Published in final edited form as:
N C Med J. 2013 ; 74(6): 505–508.

Understanding Genetic Epidemiology: The Potential Benefits and Challenges of Genetics for Improving Human Health

Amanda A Seyerle, MSPH and
Cardiovascular Disease Epidemiology Trainee, Department of Epidemiology, Gillings School of Global Public Health, University of North Carolina at Chapel Hill, North Carolina

Christy L Avery, PhD
Assistant professor, Department of Epidemiology, Gillings School of Global Public Health, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina

Abstract

Genetic epidemiology has the potential to significantly impact human health. Here, we examine the major developments in the field's history and the current state of the field, including both promising avenues of research and potential challenges genetic epidemiologists face.

Understanding the distribution and determinants of human disease is of increasing priority in an age of burgeoning healthcare costs and rising burden of disease. At the crossroads of genetics and epidemiology is genetic epidemiology, which examines the role of inherited factors in disease etiology. Current public health benefits of genomics research are numerous, including an improved understanding disease mechanisms [1], targeted cancer treatments [2], and dosage regimens for pharmaceuticals [3]. However, to fully appreciate the current and potential future contributions of genetic epidemiology research to improving human health, one must first understand the major milestones in genetic epidemiology and the current challenges researchers face in the ongoing process of deciphering the human genome.

A relatively new field, genetic epidemiology was first described in 1954 by Neel and Schull [4,5]. At the time, molecular genetics was still in its infancy. Although direct measurement of genotypes was not yet possible, genetic epidemiologists investigated the inheritance of disease by examining whether patterns of inheritance (e.g. dominant, recessive, X-linked) were consistent with phenotype patterns observed in large families. For example, early segregation analyses of breast and ovarian cancer suggested a strong genetic etiology with an autosomal dominant mode of inheritance [6,7].

Linkage analysis, the framework of which was first published in 1980 [8], was a natural extension of segregation analysis. Made possible by technological advances that allowed the direct measurement of genotypes, linkage analysis used populations of related individuals to

Corresponding Author: Amanda A. Seyerle, Department of Epidemiology, University of North Carolina Chapel Hill, Franklin Street Plaza, 137 E. Franklin St., Suite 306, Chapel Hill, NC 27514, (919) 966-9304 (voice), (919) 966-9800 (fax), aseyerle@email.unc.edu.
Potential conflicts of interest: A.A.S. and C.L.A. have no relevant conflicts of interest to disclose.

assess the genetic basis of disease and successfully identified the genes responsible for numerous monogenic disorders, including Tay-Sachs, Huntington's, and cystic fibrosis [9]. However, linkage analysis failed to make inroads in identifying genes which were associated with complex, chronic diseases on a population level. For example, genes that were associated with rare familial forms of breast cancer in specific families via linkage, including variants in the *GPT* and *ACP* genes, showed no association with breast cancer in the general population [10].

Limitations of linkage analysis led researchers to investigate other approaches for the identification of genes associated with complex diseases, including candidate gene studies. Candidate gene studies, which evaluated evidence for association between the outcome of interest and variants in or near genes selected using *a priori* hypotheses became popular because they provided greater power for detecting associations for complex traits and could be performed in population-based cohort and case-control studies [11]. Despite leveraging biologic knowledge, few candidate gene studies successfully identified associations. For example, a 2009 review of candidate gene studies for obesity found that, of 21 genes examined in the candidate gene literature, only nine were found to have any association with obesity [12]. Additionally, very few positive results were replicated in subsequent studies [13,14].

In parallel with the increased rise in popularity of candidate gene studies was the publication of the first draft of the human genome by the Human Genome Project (HGP) in 2001 [15], which has had a lasting influence on the practice of genetic epidemiology. Briefly, the HGP sought to catalog human genetic variation by identifying all human genes and sequencing the three billion bases in the human genome [16]. One significant advance made possible by the HGP were large-scale human genome studies, including genome-wide association studies (GWAS), which allowed researchers to test associations between traits of interest and single base pair changes, called single nucleotide polymorphisms (SNPs) that are spaced throughout the human genome [17]. Large scale GWAS greatly increased the coverage provided by candidate gene studies, which typically evaluated a small number of SNPs at a handful of pre-specified genes. Using GWAS, genetic epidemiology has made significant progress unraveling the genomic etiology of complex diseases. As of August, 2013, GWAS have successfully reported over 11,000 SNPs to be associated with a wide variety of diseases and their risk factors [18].

Notably, GWAS often identify pathways largely ignored by candidate gene studies. One such example is the association between obesity and the *FTO* gene, which encodes an mRNA demethylase [19] and was identified through a fused-toe phenotype in mice [20]. Although not a compelling obesity candidate gene, *FTO*, one of the first obesity loci identified by GWAS, has been successfully replicated across studies and populations [20–25]. Novel findings like the *FTO* gene also have prompted new avenues of inquiry in obesity research, including closer examination of the role between the control of energy expenditure and obesity [12], underscoring the ability of GWAS to illuminate novel biologic pathways underlying disease etiology.

Despite notable successes, genetic epidemiology has yet to fully elucidate the genetic basis of disease. With obesity, for example, studies have consistently suggested that a substantial fraction (16–85%) of the variation in body mass index (BMI), a common obesity metric, is genetic in origin [26–31]. Although GWAS have identified 36 genetic regions associated with BMI, these regions only explain a tiny proportion (<2%) of the estimated heritability of BMI [32,33]. One possible source of the missing heritability is gene-environment interaction [34]. In epidemiology, it is commonly understood that disease is caused by *both* genetics and the environment, even for monogenic disorders like phenylketonuria (PKU). For example, a child may have the genotype for PKU but he or she will not develop the disease unless the child is exposed to phenylalanine. On a population-wide context, gene-environment interaction evaluates whether the magnitude of association between a genetic variant and complex trait differs by an environmental factor. As environmental exposures are more amenable to intervention than genetics, gene-environment studies may not only help identify the missing heritability for complex diseases but also offer the best avenue by which genomics research can contribute to improving public health [17].

One promising area of gene-environment interaction study is pharmacogenomics. Briefly, pharmacogenomics evaluates the genomic underpinnings of drug response to better understand adverse drug reactions and tailor individualized treatment [35]. As of 2011, there were 70 drugs for which the U.S. Food and Drug Administration (FDA) had approved new labeling to include information on genetic variants which affect the metabolism of the drug [17]. The classic example is warfarin, an anticoagulant used to prevent blood clots and embolisms. Today, genes encoding the VKORC1 and CYP2C9 proteins are routinely evaluated in clinical settings when assigning dosage regimens [3,36,37]. Similarly, the discovery of variants in the *CYP2C19* gene that reduce patients' ability to metabolize clopidogrel, an anti-clotting medication used to prevent stent thrombosis, have prompted an FDA announcement which warns that clopidogrel may not be effective in patients with specific genetic variants [38]. In response to the FDA announcement, drug companies have developed new drug therapies that work effectively in all patients, regardless of their *CYP2C19* genotype [38]. However, expanding pharmacogenomics research to examine diverse traits is needed, as is the inclusion of racial/ethnic minorities. The latter is especially important, as most pharmacogenetics research has examined European descent populations, although a great deal of genetic variation, including many of the clinically actionable pharmacogenetic SNPs, have been shown to vary substantially in frequency among ethnically diverse populations [39].

Although warfarin and clopidogrel highlight the translational potential of genetic epidemiology research, most of the findings in genetic epidemiology to date have not made the impact on public health that many believed were possible when the human genome was first sequenced. One reason for the perceived lack of impact is that many of the benefits currently found in genomics research are indirect. In 2011, the National Human Genome Research Institute published a perspective on genetic medicine, observing that "the most effective way to improve human health is to understand normal biology (in this case, genome biology) as a basis for understanding disease biology, which then becomes the basis

for improving health" [1]. Therefore, it remains difficult to fully ascertain the future promise of genetic epidemiology for the advancement of public health.

Equally important when assessing the future contributions of genetic epidemiology is the realization that disease etiology is complex and genetic risk does not equate to genetic determinism. The complex relationship between genetics and disease poses an ethical dilemma for healthcare practitioners regarding the return of genetic test results. Genetic tests may yield incidental findings, in addition to the results for which the test was ordered, as it is often more economical to sequence the entire genome rather than genotype specific regions. For example, a test for the Huntington's mutation could also yield results for diseases such as Tay-Sachs, breast cancer, or hereditary hemochromatosis [40]. There is a question of whether these incidental findings, which may be of potential medical value, should be returned to patients. Recently, the American College of Medical Genetics and Genomics (ACMG) published a set of guidelines which emphasized the need to alert patients to such incidental findings and even outlined what incidental findings should be returned to patients [41]. However, the ACMG also pointed out that there was insufficient data to make a clear recommendation for many previously identified genetic risk variants [41]. Incidental findings also are problematic in research settings where many large studies have performed extensive genotyping on study participants. For example, in 2012, the Electronic Medical Records and Genomics (eMERGE) Network, a collection of archives and biorepositories, some of which perform GWAS, identified a subset of genetic abnormalities with sufficient data to warrant returning results to patients; the decision on returning results was the responsibility of individual repositories [42]. Unfortunately, both physician and patient education is severely lacking regarding interpretation of incidental findings [43]. Returning results will only be valuable when both the physician and the patient understand what genetic information means and if or how it can be incorporated into clinical care [1].

Even with the potentials for benefit, it is important to understand that the role of genetic epidemiology in medical care, apart from a few notable exceptions, is largely undetermined. Even with advances like the HGP, many important questions remain unanswered, including how the genome differs in diverse populations, how the environment affects the genome, and what role epigenetics, or the study of heritable mechanisms which alter gene expression without altering the underlying genetic sequence, plays. It is therefore not surprising that it could easily be 2020 or beyond before genetics makes any significant impact on public health [1]. Even as genetic findings become more translational, the general public and even many healthcare professionals do not yet have a sufficient knowledge to utilize genetic information effectively. Educational efforts for both patients and practitioners that parallel current advances in genetic epidemiology are therefore integral to realizing the full potential of this quickly advancing field.

Acknowledgments

Financial Support: A.A.S. is supported with funding provided by NHLBI training grant T32HL7055. C.L.A. is supported with funding from NHLBI grant R00HL098458.

References

1. Green ED, Guyer MS. National Human Genome Research I. Charting a course for genomic medicine from base pairs to bedside. Nature. 2011; 470:204–213. [PubMed: 21307933]
2. Xing M, Haugen BR, Schlumberger M. Progress in molecular-based management of differentiated thyroid cancer. Lancet. 2013; 381:1058–1069. [PubMed: 23668556]
3. Perera MA, Cavallari LH, Limdi NA, Gamazon ER, Konkashbaev A, et al. Genetic variants associated with warfarin dose in African-American individuals: a genome-wide association study. Lancet. 2013
4. DeWan AT. Five classic articles in genetic epidemiology. Yale J Biol Med. 2010; 83:87–90. [PubMed: 20589189]
5. Neel, J.; Schull, W. Human Heredity. Chicago: Univ Chicago Pr; 1954. p. 361
6. Go RCP, King M-C, Bailey-Wilson J, Elston RC, Lynch HT. Genetic Epidemiology of Breast Cancer and Associated Cancers in High-Risk Families. I. Segregation Analysis. Journal of the National Cancer Institute. 1983; 71:455–461. [PubMed: 6577220]
7. Williams WR, Anderson DE, Rao DC. Genetic epidemiology of breast cancer: Segregation analysis of 200 Danish pedigrees. Genet Epidemiol. 1984; 1:7–20. [PubMed: 6544234]
8. Botstein D, White RL, Skolnick M, Davis RW. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. Am J Hum Genet. 1980; 32:314–331. [PubMed: 6247908]
9. Grace VM. Human genome epidemiology: reviewing the stakes. Int J Health Serv. 2008; 38:143–159. [PubMed: 18341127]
10. King M-C, Go RCP, Lynch HT, Elston RC, Terasaki PI, et al. Genetic Epidemiology of Breast Cancer and Associated Cancers in High-Risk Families. II. Linkage Analysis. Journal of the National Cancer Institute. 1983; 71:463–467. [PubMed: 6577221]
11. Tabor HK, Risch NJ, Myers RM. Candidate-gene approaches for studying complex genetic traits: practical considerations. Nat Rev Genet. 2002; 3:391–397. [PubMed: 11988764]
12. Loos RJ. Recent progress in the genetics of common obesity. Br J Clin Pharmacol. 2009; 68:811–829. [PubMed: 20002076]
13. Chanock SJ, Manolio T, Boehnke M, Boerwinkle E, Hunter DJ, et al. Replicating genotype-phenotype associations. Nature. 2007; 447:655–660. [PubMed: 17554299]
14. Siontis KC, Patsopoulos NA, Ioannidis JP. Replication of past candidate loci for common diseases and phenotypes in 100 genome-wide association studies. Eur J Hum Genet. 2010; 18:832–837. [PubMed: 20234392]
15. Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, et al. The Sequence of the Human Genome. Science. 2001; 291:1304–1351. [PubMed: 11181995]
16. Collins FS, McKusick VA. Implications of the human genome project for medical science. JAMA. 2001; 285:540–544. [PubMed: 11176855]
17. Khoury MJ, Gwinn M, Clyne M, Yu W. Genetic epidemiology with a capital E, ten years after. Genet Epidemiol. 2011; 35:845–852. [PubMed: 22125223]
18. Hindorff, LA.; MacArthur, J.; Morales, J.; Junkins, HA.; Hall, PN., et al. A Catalog of Published Genome-Wide Association Studies.
19. Jia G, Fu Y, Zhao X, Dai Q, Zheng G, et al. N6-methyladenosine in nuclear RNA is a major substrate of the obesity-associated FTO. Nat Chem Biol. 2011; 7:885–887. [PubMed: 22002720]
20. Fawcett KA, Barroso I. The genetics of obesity: FTO leads the way. Trends Genet. 2010; 26:266–274. [PubMed: 20381893]
21. Dusatkova L, Zamrazilova H, Sedlackova B, Vcelak J, Hlavaty P, et al. Association of obesity susceptibility gene variants with metabolic syndrome and related traits in 1,443 czech adolescents. Folia Biol (Praha). 2013; 59:123–133. [PubMed: 23890480]
22. Frayling TM, Timpson NJ, Weedon MN, Zeggini E, Freathy RM, et al. A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity. Science. 2007; 316:889–894. [PubMed: 17434869]

23. Ibba A, Pilia S, Zavattari P, Loche A, Guzzetti C, et al. The role of FTO genotype on eating behavior in obese Sardinian children and adolescents. J Pediatr Endocrinol Metab. 2013; 26:539–544. [PubMed: 23509208]
24. Scuteri A, Sanna S, Chen WM, Uda M, Albai G, et al. Genome-wide association scan shows genetic variants in the FTO gene are associated with obesity-related traits. PLoS Genet. 2007:3.
25. Xi B, Cheng H, Shen Y, Chandak GR, Zhao X, et al. Study of 11 BMI-associated loci identified in GWAS for associations with central obesity in the Chinese children. PLoS One. 2013; 8:e56472. [PubMed: 23424664]
26. Adeyemo A, Luke A, Cooper R, Wu X, Tayo B, et al. A genome-wide scan for body mass index among Nigerian families. Obes Res. 2003; 11:266–273. [PubMed: 12582223]
27. Allison DB, Kaprio J, Korkeila M, Koskenvuo M, Neale MC, et al. The heritability of body mass index among an international sample of monozygotic twins reared apart. Int J Obes Relat Metab Disord. 1996; 20:501–506. [PubMed: 8782724]
28. McQueen MB, Bertram L, Rimm EB, Blacker D, Santangelo SL. A QTL genome scan of the metabolic syndrome and its component traits. BMC Genet. 2003; 4(Suppl 1):S96. [PubMed: 14975164]
29. Pietilainen KH, Kaprio J, Rissanen A, Winter T, Rimpela A, et al. Distribution and heritability of BMI in Finnish adolescents aged 16y and 17y: a study of 4884 twins and 2509 singletons. Int J Obes Relat Metab Disord. 1999; 23:107–115. [PubMed: 10078843]
30. Platte P, Papanicolaou GJ, Johnston J, Klein CM, Doheny KF, et al. A study of linkage and association of body mass index in the Old Order Amish. Am J Med Genet C Semin Med Genet. 2003; 121C:71–80. [PubMed: 12888987]
31. Yang W, Kelly T, He J. Genetic Epidemiology of Obesity. Epidemiologic Reviews. 2007; 29:49–61. [PubMed: 17566051]
32. Xia Q, Grant SF. The genetics of human obesity. Ann N Y Acad Sci. 2013; 1281:178–190. [PubMed: 23360386]
33. Monda KL, Chen GK, Taylor KC, Palmer C, Edwards TL, et al. A meta-analysis identifies new loci associated with body mass index in individuals of African ancestry. Nat Genet. 2013; 45:690–696. [PubMed: 23583978]
34. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, et al. Finding the missing heritability of complex diseases. Nature. 2009; 461:747–753. [PubMed: 19812666]
35. Bochud M. Genetics for clinicians: from candidate genes to whole genome scans (technological advances). Best Pract Res Clin Endocrinol Metab. 2012; 26:119–132. [PubMed: 22498243]
36. Daneshjou R, Tatonetti NP, Karczewski KJ, Sagreiya H, Bourgeois S, et al. Pathway analysis of genome-wide data improves warfarin dose prediction. BMC Genomics. 2013; 14(Suppl 3):S11. [PubMed: 23819817]
37. Niinuma Y, Saito T, Takahashi M, Tsukada C, Ito M, et al. Functional characterization of 32 CYP2C9 allelic variants. Pharmacogenomics J. 2013
38. Holmes DR Jr. Pharmacogenomic testing and antithrombotic therapy: ready for prime time? Rambam Maimonides Med J. 2013; 4:e0005. [PubMed: 23908855]
39. Ramos E, Doumatey A, Elkahloun AG, Shriner D, Huang H, et al. Pharmacogenomics, ancestry and clinical decision making for global populations. Pharmacogenomics J. 2013
40. Green RC, Berg JS, Berry GT, Biesecker LG, Dimmock DP, et al. Exploring concordance and discordance for return of incidental findings from clinical sequencing. Genet Med. 2012; 14:405–410. [PubMed: 22422049]
41. Green RC, Berg JS, Grody WW, Kalia SS, Korf BR, et al. ACMG recommendations for reporting of incidental findings in clinical exome and genome sequencing. Genet Med. 2013; 15:565–574. [PubMed: 23788249]
42. Fullerton SM, Wolf WA, Brothers KB, Clayton EW, Crawford DC, et al. Return of individual research results from genome-wide association studies: experience of the Electronic Medical Records and Genomics (eMERGE) Network. Genet Med. 2012; 14:424–431. [PubMed: 22361898]
43. Attia JIJATA, et al. How to use an article about genetic association: C: what are the results and will they help me in caring for my patients? JAMA. 2009; 301:304–308. [PubMed: 19155457]