang, Z., Wang, S and Li, H. (2012) Progress of genome wide association study in domestic animals. J. Anim. Sci. Biotech., **3(1):** 26.