Advances in Bioresearch
Adv. Biores., Vol 9 (6) November 2018: 163-170
©2018 Society of Education, India
Print ISSN 0976-4585; Online ISSN 2277-1573
Journal's URL:http://www.soeagra.com/abr.html
CODEN: ABRDC3
DOI: 10.15515/abr.0976-4585.9.6.163170

Advances in Bioresearch

REVIEW ARTICLE

Significance and Role of Genome wide Association Studies in Relation to Current Scenario of Human Genomic Framework

Hend Alkwai[1], Fahmida Khatoon[2]
[1] College of Medicine ,Pediatric department ,Hail University, KSA
[2]Faculty of Medicine, Department of Biochemistry, College of Medicine, University of Hail, KSA
Email: f.khatoon@uoh.edu.sa

ABSTRACT
The inductive spurious associations between a phenotype for the pathology and a variance subsequently assist in evaluating true relationships between genetic variables. Most of the Genome Wide Association Studies are designed in ways so as to include the single individual from the same population in order to minimize errors and biasness caused by population stratification. The phenotype traits which are useful tools for anthropometric measurements for forensic purposes include the color of eyes and hair, freckling, hair structure; elicit response to better tastes, urinary and creatinine excretion and metabolic rates. This search strategy was based on evidence available within a range of past ten years and was basically primary researches. The methods for research for this particular study is to demonstrate a review of relevant literature. The effects of epitasis can be favorable in producing no random phenotype and can be harmful in various ways. In the current era, epitasis is now becoming widely accepted as a tool in determining the relations between phenotype and genotype. It has also been shown that simple Mendelian traits like identification of metabolic disorders in body secretions samples can be made easy to detect when interactions between other genes and the mutant variations that serve to produce phenotypes for a particular genetic expression are assessed. The correct identification of a particular genotype in varying phenotypical expressions can be convoluted that can be one of the reasons epitasis can be challenging in giving success. Several methods have been developed in order to study the epitasis with regards to Genome association studies, including logistic regression methods, penalized methods and data mining approaches such as Random Forests, Recursive Elimination of Feature-F, Grammatical Evolution Neural Networks, Multifactor Dimensionality Reduction and Combinational Partitioning. Stratification acts like a tool for confounding subgroups for various populations within the sample that have different occurrences for a particular trait. The relevant rationale for assessing the human behavior helps to legalize and authenticate the data selectively collected from the secondary sources. The aim of this study is to explore the application of Genome Wide Association Studies in accordance with Biochemical investigation of human population.
Key Words: *Human Genome Project, Genome Wide Association Studies, Genome,Gene, Research Design*

Received 21.07.2018 Revised 19.09.2018 Accepted 21.10.2018

How to cite this article:
Hend Alkwai, Fahmida Khatoon. Significance and Role of Genome wide Association Studies in Relation to Current Scenario of Human Genomic Framework. Adv. Biores., Vol 9 [6] November 2018.163-170.

INTRODUCTION

The Human genome is now capable to incite in depth and precise data information and allows access into processes that extract detailed sequences for genomics in order to analyze and answer not only theoretical questions but also identifies the practically feasible genetic characteristics [1].

Genomic sequences aid in appreciating the origination and nature of the particular mutations in genetic data and thus, incorporate functional genetics, biological sciences and bioinformatics to extract the cause of a variation in genetic framework. In the evolutionary field of forensics, the combination of information and the understanding of genetic architecture lead to dissect the unidentified entities and translate for making connections in solving ongoing genotype and phenotype puzzles of unknown nature [1].

The last five years in the field of genomic studies have been an era of insurgent of discoveries for epidemiological studies associated with the genetic and genomic technological advancements. High-Throughput genotyping and sequencing centers are containing torrents of data and information related to surveying genetic variants across the genome for association of complex diseases , human phonotypical

quantitative traits like hair colour and eye colour and quantitative traits [2]. The Genome Wide Association Studies have been the work force for these efforts and now is considered as a standard technique for disease gene mapping. Thus, the use of already known Mandelian relationships between pairs of individuals included in the study sample is for purpose to infer the tendency to make mistakes which is not exclusively possible to elicit successful outcomes. Instead, misidentification of samples can be detected by following comparisons between: (a). Sex characteristics associated with X and Y chromosomes data for genome that have not been established to identify/ recognize sex and/or sex characters, (b) The inferred data for constructing ancestry details based on principle component analysis with self proclaimed ancestry information [3, 4]. Most of the Genome Wide Association Studies are designed in ways so as to include the single individual from the same population in order to minimize errors and their biased nature resulted due to population stratification [5].

The utilization of Genome Wide Association Studies for detecting many diseases of forensic interest and traits for sex distributions are due to sex associated variations in prevalence [3] The sex and ancestry related data and genetic information are important components for Quantity Check (QC) in GWAS and can be implemented in practice to remove misidentifications of samples and stratifications in population samples. The approach to identify the misidentification in processes where strong associations are detected between genotype and phenotype characteristics can be inferred if the observed phenotype is expected to be caused by the observed genotype for every subject within the study [5]. This observed phenotype if detected to be too extreme in accordance with the genotype then the person is flagged as a possible subject for misidentification [3]. The genetic framework for the European derived genomes further helped in designing the first episode for GWA studies [6].

The study of genomic and associated genetic markers requires the step wise following of techniques like some of these aspects have been examined on a smaller scale for individual diseases such as type 2 diabetes , inflammatory bowel disease , and cancer, a comprehensive genome-wide analysis across all GWAS published to date has not been conducted. a set of information about genotype and phenotype relationship need to be established and identified [5]. Secondly, phenotypes need to be ascertained at additional cost depend on experimental study and time duration before the initiation of study, possibly for reasons of extracting usual phenotype data during the collection method. The modeling for the mixture sample that is under consideration for estimating phenotype and genotype relationship needs to be optimized to deliver most suitably applicable information. Finally, the combination of all the information assessed from the relationships should be higher enough to give better sensitivity and specificity. It can also be possible to assess Single Nucleotide Polymorphisms (SNPs) in determining genotype for each sample, before the execution of costly Genome Associated Studies or Genome Sequencing into practice [5]. This study is to critically evaluate the Genome Wide Association Studies in accordance with population based forensic investigations and to examine the population stratification with respect to Genome Wide Association Studies and its related challenges in illustrating error-free determination.It is also evaluating the challenges in developing statistical techniques and implementation of computational efficacy in judging genetic variance and their interactions with genomic expression during 'logistic regression' (LR), 'multi-factorial dimensionality reduction' (MDR- polygenic mode of inheritance dependent on a number of genes at different loci), ' random forest' (RF) and 'evaporating cooling (EC) methods (mention the basic concepts underlying these interactions to illustrate their applications in this study).

LITERATURE REVIEW

These aspects of the study are concentrated on the evaluation and confirm the applications of the research related to genome wide association studies. This step enhances efforts to establish the relationship between genotype and phenotype observation with genome wide association studies. Give some related references to validate this fact.

With the embankment of the technologies in genetic sequencing and phenotyping, the execution of genome wide studies to evaluate relationships between phenotypical and genotypic variants can discover reliable inferences. Genome Wide Association Studies typically utilizes the high tech and thorough input platform to gather large quantity of valid data from millions of genetic variations [7].

This alternative approach is mostly applied to correlate and elaborate the genetic causes of the complex trait/s. These studies focus on the establishing the correlation between genetic markers and relevant traits among unrelated population samples [8]. Few examples of entrepreneurial work situations permeated by emotions. These studies also play a pivotal role in shaping human cognition, motivation and behavior [9] and, consequently, they influence entrepreneurial performance (EP) [10]. The reasons for ancestral mutation of genes that have taken place hundreds of years ago. The lineages of all such

descendents of the gene carrier of the casual variants are considered to be susceptible of the polygenic mutations. Cardon *et al* suggested that if such casual variant/s is investigated with reference to perspective of genome wide studies perspectives. Individuals carrying the derived allele are likely to disperse it in different phenotypes as compared to those who carry ancestral alleles [8] give the reported examples by this research group. Global studies [11] reveal that genome wide population structure on various levels can be assessed using statistical approaches. Individuals in different continents get differentiated in genetic framework through genome-wide SNP data but however, some level of over-lapping do exists between continental regions [11]Key characteristics of the observed associations and the trait/disease associated SNPs (TASs) have been examined on a smaller scale for individual diseases such as type 2 diabetes and IBS (inflammatory bowel disease) a comprehensive genome-wide analysis across all GWAS published to date has not been conducted. For example, there is a strong relationship between geographic correlation and genetics similarity such as populations from Oceania and the Americas—the global distribution of CNVs largely accords with population structure analyses for SNP data sets of similar size. Our results produce new inferences about inter-population variation, support the utility of CNVs in human population-genetic research, and serve as a genomic resource for human-genetic studies in diverse worldwide populations[11] which determines accurate determinations for genetic variability [11]. Such data obtained by the researchers by adopting this technique and the genes that are responsible for predisposing a person at risk of developing disease or a trait can be an important in terms of forensic significance [12].

Selective genotyping is an application used to identify different ethnic groups like African and Europeans or a mix population and can be beneficial in identifying specific traits in a homogenous ancestry distribution. Similarly, by comparing relative allele frequencies between different phenotypic groups, Genome Wide Studies can detect SNPs that are associated specifically to a disease process or an ethnic trait. GWAS are preferred by many researchers because they provide more accurate localization of casual genes and also they provide unbiased detection of whole genome for genetic association [12]. The analysis for identifying the linkages in familial expressions allows the execution of frequency of polymorphism among 71 proteins, and is studied in European populations[12], and the inferences deduced state that 0.28 base changes per 1000 base pairs suggesting that DNA Sequence Polymorphism depends on the size of the average gene. The visible Electromorphs represents about 1/3 rd of Nucleic acid sequence changes and out these third base changes are not reflected in corresponding amino acids. This depicts that the DNA sequence polymorphism is about 0.001 for each of the base pairs among various protein coding sequences [13].

COMPLEXITY OF GENOMIC FRAMEWORK

There are numerous forensic cases in which DNA-based inferences of biogeographic ancestry information are vital that facilitate police investigations to find probable and/or suspect individuals or unidentified victims under consideration;[14]. However, it is important to acknowledge that when and how biogeographic ancestry is applied in answering forensic queries and where DNA testing techniques can be introduced to determine the level of ethnic disparity using specific DNA marker sets in association studies on neuroticism, acculturation and the Cortisol Awakening It is evident to keep in mind that the genetic diversity allows the involvements of various appearance traits likely to be error prone as no phenotype is restricted to a certain geographic region. In similar terms, appearances of the unidentified individuals can be known by utilizing markers from genes incorporated into data sets for genomic studies that are functionally active and strongly associated with the appearances and skin color of the subject under consideration, [14].

The genetic effect of Single Nucleotide Polymorphisms (SNPs) on the phenotype of an individual depend on the number of contributing SNPs and non-genetic influences like environmental effects, in determining accurately an identity of a person. Phenotype studies illustrate that eye color is the most successfully predictable phenotype essential in accurately identifying basic physical factors like black and white eye colourin an individual [14]. The challenges undertaken in applying DNA prediction into categorizing appearance traits and eye color is its expected variability in conceptual understanding of trait information. For example, people assign same eye color to various color categories and therefore look different from others using an eye color provided by DNA predictions. In order to minimize this problem, studies investigated about the genetic basis of variation in eye color utilizing SNP data analysis [14].

Two of the SNPs are believed to be highly specific to hair color while the third relates to reflect biogeographic ancestry rather than emphasizing on the hair color [14]. Another trait that expresses a physically visible characteristic can be useful in successfully predicting appearance of an individual specially age. Two DNA- based approaches for age predictions are based on mtDNA (mitochondrial DNA)

deletions and age-dependent telomere shrinkage have been suggested for forensic assistance but further research is needed to diversify their effects[14] Genome wide studies, on age related features for analyzing DNA methylation pattern may provide extensive benefits for establishing suitable age-predictive biomarkers [14]. By targeting the Genome data that is relevant and informative can be obtained and is generally applicable. This is because of the fact that the specific information contained within the Genome is generally correlated with all the normal cells in the human body as all cells contain similar DNA sequences, same mutations and polymorphisms.

PRINCIPLE COMPONENT AND ASSOCIATION STUDY ANALYSIS

The genomic studies has also been carried within the European population for achieving better inferences in observing a common group of individuals for genomic variance in order to reduce the expenses and efforts for collecting samples and genotyping analysis [15]. Another reason for European population based studies was the availability of large and homogeneous sample for populations, such as in cohorts for Finland and Sardina derived population samples [15].These cohorts are believed to contain extensive collaborations and the prior genetic and phonotypical information to help in analysis and interpretation of Genome Wide Association Studies [15].

MATERIAL AND METHODS

The main objective of the literature review is to draw conclusions for Genomics and its relevance in applying techniques in forensic investigations. This chapter will highlight the definition of literature review and its significance as a method for this study. Then, it will outline the way to carry out this study including the searching strategy process, limitations, and the qualitative design employed in this study. Finally, it will conclude with ethical issues along with the inclusion and exclusion criteria.

For establishing any search of literature, it is important to comprehend any research and its role in informing the clinical practice as well as questioning [16]. This systemic process was defined by the researcher as utilizing the definite methods for cracking the problem and answering to the research question [16].

Search Strategy

For research evaluation on the molecular DNA and forensics, the inclusive review of most of the studies available was important. The databases used and accessed for this purpose included: ProQuest, PubMed, Cochrane Library, Science Direct and CINAHL. For the data bases searches following keywords were utilized including Molecular genetics, Genome-wide studies, Polymerase Chain Reaction, DNA Phenotyping, DNA-amplification and sources for DNA-extraction. The bibliographic references of studies selected for the literature review were also searched from these databases. All studies meeting up with the explicit inclusion & exclusion criteria for each literature review section were retained at the end of the search.

Qualitative Design

The qualitative research design is the most appropriate research design to examine the human behavior and it assists to legalize and authenticate the data which is selectively collected from the secondary sources. It helps in refining the research and additionally it adds a primary hand worth to it. For many of the researchers the secondary data is crucial in the health care and medicine field as it suggest the previous researches [17], stated that a qualitative research comprise of using exclusive steps of strategies and analysis of inquiry with researchers interpreting what they hear, see, and understand. The method used in this study is qualitative. As compared to quantitative research, Qualitative research is more subjective, and it is based on unlikely methods of gathering information. This research is more or less based on the literature review and the conclusions are drawn on the basis of actual resources.

Critical Analysis

The critical analysis is central process of any research study. It involves critical thinking which applies logical and rational thinking while deconstructing the text. It is a complex, intellectual activity involving analysis and critical comment on the material formerly gathered. [18] defined the critical thinking as a set of awareness which interrelates with the research question under study which is expected to be critically analyzed. It is an ability to answer and ask the critical question at a specific time and it desires to use the critical questions actively).

Inclusion criteria

The search of literature only found out the material published in English and for this reason it created no issues to reject any possible study for the reason of the translating complexities of the papers

Exclusion Criteria

Exclusion criteria included the articles not published as full manuscripts or not in the peer review literature. The studies including controversial ethical or legislative materials are not included. The studies which were conference papers or studies in progress, or government reports will be excluded. The searches will be restricted to retrieve the literature available in English

Limitation

In this review there was a likelihood of publication bias. By rejecting those studies with negative outcomes and were unpublished it was probable to over rate the effects of techniques used, nevertheless, for the published literature the comprehensive search for potentially appropriate articles was undertaken by means of a strategy for systemic review for the purpose of avoiding bias.

Ethical Considerations

This study lies on primarily on the original and primary sources as it is known that the unethical, fraudulent and dishonest researchers can circumvent the scientific method. It is important to properly cite the person whose work is being used. This is because the person who has conducted an actual research may have served a lot of effort and time to extract the outcomes of the study and hence it is one's responsibility to give credit legally to the individual who has conducted that particular work.

RESULT and DISCUSSION

This section will include the thematic analysis and critique of the selected articles. This chapter will also discuss the selected researches in detail and their findings, based on the findings of the reviewed literature.

Critical Analysis

In 1999, Pritchard and Rosenberg proposed a technique which was based on the concept of differences in the genetic framework between cases and their controls that could be tested. The variations in the genetic details of cases and controls can be measured by utilizing markers which are unrelated with the mutations detected [19]. This method has been successful in evaluating the specific structural characteristics in the population; however, it is unable to correct the detected abnormalities in the provided data information about genetic mutations. In the same year, Devlin and Roeder presented Genomic Control method which was also based on the concept of biological markers for genetic detections [20]. Genomic Control also detects significant stratification but effectively compensates for it to be corrected. Their technique was based on the assumption of population structure with respect to case and control data that can be alter by implying multiplicative factorization in proportion to structure of population in unidentified genetic markers.

The inflation factor can later be incorporated into analysis for detecting associations and for correcting stratifications. One benefit for this method is that it can easily be implemented for analyzing the various categories for DNA data information and sequestering. Further, it also allows the implication of more accuracy in results as the number of genetic markers used to calculate the inflation factor, which makes it an efficacious technique to be followed in Genome Wide Association Studies [19]. But on the downhill, it cannot on consistent basis detect and correct for relations with the family structure when they are responsible for causing stratification and can lead to corresponding loss of power when markers are strictly differing across the ancestry of subsequent population framework [19].

The serious issue in Genomic Association studies is studying the influence of genomic ancestry on the design for Genome Wide Association Studies. The confounding effects of genetic ancestry in association studies play a significant effect in understanding the population based characteristics that differentiates one subset from the other regarding genotype and phenotype information [21]. Most of the Genome Wide Association Studies are been conducted on European derived populations [21]. In the initiation stages of GWAS, extensive researches were under taken on Europeans mostly due to logistical and research design reasons. In logistical terms, the financial burden for the conduction of GWAS lead to many early studies to be conducted on same general populations with extensive variance in physical characteristics [21].

Epitasis and its Relevance with Genome Wide Association Studies

The first definition of epitasis was provided by William Bateson in 1909, and was described in terms of deviation from Mendelian inheritance. Currently, epitasis is commonly defined as interaction among genes at various loci where the impact of allele at one locus is inhibited by an allele on the other locus [1]The effects of epitasis can be favorable in producing no random phenotype and can be harmful in various ways. In the current era, epitasis is now becoming widely accepted as a tool in determining the relations between phenotype and genotype [1].Biologists and forensic experts study gene to gene interactions by incorporating epitasis as physical and biochemical reactions which occur between genes and regulatory networks. Biological component of epitasis can be expressed in several ways; by interactions between factors responsible for transcriptions and by enzymes involved in metabolic pathways [1]. For example,

blood borne or metabolic abnormalities like Maple Syrup Urine Disease in forensic sampling can aid to separate the individual profile from the majority of the population, thus identity of the unknown can be made possible. Similarly, in Sickle Cell Disease there is a biomolecular interactions between globulin proteins, hence helps in differentiating sick from healthy individuals [1].

Epidemiology and Significance of Epitasis

Epitasis has been increasingly significant factor in understanding the genotype for complex abnormality [22]. These inferences depict that epitasis can be more prevalent not only in failed replications for Genome Wide Studies but also have main effects in detecting heritability in samples with unique genetic characteristics [23]. It has also been shown that simple Mendelian traits like identification of metabolic disorders in body secretions samples, Cystic Fibrosis in blood samples etc. can be made easy to detect when interactions between other genes and the mutant variations that serve to produce phenotypes for a particular genetic expression are assessed.

Challenges in Identifying Epitasis

The correct identification of a particular genotype in varying phenotypical expressions can be convoluted that can be one of the reasons epitasis can be challenging in giving success. This is mostly applicable on human subjects which limit the benefits of methods such as genetic manipulation on organisms that serves as models for humans, hence making the execution of techniques even more difficult. In order to overcome this problem, characterization of epitasis on genome wide scale allows numerous biological, computational as well as statistical challenges not only in analyzing genomic data but also to prove functional validity of such inferences. This method presents even more challenges when human populations are assessed as researchers can not benefit from all tools that could be tested on model organisms [23].

The conventional methods for applying association theory comprised of using linear models, while these models work better in finding main effects, but use to fall short in finding interactive effects [23]. The detailed and diverse data framework that needs to be analysed also presents analytical challenge and demand the development of better statistical techniques to cater non-handle components for these interactions which are influenced on by environmental factors. This technique requires the filtration and priotization of biological relevance along with maintaining computational efficacy in evaluating genetic variance and their interactions with genetic environment [24]. Several methods have been developed in order to study the epitasis with regards to Genome association studies, including logistic regression methods, penalized methods and data mining approaches such as Random Forests, Recursive Elimination of Feature-F, Grammatical Evolution Neural Networks, Multifactor Dimensionality Reduction and Combinational Partitioning [25, 26, 1].

Logistic Regression Method in Determining Epitasis:

Regression based techniques have been deployed for analyzing epitasis interactions in genomic association studies. They use the statistical approaches to determine the relationships between linear values for predictor variables for example, genetic variants and the possible phenotypical expression. Logistic regression is also been in use when the outcome for the population is binary in nature that is either case or control. This method can also be applied for variable selections for example, in step wise regression, where in variables which elicit main effects are selected first followed by interactive factors that lie between them. The utilization of regression method, however, eliminates the possibility of finding interactions between variations with no profound effects. Logistic regression also has limitations when data are greatly multidimensional as is common when information is analyzed for association studies thus, can lead to higher rates for false positive results. The availability of software packages that help in performing statistical correlation for logistic regression on genomic population samples, one of such tools is named as Genome Analysis Tool (PLINK), which provides the option of logistic regression in association analysis[27].

CONCLUSION

Genome Wide Association Studies has evoked several other successful studies searching for answers for genetic variants related to complex hereditary information. [28,29]However, other studies also illustrated that results cannot be replicated in other independent sampling and these scenarios demand deeper investigations in order to cater the replication failures. The popularity and the utilization of Genome Wide Association Studies have increased over the years, along with the challenges like identification of stratifications in a population has been made feasible for forensic investigations

In this study, the analysis for Genomics yield conclusive evidences for cases when common traits have been detected in genomic samples like detection of familial tendency of pathologies, mutation at fixed loci and phenotypical similarities.[30] For example, some of the phenotypic appearances are specific to an ethnic

group, for example the hair color, skin color and morphological features. It has now been suggested that the prevalence of the minor allele frequencies have higher penetrance and may be responsible for familial diseases. [31].

In 2005, Genome Wide Association Studies has evoked several other successful studies searching for answers for genetic variants related to complexed hereditary information.[28,29] However, other studies also illustrated that results cannot be replicated in other independent sampling and these scenarios demand deeper investigations in order to cater the replication failures. The popularity and the utilization of Genome Wide Association Studies have increased over the years, along with the challenges like identification of stratifications in a population has been made feasible for forensic investigations.

Future implications for minimizing the failures in replicative studies pertaining to Genome Association Studies can led to development of methods to define and evaluate the genomics and to prevent errors. The robust application of Genomics and associated studies can also have the confounding impacts as factors that can improve the reliability of the inferences extracted. On the other hand, the novelty of the technique has limitations to the amount of genetic variations detailed by the risk variants found to be related to the genetic factors under investigation.

The Genome Wide Association Studies has also suggested the limited approach of this method as it may not serve as a powerful tool to discover the complete genetic predisposition of the mutations unique to a familial trait, or even can predict a disease risk which has not been noted due to insufficient knowledge about genetic framework. This limitation can be attributed to the held assumption that the effects of main factors that act independently but chiefly contribute to the genotype for causation of mutations. However, this notion has now been challenged as more empirical studies have depicted that epistatic impact can occur despite of the presence of any main effect and thus, can essentially effect the phenotype, as has been proved solvable in Genome Wide Association Studies.

REFERENCES

1. Moore,J. H. and Williams,S. M. (2005). Traversing the Conceptual Divide between Biological and Statistical Epistasis: Systems Biology and a More Modern Synthesis. *Bioessays 27 (6)* 637-646.
2. Park Lee,S.Y., Chung,Y., Elston,R. C., Kim,Y., and Park,T. (2007). Log-Linear Model-Base Multifactor Dimensionality Reduction Method to Detect gene–gene Interactions. Bioinformatics 23 (19) 2589-2595.
3. Laurie, C., K. Doheny, D. Mirel, E. Pugh, L. Bierut, T. Bhangale, F. Boehm, N. Caporaso, M.Cornelis, H. Edenberg, et al. (2010): "Quality control and quality assurance in genotypic data for genome-wide association studies," Genetic Epidemiology, 34, 591–602.
4. Anderson, C., F. Patterson F, G. Clarke, L. Cardon, A. Morris, and K. Zondervan (2010): "Data quality control in genetic case-control association studies," Nature Protocols, 5, 1564–1573.
5. Hindorff, L., H. Junkins, P. Hall, J. Mehta, and T. Manolio (2009): "A catalog of published genome-wide association studies. Available at: http://www.genome.gov/gwastudies
6. Jakkula, E., etal., (2008). The genome-wide patterns of variation expose significant substructure in a founder population. Am J Hum Genet. 83 (6): p.787-94.
7. Manolio,T. A., Brooks,L. D., and Collins,F. S. (2008). A HapMap Harvest of Insights into the Genetics of Common Disease. J.Clin.Invest. 118 (5) 1590-1605.
8. Cardon, L.R. and Bell, J.I., 2001 Association study designs for complex diseases. Nat Rev Genet; 2(2):p.91-9.
9. Judge, T.A. and Larsen, R. (2001). Dispositional affect and job satisfaction: a review and theoretical extension. Organizational Behavior and Human Decision Processes, 86, pp. 67-98.
10. Baron R, Rowe W, Schaefer C, Zhang J. Buetow K (2008) Needles in the haystack: identifying individuals present in pooled genomic data. PLoS Genet 5(9): e1000668. doi:10.1371/journal.pgen.1000668
11. Jakobsson, M et al. (2008). Genotype, haplotype and copy number variation in worldwide human populations. Nature 451, 998-1003.
12. Lueders, T., & Friedrich, M. W. (2003). Evaluation of PCR amplification bias by terminal restriction fragment length polymorphism analysis of small-subunit rRNA and mcrA genes by using defined template mixtures of methanogenic pure cultures and soil DNA extracts. Applied and Environmental Microbiology, 69(1), 320-326.
13. Osborn, A. M., E. R. B. Moore, and K. N. Timmis. 2000. An evaluation of terminal restriction fragment length polymorphism (T-RFLP) analysis for the study of microbial community structure and dynamics. Environ. Microbiol. 2: 39–50
14. Mangold E. et al. 2010. Genome wide association study identifies two susceptibility loci for nonsyndromic cleft lip with or without cleft plate. Nature Genet. 42; 24-26.
15. Rosenberg, N.A., et al., 2010. Genome-wide association studies in diverse populations. Nat Rev Genet. 11(5): p.356-66.
16. Parahoo K (2006) Nursing Research: Principles, Process and Issues,2nd edn. Palgrave Macmillan, Basingstoke
17. Creswell, J.W. (2009) Research Design Qualitative, Quantitative, and Mixed Methods Approaches Third Edition, SAGE Publications, Inc, pp. 296 retrieved from http://www.sagepub.com/books/Book232401.
18. Browne, M & Keeley, S (2001), Asking the right questions: a guide to critical thinking, 6th edn, Prentice-Hall, Upper Saddle River, N.J.

19. Weir,B. S., Anderson,A. D., and Hepler,A. B. (2006). Genetic Relatedness Analysis: Modern Data and New Challenges. Nat.Rev.Genet. 7 (10) 771-780.
20. Alexander,D. H., Novembre,J., and Lange,K. (2009). Fast Model-Based Estimation of Ancestry in Unrelated Individuals, Genome Res. 19 (9): 1655-1664.
21. Need, A.C. and Goldstein, D.B., 2009.Next generation disparities in human genomics: concerns and remedies. Trends Genet. 25(11): p .489-94
22. Shao,H., Burrage,L. C., Sinasac,D. S., Hill,A. E., Ernest,S. R., O'Brien,W., Courtland,H., Jepsen,K. J., Kirby,A., Kulbokas,E. J., Daly,M. J.,Broman,K. W., Lander,E. S., and Nadeau,J. H. (2008). Genetic Architecture of Complex Traits: Large Phenotypic Effects and Pervasive Epistasis. Proc.Nat.Acad.Sci. 105 (50) 19910-19914.
23. Greene,C. S., Penrod,N. M., Williams,S. M., and Moore,J. H. (2009). Failure to Replicate a Genetic Association may Provide Important Clues about Genetic Architecture. PLoS ONE 4 (6) e5639.
24. Lewontin,R. C. (2006). Commentary: Statistical Analysis Or Biological Analysis as Tools for Understanding Biological Causes. International Journal of Epidemiology 35 (3) 536-537.
25. Motsinger-Reif,A. A., Dudek,S. M., Hahn,L. W., and Ritchie,M. D. (2008). Comparison of Approaches for Machine-Learning Optimization of Neural Networks for Detecting GeneGene Interactions in Genetic Epidemiology. Genet.Epidemiol. 32 (4) 325-340.
26. Wang,Y., Liu,G., Feng,M., and Wong,L. (2011). An Empirical Comparison of several Recent Epistatic Interaction Detection Methods. Bioinformatics 27 (21) 2936-2943
27. Purcell,S., Neale,B., Todd-Brown,K., Thomas,L., Ferreira,M. A. R., Bender,D., Maller,J., Sklar,P., de Bakker,P. I. W., Daly,M. J., and Sham,P. C. (2007). PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. 81 (3) 559-575.
28. Scott CM, et al. (2007) ADD66, a gene involved in the endoplasmic reticulum-associated degradation of alpha-1-antitrypsin-Z in yeast, facilitates proteasome activity and assembly. Mol Biol Cell 18(10):3776-87
29. Hunter DJ, Kraft P, Jacobs KB, Cox et all A genome-wide association study identifies alleles in FGFR2 associated with risk of sporadic postmenopausal breast cancer Eur J Clin Nutr. 1998 Mar;52(3):186-92.
30. Fernando,M. M. A., Stevens,C. R., Walsh,E. C., De Jager,P. L., Goyette,P., Plenge,R. M., Vyse,T. J., and Rioux,J. D. (2008). Defining the Role of the MHC in Autoimmunity: A Review and Pooled Analysis. PLoS Genet 4 (4) e1000024.
31. Schork,N. J., Murray,S. S., Frazer,K. A., and Topol,E. J. (2009). Common Vs. Rare Allele Hypotheses for Complex Diseases. Curr.Opin.Genet.Dev. 19 (3) 212-219.