NIH Public Access
Author Manuscript
Am J Ophthalmol. Author manuscript; available in PMC 2013 July 15.

Published in final edited form as:
Am J Ophthalmol. 2010 October ; 150(4): 450–452.e2. doi:10.1016/j.ajo.2010.06.012.

GENETIC EPIDEMIOLOGY: SUCCESSES AND CHALLENGES OF GENOME-WIDE ASSOCIATION STUDIES USING THE EXAMPLE OF AGE-RELATED MACULAR DEGENERATION

Inga Peter, PhD[1] and **Johanna M. Seddon, MD, ScM**[2]

[1]Department of Genetics and Genomic Sciences, Mount Sinai School of Medicine, New York, NY

[2]Ophthalmic Epidemiology and Genetics Service, Tufts Medical Center, and Department of Ophthalmology, Tufts University School of Medicine, Boston, MA

With increasing evidence of importance of genetic factors and gene-environment interactions in the etiology of common eye-related disorders, identification of genetic variation associated with a disease risk may help provide insight into the mechanism of disease pathogenesis and reveal novel targets for preventive and therapeutic interventions. Family-based studies have identified genomic regions associated with highly penetrant genes related to rare familial forms of eye diseases. However, late onset complex phenotypes, such as age-related macular degeneration (AMD), appear to be polygenic, with the involvement of multiple genes, with varying levels of effect.

It has been estimated that about 90% of sequence variants in humans are differences in single bases of DNA, called single nucleotide polymorphisms (SNPs). According to dbSNP, (available from: http://www.ncbi.nlm.nih.gov/projects/SNP) more than 14 million uniquely mapped SNPs have been identified and assembled into a genome-wide database. The HapMap effort (available from: http://hapmap.org) allowed reducing the number of SNPs required for the examination of the entire genome to roughly a million *representative* SNPs, also called tagging SNPs, making genome scan approaches finding regions with genes that affect diseases more efficient and less costly.

During the past several years, genome-wide association studies (GWAS) have been designed to assess association between traits and large numbers of DNA sequence variants distributed across the genome and detect novel disease-associated pathways using an unbiased hypothesis-free approach. By design, GWAS are aimed at identifying *common* SNPs with allele frequency of >5% that may only modestly increase the disease risk consistent with the common variant-common disease hypothesis of disease pathogenesis.

© 2010 Elsevier Inc. All rights reserved.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

The coauthors have seen and agree with each of the changes made to this manuscript in the revision and to the way their names are listed.

Financial Disclosures: None (IP); Pfizer and Genentech (JMS) Tufts Medical Center has filed a patent application for some materials related to this work (JMS)

Contributions to Authors in each of these areas: design and conduct of the study (IP, JMS); collection and interpretation of the data (IP, JMS); and preparation, review, or approval of the manuscript (IP, JMS).

Statement about Conformity with Author Information: This is an editorial that did not include any original data. IRB approval is irrelevant.

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Currently, 561 GWAS have been reported and catalogued at http://www.genome.gov/gwastudies. More than 250 genetic loci in which common variants were reproducibly associated with a number of polygenic traits have been identified during the last two years only (reviewed in 1). However, despite significant successes achieved by GWAS, their inconsistency is a generally recognized limitation.

Challenges for Genome-Wide Association Studies

Population stratification

The most common factor that can bias the results of genetic association studies refers to differences in frequency of candidate genetic variants in the study cohort due to sub-populations with different ancestry. In the presence of population stratification, the significant associations could be exclusively due to the different prevalence of genetic variants in various ethnic and racial sub-groups, whereas the real disease causing locus might be missed. This issue can be addressed by matching cases and controls for individual's origin or including ancestry informative markers in the analysis.

Multiple hypothesis testing

In traditional statistics, with just one test performed at the p-value level of 0.05, there is a 5% chance of incorrectly rejecting the null hypothesis. However, with 100 independent tests, the expected number of such rejections is 5 (5% of 100). The number of false positives may increase to 50,000 when a million SNP chip array is considered. In order to minimize the false positive error rate associated with carrying out multiple statistical comparisons, a recommended genome-wide statistical significance level is $p < 5\times10^{-8}$. This level of significance can be achieved when the genetic effect is considerable (e.g., large odds ratio, OR) or when a sample size for the study is large.

Sample size/Power

It has been reported that most of the common variants found in the recent GWAS are associated with ORs of only between about 1.2 and 1.5, with the mean OR of 1.36.2 This effect size translates to sample sizes of about 4000 cases and 4000 controls required to detect genetic associations with 80% statistical power if a minor allele frequency (MAF) is 10% and almost 7400 cases and 7400 controls if the MAF is 5%. Therefore, small studies are often underpowered to detect small to modest effects and may result in misleading conclusions.

Functional relevance

Significant associations are often detected with genetic variants which are located in non-coding genetic regions as well as outside of known genes, which create a challenge in identifying a causal pathway. While some of the detected associations can be false positives, SNPs not resulting in amino acid change can still affect alternative splicing, gene expression or protein folding. Also, SNPs found to be associated with the phenotype of interest may not be causal but surrogates for causal SNPs, as the result of correlation due to linkage disequilibrium (LD). This has an implication in replication findings, if the replication populations have different LD structures in the regions of interest.

Rare variants

SNPs in the coding or regulatory regions of genes are likely to cause functional differences. Nevertheless, many functional variations can be missed due to their rare frequencies which are not well tagged by commercial genotyping panels. Identification of rare variants with

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

impact on disease susceptibility can be accomplished by follow-up re-sequencing of the refined gene regions.

Replication

Given the high rate of false positive findings due to phenotype heterogeneity, population stratification, multiple hypothesis testing and other factors, replication of association findings in an independent cohort followed by functional studies is a gold standard in the GWAS field.

Genome-Wide Association Studies of Age-Related Macular Degeneration

One of the earliest GWAS examined AMD in a screen of 96 cases and 50 controls genotyped for 116,204 SNPs in the Age-Related Eye Disease Study.3 Despite the small cohort size, an intronic and common variant in the complement factor H gene (*CFH*) was found to be strongly associated with AMD. In individuals carrying two copies of the risk allele, the likelihood of AMD was increased 7.4-fold.3 Re-sequencing revealed a polymorphism in LD with the risk allele representing an acid change in a region of *CFH* that binds heparin and C-reactive protein. Other groups discovered similar findings 4 and later replicated findings and reported new *CFH* variants.5, 6 Additional AMD-susceptibility loci were found in the *ARMS2/HTRA1* region 7 as well as other complement pathway genes. 8–11

Recent GWAS with more statistical power, examining much larger numbers of individuals as well as replication cohorts, have identified additional pathways associated with the disease. Hepatic lipase-C, *LIPC*, was discovered to effect AMD susceptibility.12 The allele which raises high density lipoprotein (HDL) cholesterol reduced risk of AMD. This association was corroborated by another GWAS.13 Furthermore, both GWAS implicated a number of other genes in the HDL pathway. Those genes did not all have the same direction of effect, and therefore mechanisms other than a direct effect of serum HDL metabolism could be involved in the pathogenesis of AMD. Also, a susceptibility locus near the *TIPM3* gene, a metalloproteinase involved in degradation of the extracellular matrix, previously discovered to be associated with Sorsby's fundus dystrophy, an early-onset maculopathy, was implicated12 and corroborated13 in these recent GWAS.

While the GWAS design has borne much fruit for AMD research, many diseases and traits including other eye diseases, such as glaucoma and diabetic retinopathy, to date have not been as successful. It has been suggested that rare variants and structural variation may help find “missing heritability” – unexplained portion of phenotypic variance attributable to genetic factors. 14 The targeted re-sequencing of pathways detected by genome scans as well as whole-genome sequencing in people with extreme phenotypes from diverse ethnic groups will help detect new functional variants in novel and previously identified loci and define their associations with a disease. Large studies will help evaluate the role of gene-gene, gene-environment and gene-treatment interactions and examine their contributions to disease risk, progression, as well as prevention and treatment.15

Acknowledgments

Funding/Support (including none): Grant RO1-EY11309 from the National Institutes of Health, Bethesda, MD; Massachusetts Lions Eye Research Fund, Inc.; the Macular Degeneration Research Fund of the Ophthalmic Epidemiology and Genetics Service, New England Eye Center, Tufts Medical Center, Tufts University School of Medicine, Boston, MA.

Other Acknowledgments: None

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

References

1. Hirschhorn JN. Genomewide association studies--illuminating biologic pathways. N Engl J Med. 2009; 360(17):1699–1701. [PubMed: 19369661]
2. Bodmer W, Bonilla C. Common and rare variants in multifactorial susceptibility to common diseases. Nat Genet. 2008; 40(6):695–701. [PubMed: 18509313]
3. Klein RJ, Zeiss C, Chew EY, et al. Complement factor H polymorphism in age-related macular degeneration. Science. 2005; 308(5720):385–389. [PubMed: 15761122]
4. Hageman GS, Anderson DH, Johnson LV, et al. A common haplotype in the complement regulatory gene factor H (HF1/CFH) predisposes individuals to age-related macular degeneration. Proc Natl Acad Sci U S A. 2005; 102(20):7227–7232. [PubMed: 15870199]
5. Souied EH, Leveziel N, Richard F, et al. Y402H complement factor H polymorphism associated with exudative age-related macular degeneration in the French population. Mol Vis. 2005; 11:1135–1140. [PubMed: 16379025]
6. Maller J, George S, Purcell S, et al. Common variation in three genes, including a noncoding variant in CFH, strongly influences risk of age-related macular degeneration. Nat Genet. 2006; 38(9):1055–1059. [PubMed: 16936732]
7. Rivera A, Fisher SA, Fritsche LG, et al. Hypothetical LOC387715 is a second major susceptibility gene for age-related macular degeneration, contributing independently of complement factor H to disease risk. Hum Mol Genet. 2005; 14(21):3227–3236. [PubMed: 16174643]
8. Gold B, Merriam JE, Zernant J, et al. Variation in factor B (BF) and complement component 2 (C2) genes is associated with age-related macular degeneration. Nat Genet. 2006; 38(4):458–462. [PubMed: 16518403]
9. Fagerness JA, Maller JB, Neale BM, Reynolds RC, Daly MJ, Seddon JM. Variation near complement factor I is associated with risk of advanced AMD. Eur J Hum Genet. 2009; 17(1):100–104. [PubMed: 18685559]
10. Maller JB, Fagerness JA, Reynolds RC, Neale BM, Daly MJ, Seddon JM. Variation in complement factor 3 is associated with risk of age-related macular degeneration. Nat Genet. 2007; 39(10): 1200–1201. [PubMed: 17767156]
11. Yates JR, Sepp T, Matharu BK, et al. Complement C3 variant and the risk of age-related macular degeneration. N Engl J Med. 2007; 357(6):553–561. [PubMed: 17634448]
12. Neale BM, Fagerness J, Reynolds R, et al. Genome-wide association study of advanced age-related macular degeneration identifies a role of the hepatic lipase gene (LIPC). Proc Natl Acad Sci U S A. 2010; 107(16):7395–7400. [PubMed: 20385826]
13. Chen W, Stambolian D, Edwards AO, et al. Genetic variants near TIMP3 and high-density lipoprotein-associated loci influence susceptibility to age-related macular degeneration. Proc Natl Acad Sci U S A. 2010; 107(16):7401–7406. [PubMed: 20385819]
14. Manolio TA, Collins FS, Cox NJ, et al. Finding the missing heritability of complex diseases. Nature. 2009; 461(7265):747–753. [PubMed: 19812666]
15. Seddon JM, Reynolds R, Maller J, Fagerness JA, Daly MJ, Rosner B. Prediction model for prevalence and incidence of advanced age-related macular degeneration based on genetic, demographic, and environmental variables. Invest Ophthalmol Vis Sci. 2009; 50(5):2044–2053. [PubMed: 19117936]

Biographies

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Johanna M. Seddon, MD, ScM. is Professor of Ophthalmology at Tufts University School of Medicine, Founding Director of the Ophthalmic Epidemiology and Genetics Service at Tufts Medical Center, and retina specialist at New England Eye Center, Boston, MA. Her primary interests are genetic-epidemiology of ophthalmic diseases, and identifying preventive factors and cures for macular degenerations.

Inga Peter is Associate Professor of Genetics and Genomic Sciences at Mount Sinai School of Medicine, and Associate Director of Statistical Genetics and Genetic Epidemiology at the Mount Sinai Genomic Institute. Her primary interests are genetic epidemiology of complex diseases and pharmacogenetics aimed to identify genetic markers for early diagnosis and personalized treatment.