Verma et al., *Journal of Innovation in Applied Research* (2023), 6(1): 56-64
https://doi.org/10.51323/JIAR.2023/6/1/56-62

Journal of Innovation in Applied Research
eISSN: 2581-4206

Genome Wide Association Studies: An Outlook.

Apoorva Verma, Raj Kumar, Shubhang Shukla, Swarnima Singh*, Parul Johri and Manish Singh Rajput

Received: 28 December 2022
Accepted: 27 January 2023
First Published: 18 February 2023

*Corresponding author: Swarnima Singh, Department of Biotechnology, Ambedkar Institute of Technology for Handicapped Kanpur, Dr
E-mail:
swarnimasinghshash@gmail.com

Reviewing editor: Kamlesh Choure

Additional information is available at the end of the article

Abstract: Genome-wide association studies (GWAS), which have been used in human disease research for the past ten years, are now the de facto method for finding new genes. While although there is ongoing discussion about how to make the most of these studies and occasionally about how much value they actually add, it is evident that many of the most compelling findings have come from large-scale mega-consortia and/or meta-analyses that pool information from several studies and tens of thousands of participants. Even though these studies are growing more and more prevalent, statistical techniques have lagged behind. There are effective meta-analysis techniques available, but even when they are used carefully and to their full potential, some statistical problems persist. The GWAS meta-analysis literature is comprehensively reviewed in this article, with a focus on methodology, software choices, and methodologies that have been applied in actual investigations. Using a case study, we highlight how various approaches differ from one another. We also talk about some of the open questions and possible directions in the future.

Keywords: GWAS; Gene metaanalysis; PAST; QTL analysis; SNP

1.

2. Introduction

Here, we survey the striking scope of revelations that Genomic wide association studies (GWASs) have worked with in population and complex-quality hereditary qualities, the science of sicknesses, and interpretation toward new therapeutics. In the early on segments, we give a foundation to this background, sum up its degree and design, and return to the logical reasoning for GWASs. We then audit general ends that can be drawn from GWAS revelations across many attributes. We in this manner feature more unambiguous aftereffects of disclosures and techniques on the way from GWAS to science and survey progress in three exemplar diseases, namely type 2 diabetes, auto-immune diseases, and schizophrenia. We end the review with various areas on the restrictions of current exploratory planswhat's more, potential ways of conquering these and an expectation onthe eventual fate of GWASs for human characteristics.

GWAS Meta analysis and methods

Candidate gene association studies have been generally used to study Hereditary defencelessness to complex diseases, including cancer (1). Critics of quality investigations have highlighted non-replication of results, misleading positives, inadequate sample sizes, and restricted earlier information on naturally pertinent up-and-comer genes (2). These worries have incited the utilization of efficient audits, particularly meta-investigations of different studies, to limit misleading positive affiliations and evaluate the believability of findings (Ioannidis JP 2006 et al). In recent, genome wide association studies (GWAS) have incredibly sped up the speed of disclosure and found numerous clever hereditary affiliations that were not expected by the quality approach (3,4). Associations found by GWAS bring up unexpected issues, especially on the grounds that noticed impacts are ordinarily very small (5). Besides, the implicated SNPs address markers that require further examination to distinguish causal variants (6), although this might turn out to be to a lesser degree an issue as strategies for fine mapping affiliations get to the next level.

© 2023 The Author(s). This open access article is distributed under a Creative Commons Attribution (CC-BY) 4.0 license.

To assist the immense measure of data from both candidate gene and GWAS of malignant growth, the the Centers for Disease Control and Prevention's (CDC) Office of General Well-being Genomics and the Public Disease Establishment's Division of Malignant growth Control and Populace Sciences sent off the Disease GAMAdb in2010 (7). This persistently refreshed data set indexes distributed GWAS and meta-investigations and pooled examinations that have assessed relationship of hereditary polymorphisms and malignant growth risk since January 1, 2000. Disease GAMAdb assembles on a distributed informational index by Dong et al (Dong LM 2008), which incorporated meta-examinations and pooled investigations of hereditary polymorphisms, and disease risk distributed. Relationship in the data set distributed after that date have been distinguished utilizing the Human Genome the study of disease transmission (Tremendous) Guide database (8) and the Public Human Genome Exploration Establishment (NHGRI) GWAS catalog (9). The Habitats for Infectious prevention and Avoidance's Enormous Guide is a consistently refreshed information base in HuGE (10). The NHGRI GWAS index extricates information from GWAS publications. Hereditary affiliations with malignant growth are chosen from these two data sets for curation in the Disease GAMAdb. Information depicting the association(s) — including concentrate on populace, minor allele frequencies, and impact sizes — are physically separated from each article and went into the Malignant growth GAMAdb. The ongoing investigation depends on the information that were remembered for Malignant growth GAMAdb as of February 26, 2011.

Our analysis considered the extent to which associations reported in meta analyses and GWAS overlapped. When both types of studies reported associations with the same variant, we called the overlap direct. When they reported associations with variants separated by less than 1 million base pairs, we called the overlap indirect. In an additional analysis, we also examined noteworthy associations, which we defined as those with false-positive report probabilities (FPRP) r0.2, a stringent threshold suggested by Wacholder et al,12 and used in the analysis by Dong et al.8 We calculated FPRPs at two levels of prior probability and at two levels of association (OR 1.5 and OR 1.2). As in the analysis by Dong et al,8 we chose to evaluate the associations using a low-prior probability of 0.001 (expected for a candidate gene) and a very low-prior probability of 0.000001 (expected for a random SNP). An association was considered noteworthy if it passed the FPRP threshold in one or more of these four categories.

We identified 5131 gene-variant associations with incident cancer from 386 meta-analyses and pooled analyses published after the review by Dong et al review. We excluded 3828 (74.6%) associations because their reported P-values were $\geq$0.05; 1026 more were excluded

for reasons. After applying all exclusion criteria, we found 277 significant associations; the review by Dong et al included 98 significant associations. Twenty-six (7.4%) of these were also found in meta-analyses published since the paper by Dong et al. Thus, there were 349 unique variant-cancer associations in all, involving 264 genes (76 with more than one associated variant) and spanning 25 different cancer types. The largest number of candidate gene associations was found for breast cancer (n=80) followed by prostate cancer (n=53). Significant associations from meta-analyses and pooled analyses of candidate genes are listed in Supplementary Table 1.

Abbreviations: ALL, adult lymphoblastic leukemia; GWAS, genome-wide association studies; MA, meta-analyses or pooled analyses; MCL, myeloid cell leukemia; NHL, non Hodgkin lymphoma.

a - Total significant associations reported in previous systematic review of meta-analyses (Donget al8) and meta-analyses and pooled data of individual studies published from 20 March2008 through 26 February 2011. Meta-analyses were defined as those of candidate genestudies. Significance threshold was 0.05.

b - From GWAS catalog. Excludes variants that were not reported. GWAS with meta-analyses included were considered GWAS. Significance threshold was 1_10_5.

c - Some variants may be linked to one another due to proximity. Associations with combinations

of two or more variants were considered unique, even if listed standalone variants were also reported.

d - Intergenic regions used if no gene provided by paper or associated with variant**.**

Table 1. umber of significant associations (in variants and genes) reported in candidate gene meta-analysis and pooled analysis and GWAS, by cancer site.

	MA [a]		GWAS [b]	
Cancer Site	**Variants [c]**	**Genes [d]**	**Variants [c]**	**Genes [d]**
Bladder	15	14	10	10
Blood related	1	1		
Breast	80	59	36	30
Cervical	4	4		
Colorectal	30	23	17	14
Endometrial	2	1		
Esophageal	9	9	4	4
Gastric	21	17	2	2
Genitourinary	2	2		
Glioma	18	13	9	8
Head and Neck	14	11		
Hepatocellular	8	4	4	6
Hodgkin lymphoma			4	3
Laryngeal	2	2		
Leukemia	4	4	32	27
Lung	32	23	25	22
Meningioma	1	1		
Myeloprolifrative			1	1
Nasopharyngeal	4	3	6	6
Neuroblastoma			5	3
Non-hodgkin lymphoma	10	8	2	2
Oral	1	1		
Ovarian	14	12	10	10
Pancreatic			21	21
Prostate	53	40	56	35
Renal Cell			3	3
Skin	20	8	8	7
Testicular			12	10
Thyroid			2	2
Upper aero-digestive tract	2	2		
Upper aero-digestive tract and lungs	1	1		
Urothelial	1	1		
Total	**349**	**264**	**269**	**223**

Complications

Nowadays, GWAS meta-analysis is frequently employed and, in general, has been successful in identifying genetic effects that were not identified in individual research. But there are still certain obstacles and unresolved methodological problems.

2.1. Genotype data cleaning

It is imperative that all data sets go through extensive standard GWAS data cleaning procedures before meta-analysis, such as removing ”poor” SNPs and samples using genotyping call rates, tests of Hardy-Weinberg equilibrium (HWE), etc. (11). The significance of having consistent data cleaning procedures and standards across all data sets is not totally obvious. Can using various genotype call rate cutoffs in various data sets lead to issues, for instance? To our knowledge, this has not been systematically investigated. There are three strategies to deal with HWE in genetic association studies for specific SNPs: include all studies regardless of the HWE tests (12), sensitivity analysis is being performed to confirm different genetic influences in subgroups (13-16), and eliminating research with statistically significant HWE divergenceand eliminating research with statistically significant HWE divergence (13,17). Recent large consortium meta-analyses have made an effort to utilise uniform HWE cutoffs across trials, which is unquestionably the most secure method.

Furthermore, it is unclear whether implementing data cleansing procedures that compare various data sets is required or desirable. It is usual practise to search through data sets for SNPs with drastically varying allele frequencies and remove them before combining because the same SNP assay can react differently on several chips, or even on the same chip in different batches. However, there will be SNPs for which there are ”real” differences in allele frequency if the data sets are from various ethnic groups. It is unclear how to distinguish between the artefacts and the actual differences, making it challenging to suggest the best cleaning method. Similar problems arise with HWE testing when data sets are merged (as was mentioned above), but it is fairly obvious that HWE tests on combined data sets would be too conservative. These difficulties are especially crucial when there are disparate phenotypic distributions between research (or, equivalently, different case : control ratios).

3. Imputation

Direct SNP-by-SNP meta-analysis is not practicable when studies are genotyped on multiple chips since there may be very little overlap in the SNP sets. For instance, only roughly 100K or 20% of the SNPs in the Illumina 550K SNP set and the Affymetrix 500K SNP set match. The typical approach to this issue is to impute the genotypes of all SNPs in all samples, and there are several effective ways for doing so (18). Imputed genotypes have a little bit greater error rates and variances than non-imputed genotypes, which is a challenge that hasn’t been properly addressed in the literature. Imputation error rates are often quite low when done carefully. However, error rates may be higher in regions of the genome with poor SNP coverage or in ethnic groups that are underrepresented in the reference data set for imputation (usually HapMap or 1000 genomes). This problem, like data cleaning above, can be very serious if the distribution of phenotypes varies between research. A disparity between case and control variances can result from two studies with different case:control ratios, one of which is genotyped and the other imputed for a specific SNP. This can lead to false positive results. In contrast, imputation will produce ”genotypes” if one chip has extremely low coverage of a particular location, which actually conveys very little information. In this situation, the meta-analysis may produce falsely negative results since it is averaging in unhelpful data sets. This issue might have a regionally-smoothed meta-analysis as a solution, but as far as we know, no such techniques exist. Generally speaking, it is always a good idea to examine the data quality of replication results that are mostly dependent on imputed data.

4. Choice of genetic models

The fundamental association test in GWAS analysis might be based on a comparison of allele frequencies or on different statistical contrasts of genotype frequencies, such as an additive model, a dominant model, etc. The additive model is typically employed because it is the same model that is used for each SNP (19). In a meta-analysis, it is ideal to utilise the same model across all studies, but in post hoc combinations of analyses, this may not always be feasible. As far as we know, no one has investigated how such modification in the association model affects meta-analysis. Even though it would not match a Gaussian random effects model, it would undoubtedly create some degree of effect heterogeneity that would, at the very least, violate a fixed effects model. Similar problems develop when other variables or population stratification control strategies are applied in various research.

5. Between-study heterogeneity

As was already said, between-study heterogeneity in GWAS meta-analysis should probably be regarded as the norm. It is crucial to identify and document this variability because it can provide crucial biological insights, such as variations in the genetic regulation of male and female recombination. According to accepted wisdom in the statistical literature, the random effects model is preferable to the fixed effects model when heterogeneity is present or even likely. We argue that this may not be the best strategy for GWAS, because (i) often only a small number of research are integrated, which results in an estimate of the heterogeneity that is not precise (ii) A Gaussian random effects model typically does not fit the form of the heterogeneity. While we do advocate using forest plots as a key heuristic tool for identifying and comprehending heterogeneity, we also indicate that future research on random or mixed-effects models that offer a better fit to GWAS data may enhance analysis. We could fit a model that explicitly has different fixed male and female effects in our recombination scenario as we know that males and females are likely to be different.

PAST: Pathway Association Studies Tools

Finding the parts of the genome that affect complex features in maize and other crops has become highly popular because to genome-wide association studies (GWAS) (20-22). F statistics are typically used to evaluate thousands of single nucleotide polymorphism (SNP) markers. a p-value for the SNP-trait correlation and a connection with the trait. Individual connections between markers and traits that are significant enough to meet the false discovery rate (FDR, the percentage of then, a more thorough examination is conducted to identify false positives among all significant data for some level. Indications about the genetic foundation of the trait and suggestions for future improvements. Because the FDR threshold in GWAS may be as low as divided by the total number of SNPs being analysed, many real relationships may go undetected. The FDR threshold may not be met in complicated, polygenic traits where genes that have tiny eects on a trait, particularly if the association's eect value is modified by the environment. Additionally, the positive alleles of other genes in the same pathway must also be present for an allele to be useful to be detected because many alleles of genes may only express themselves in particular genetic backgrounds (22). There's a chance that the small sample size of the GWAS panel prevents these allelic combinations from existing. Thus, the strict FDR thresholds and the insufficient numbers of high-frequency polymorphisms present in most panels limit the statistical power of GWAS for discovering genes of minor effect.

The combined effects of several genes that are grouped together based on their similar biological function are the subject of metabolic pathway analysis (23-25). This is a potential strategy that can support GWAS by providing hints about the genetic underpinnings of a trait. Pathway analysis and association mapping have been used in medical research to discover biological insights missed when focusing on only one or a small number of genes that have highly significant associations with a trait of interest (23,26-28). These methods were initially developed to study differences in gene expression data in human disease studies (Subramanian et. al 2005). Studies on both plants and animals have just recently started to apply pathway analysis (29-30). Additionally, enormous data sets generated by other high-throughput techniques like

RNA sequencing, proteomics, and metabolomics can be interpreted using biologically relevant pathways.

In recent years, the genetic basis of complex characteristics in plants has been studied using GWAS-based metabolic pathway analysis as a discovery technique. Aflatoxin build up (31), corn ear worm resistance (32), and oil production (33) in maize were all studied using a pathway-based methodology. By combining GWAS and metabolic circuit analysis, all genetic sequences are taken into account. Irrespective of their magnitude, and collectively they may highlight which sequences are lead to systems that improve crops and that demand additional research and manipulation, instance as in the case of gene editing. Even though combined GWAS and pathway analyses were very effective in identifying related pathways, the studies were time-consuming and labor-intensive because the analysis tools were created in a mix of R, Perl, and Bash, and the results of one analysis had to be manually fed into the next. There was a dearth of a single, unified, and user-friendly instrument to carry out this pathway analysis.

The Pathway Association Study Tool (PAST) was created to make GWAS-based metabolic pathway analysis simpler and more effective. Although PAST was created for use with maize, it can also be applied to other species. Regardless of their strength or relevance, it tracks all relationships between SNP markers and traits. Based on linkage disequilibrium (LD) data, PAST divides SNPs into linkage blocks and selects a tagSNP from each block. The features of the tag SNP, such as the allele effect, R^2, and p-value of the original SNP-trait relationship discovered through the GWAS analysis, are then transferred to the gene(s) within a user-defined distance of the tagSNPs using PAST. The enrichment score (ES) and p-value for each pathway are calculated by PAST using the gene eect values. PAST is simple to use as a standalone R script, an online tool, or a downloaded R Shiny application. It takes as input TASSEL (Bradbury et. al 2007) files produced as output from General Linear or Mixed Linear Models (GLM and MLM) in table format, or files from any association analysis that have been similarly formatted, along with genomic annotations in GFF format and a metabolic pathways file. One line should represent each gene in the metabolic pathways file, and the columns should list the pathway ID, pathway name, and gene ID.

Figure 1. The process through which PAST processes genome-wide association study (GWAS) output data to identify metabolic pathways significantly associated with a trait of interest

QTL analysis and GWAS of plant breeding

One of the second-most important oilseed crops, rapeseed (Brassica napus L., genome AACC, 2n = 38) is primarily produced for vegetable oil and protein meal around the world (34-37). The ever-increasing demand from people makes it difficult to feed the growing population, and the generation of biofuels improves food security globally (38-39). Grain output is anticipated to rise by up to 50% by 2025 in order to meet the increased demand for food worldwide (40). Therefore, new plant varieties with superior agronomic features will continuously need to be produced in order to meet this growing need. Stress-inducing response, yield, and yield-related features are just a few examples of several agronomic variables that are governed by numerous genes and strongly influenced by the environment (41). The separation and isolation of complex qualities into a single chromosome locus, as well as their characterisation for each quantitative trait locus (QTL), are therefore imperatively noticed in order to develop the real mechanism of agronomic traits.

With the rapid advancement of sequencing technology and bioinformatics tools, QTL analysis has emerged as a highly significant, precise, and effective genotyping method that makes use of molecular markers (such as single nucleotide polymorphisms, or SNPs), allowing for the strict control of complex genomic traits (42). The 60 K illumines Infinium SNP array population for B. napus may be successfully translated to a gene-based, low-cost, and high-throughput genotype-based screening technique for gene mapping (43). This technique is incredibly effective at mapping the QTLs at a narrow-range genomic level to control the intended trait, and it may also be the source of supply markers for the required traits (44). The quick developments between the Brassica and Arabidopsis species have been extraordinarily sparked by a prior study. These findings demonstrated the identification of 12 genes and the eight quantitative

trait nucleotides (QTNs) that underlie seed weight. Additionally, BnAP2, a single gene-specific marker, was found (45).

An effective method for the association mapping of QTL traits is linkage disequilibrium (LD) mapping, also known as association mapping. LD mapping shows the statistical relationship between the genetic markers and phenotypes within the natural populations. The successful and promising method of partitioning complex features is known as genome-wide association studies (GWAS) (46-47). Avena sativa (48), Sorghum bicolor (49), Hordeum vulgare (50), Triticum aestivum (51), Glycine max (52), Oryza sativa (53), Zea mays (54), Arachis hypogaea (55), and Brassica napus (56) are only a few of the crop varieties for which GWAS has lately shown promise (57).

The application of QTL/GWAS approaches to the analysis of these crops will spread quickly to include cereal crops. As a result, the current study primarily focuses on the traits of the rapeseed QTL, which serve as a useful model for subsequent research. The current analysis also offers a benchmark summary of the most recent research, with a focus on rapeseed QTLs that suggest a key role in upcoming breeding plans. This work also emphasises the thorough information regarding single-locus and multi-locus GWAS techniques, which can increase the robustness of GWAS for complex genetic characteristics.

6. Conclusion

Researchers must be aware of the most effective techniques for carrying out that meta-analysis as the GWAS literature moves away from artificial ”replication” and toward the more statistically optimal direct combining of all available data in a meta-analysis framework. Although most research now employ sound procedures, there is still opportunity for improvement in many of the previously covered specifics. Planning research in a coordinated approach from the start would be ideal for addressing many of the potential improvements, but that is not always possible. For post hoc pairings of studies that may have considerable variation in chip, research population, environmental exposures, association tests, etc., better approaches are still required. Looking even further forward, it is necessary to revisit all of the difficulties raised above in light of the impending publication in journals throughout the field of meta-analyses of SNP data produced from sequencing studies.

Author details

Apoorva Verma

Raj Kumar

Shubhang Shukla

Swarnima Singh

Parul Johri

Manish Singh Rajput
Department of Biotechnology, Ambedkar Institute of Technology for Handicapped Kanpur, Dr.

Citation information

Cite this article as: Verma, A., Kumar, R., Shukla, S., Singh, S., Johri, P., & Rajput, M. S. (2023). Genome Wide Association Studies: An Outlook.. *Journal of Innovation in Applied Research, 6(1)*, 56-64. doi: 10.51323/JIAR.2023/6/1/56-62

References

Brown, M. E., & Funk, C. C. (2008). *Food Security under Climate Change*. Lincoln, NE, USA.

Carlson, C. S., Eberle, M. A., Kruglyak, L., & Nickerson, D. A. (2004). Mapping complex disease loci in whole-genome association studies. *Nature*, *429*, 446-452.

Clarke, W. E., Higgins, E. E., Plieske, J., Wieseke, R., Sidebottom, C., Khedikar, Y., ... Li, R. (2016). A high-density SNP genotyping array for Brassica napus and its ancestral diploid species based on optimised selection of single-locus markers in the allotetraploid genome. *Theor. Appl. Genet*, *129*, 1887-1899.

Cordell, H. J., & Clayton, D. G. (2002). A unified stepwise regression procedure for evaluating the relative effects of polymorphisms within a gene using case/control or family data: application to HLA in type 1 diabetes. *Am J Hum Genet*, *70*, 124-141.

Duarte, D. A. S., Newbold, C. J., Detmann, E., Silva, F. F., Freitas, P. H. F., Veroneze, R., & Duarte, M. S. (2019). Genome-wide association studies pathway-based meta-analysis for residual feed intake in beef cattle. *Anim. Genet*, *50*, 150-153.

Freedman, M. L., Monteiro, A. N., & Gagther, S. A. (2011). Principles for the post-GWAS functional characterization of cancer risk loci. *Nat Genet*, *43*, 513-518.

Hardy, J., & Singleton, A. (2009). Genomewide association studies and human disease. *N Engl J Med*, *360*, 1759-1768.

Hindorff, L. A., Gillanders, E. M., & Manolio, T. A. (2011). Genetic architecture of cancer and other complex diseases: lessons learned and future directions. *Carcinogenesis*, *32*, 945-954.

Hindorff, L. A., Sethupathy, P., & Junkins, H. A. (2009). Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc Natl Acad Sci*, *106*, 9362-9367.

Khan, S. U., Yangmiao, J., Liu, S., Zhang, K., Khan, M. H. U., Zhai, Y., ... Zhou, Y. (n.d.).

Khush, G. S. (2001). Challenges for meeting the global food

and nutrient needs in the new millennium. *Proc. Nutr. Soc*, *60*, 15-26.
Kumar, J., Gupta, D. S., Gupta, S., Dubey, S., Gupta, P., Kumar, S., ... Zhou, Y. (2012). Identification of candidate genes of QTLs for seed weight in Brassica napus through comparative mapping among Arabidopsis and Brassica species. *Plant Cell Rep*, *36*, 105-105.
Kuo, C. L., & Feingold, E. (2010). What's the best statistic for a simple test of genetic association in a case-control study? *Genet. Epidemiol*, *34*, 246-253.
Laurie, C. C., Doheny, K. F., Mirel, D. B., Pugh, E. W., Bierut, L. J., Bhangale, T., ... Edenberg, H. J. (2010). Quality control and quality assurance in genotypic data for genome-wide association studies. *Genet. Epidemiol*, *34*, 591-602.
Li, C., Fu, Y., Sun, R., Wang, Y., & Wang, Q. (1083). Single-locus and multi-locus genome-wide association studies in the genetic dissection of fiber quality traits in upland cotton. *Gossypium hirsutum L.). Front. Plant Sci*, *9*.
Li, H., Thrash, A., Tang, J. D., He, L., Yan, J., & Warburton, M. L. (2019). Leveraging GWAS data to identify metabolic pathways and networks involved in maize lipid biosynthesis. *Plant J*, *98*, 853-863.
Lin, B. K., Clyne, M., & Walsh, M. (2006a). Tracking the epidemiology of human genes in the literature: the HuGE Published Literature database. *A J Epidemiol*, *164*, 1-4.
Lin, B. K., Clyne, M., & Walsh, M. (2006b). Tracking the epidemiology of human genes in the literature: the HuGE Published Literature database. *A J Epidemiol*, *164*, 1-4.
Ma, L., Liu, M., Yan, Y., Qing, C., Zhang, X., Zhang, Y., ... Zou, C. (2018). Genetic dissection of maize embryonic callus regenerative capacity using multi-locus genome-wide association studies. *Front. Plant Sci*, *9*, 561-561.
Marchini, J., & Howie, B. (2010). Genotype imputation for genome-wide association studies. *Nat. Rev. Genet*, *11*, 499-511.
Minelli, C., Thompson, J. R., Abrams, K. R., Thakkinstian, A., & Attia, J. (2008). How should we use information about HWE in the meta-analyses of genetic association studies? *Int. J. Epidemiol*, *37*, 136-146.
Morris, G. P., Ramu, P., Deshpande, S. P., Hash, C. T., Shah, T., Upadhyaya, H. D., ... Mitchell, S. E. (2013). Population genomic and genome-wide association studies of agroclimatic traits in sorghum. *Proc. Natl. Acad. Sci*, *110*, 453-458.
Munafo, M. R., & Flint, J. (2004). Meta-analysis of genetic association studies. *Trends Genet*, *20*, 439-444.
Nakaoka, H., & Inoue, I. (2009). Meta-analysis of genetic association studies: methodologies, between-study heterogeneity and winner's curse. *J. Hum. Genet*, *54*, 615-623.
Newell, M., Cook, D., Tinker, N., & Jannink, J. L. (2011). Population structure and linkage disequilibrium in oat (Avena sativa L.): Implications for genome-wide association studies. *Theor. Appl. Genet*, *122*, 623-632.
Pegolo, S., Mach, N., Ramayo-Caldas, Y., Schiavon, S., Bittante, G., & Cecchinato, A. (2018). Integration of GWAS, pathway and network analyses reveals novel mechanistic insights into the synthesis of milk proteins in dairy cows. *Sci. Rep*, *8*.
Rafalski, J. A. (2010). Association genetics in crop improvement. *Curr. Opin. Plant Biol*, *13*, 174-180.
Salanti, G., Sanderson, S., & Higgins, J. P. (2005). Obstacles and opportunities in meta-analysis of genetic association studies. *Genet. Med. Official J. Am. College Med. Genet*, *7*, 13-20.
Schully, S. D., Yu, W., & Mccallum, V. (2011). Cancer GAMAdb: database of cancer genetic associations from meta-analyses and genome-wide association studies. *Eur J Hum Genet*, *19*, 928-930.
Shahid, M., Cai, G., Zu, F., Zhao, Q., Qasim, M. U., Hong, Y., ... Zhou, Y. (1982). Comparative Transcriptome Analysis of Developing Seeds and Silique Wall Reveals Dynamic Transcription Networks for Effective Oil Production in Brassica napus L. *Int. J. Mol. Sci*, *20*.
Song, X. J., & Ashikari, M. J. R. (2008). Toward an optimum return from crop plants. *Rice*, *1*, 135-143.
Spindel, J., Begum, H., Akdemir, D., Collard, B., Redoña, E., Jannink, J., & Mccouch, S. (2016). Genome-wide prediction models that incorporate de novo GWAS are a powerful new tool for tropical rice improvement. *Heredity*, *116*, 395-408.
Tabor, H. K., Risch, N. J., & Myers, R. M. (2002). Candidate-gene approaches for studying complex genetic traits: practical considerations. *Nat Rev Genet*, *3*, 391-397.
Takeda, S., & Matsuoka, M. (2008). Genetic approaches to crop improvement: Responding to environmental and population changes. *Nat. Rev. Genet*, *9*, 444-457.
Tang, J. D., Perkins, A., Williams, W. P., & Warburton, M. L. (2015). Using genome-wide associations to identify metabolic pathways involved in maize aflatoxin accumulation resistance. *BMC Genom*, *16*.
Thakkinstian, A., Mcelduff, P., D'este, C., Duffy, D., & Attia, J. (2005). A method for meta-analysis of molecular association studies. *Stat. Med*, *24*, 1291-1306.
Thomson, M. J. (2014). High-throughput SNP genotyping to accelerate crop improvement. *Plant Breed. Biotechnol*, *2*, 195-212.
Torkamani, A., Topol, E. J., & Schork, N. J. (2008). Pathway analysis of seven common diseases assessed by genome-wide association. *Genomics*, *92*, 265-272.
Wang, K., Li, M., & Bucan, M. (2007). Pathway-Based Approaches for Analysis of Genomewide Association Studies. *Am. J. Hum. Genet*, *81*, 1278-1283.
Warburton, M. L., Tang, J. D., Windham, G. L., Hawkins, L. K., Murray, S. C., Xu, W., ... Williams, W. P. (2015). Genome-Wide Association Mapping of and Aflatoxin Accumulation Resistance in Maize. *Crop Sci*, *55*, 1857-1867.
Warburton, M. L., Womack, E. D., Tang, J. D., Thrash, A., Smith, J. S., Xu, W., ... Williams, W. P. (2018). Genome-Wide Association and Metabolic Pathway Analysis of Corn Earworm Resistance in Maize. *Plant Genome*, *11*.
Weng, L., Macciardi, F., Subramanian, A., Guffanti, G., Potkin, S. G., Yu, Z., & Xie, X. (2011). SNP-based pathway enrichment analysis for genome-wide association studies. *BMC Bioinform*, *12*.
Witte, J. S. (2010). Genome-wide association studies and beyond. *Ann Rev Public Health*, *31*, 20-20.
Xiao, Y., Liu, H., Wu, L., Warburton, M., & Yan, J. (2017). Genome-wide Association Studies in Maize: Praise and Stargaze. *Mol. Plant*, *10*, 359-374.
Xu, Y., Xu, C., & Xu, S. (2017). (Vol. 119). Prediction and association mapping of agronomic traits in maize using multiple omic data.
Yan, J., Warburton, M., & Crouch, J. (2011). Association Mapping for Enhancing Maize (Zea mays L.) Genetic Improvement. *Crop Sci*, *51*, 433-449.
Yang, Y., Shen, Y., Li, S., Ge, X., & Li, Z. (2017). High density linkage map construction and QTL detection for three silique-related traits in Orychophragmus violaceus derived Brassica napus population. Front. *Plant Sci*, *8*, 1512-1512.
Yang, Y., Zhu, K., Li, H., Han, S., Meng, Q., Khan, S. U., ... Zhou, Y. (2018). Precise editing of CLAVATA genes in Brassica napus L. regulates multilocular silique development. *Plant Biotechnol. J*, *16*, 1322-1335.
Yoon, S., Nguyen, H., Yoo, Y. J., Kim, J., Baik, B., Kim, S., ... Nam, D. (2018). (Vol. 46). Efficient pathway enrichment and network analysis of GWAS summary data using GSA-SNP2. Nucleic Acids Re.
Zegeye, H., Rasheed, A., Makdis, F., Badebo, A., & Ogbonnaya, F. C. (2014). Genome-wide association mapping for seedling and adult plant resistance to

stripe rust in synthetic hexaploid wheat. *PLoS ONE*, *9*, 105593-105593.

Zhang, J., Song, Q., Cregan, P. B., Nelson, R. L., Wang, X., Wu, J., & Jiang, G. L. (2015). (Vol. 16). Genome-wide association study for flowering time, maturity dates and plant height in early maturing soybean (Glycine max) germplasm. BMC Genom.

Zhang, X., Zhang, J., He, X., Wang, Y., Ma, X., & Yin, D. (1611). Genome-wide association study of major agronomic traits related to domestication in peanut. *Front. Plant Sci*.

Zhang, Y., Liu, P., Zhang, X., Zheng, Q., Chen, M., Ge, F., ... Liang, T. (2018). Multi-locus genome-wide association study reveals the genetic architecture of stalk lodging resistance-related traits in maize. *Front. Plant Sci*, *9*, 611-611.

Zhou, Q., Zhou, C., Zheng, W., Mason, A. S., Fan, S., Wu, C., ... Huang, Y. (2017). Genome-wide SNP markers based on SLAF-seq uncover breeding traces in rapeseed. *Brassica napus L.). Front. Plant Sci*, *8*, 648-648.

Zintzaras, E., & Lau, J. (2008). Trends in meta-analysis of genetic association studies. *J. Hum. Genet*, *53*, 1-9.