
Genes, Environment, Health, and Disease: Facing up to Complexity

Teri A. Manolio Francis S. Collins

National Human Genome Research Institute, NIH, Bethesda, Md., USA

The deciphering of the human genome sequence, the DNA instruction book that makes us not only uniquely human but also unique as individuals, has provided unparalleled opportunities for identifying the relationship of genetic variation to health and disease. Yet as we move from single-gene diseases, in which a single misspelled base-pair can lead to such devastating conditions as sickle cell disease or Hutchinson-Gilford progeria syndrome, to complex diseases involving interactions among many interacting genes of small effect and many potential environmental modifiers, we need to develop tools to unravel the complexity of these interacting risk factors. Even in single-gene diseases, the frequent inter-individual variability in disease manifestations and course suggests the presence of modifiers for these presumably simple diseases, whether due to differing genetic backgrounds or differing environmental exposures, that are only beginning to be identified [1].

This special issue of *Human Heredity* explores challenges in the identification and interpretation of gene \times gene and gene \times environment interactions, in anticipation of the massive amount of genomic variation data about to be generated in upcoming genome-wide association studies (websites 1–7) [2, 3]. Interactions (primarily among environmental factors or physical characteristics) were of course known to be important long before the advent of genome-wide association studies, with classic examples such as aflatoxin and chronic hepatitis B infection on liver cancer [4] and smoking and asbestos exposure on lung cancer [5].

Early gene-disease association studies followed these leads by looking for variations in risk relationships by presence or absence of an environmental factor, such as the angiotensin converting enzyme insertion/deletion polymorphism and myocardial infarction in persons with low body mass index and low plasma apolipoprotein B levels [6], or the methylenetetrahydrofolate reductase thermolabile variant and homocysteine levels in persons with low folate intake [7]. Variation in genotype-phenotype associations has also been noted in relation to other genetic variants, such as the interaction of alpha- and beta-adrenergic receptor variants on risk of heart failure [8] or the interaction of interleukin 4 receptor- α variants and interleukin 13 promoter variants on risk of asthma [9]. Sometimes it may have seemed that such interactions were more avidly sought when initial attempts to identify gene-disease associations were unsuccessful [10, 11], and vigilance is needed to avoid the perils of data-dredging [12, 13]. The critical importance of gene-environment interactions for designing therapies, however, has been amply demonstrated from a large body of experience in Mendelian genetics, most notably from environmentally modulated single-gene conditions such as phenylketonuria [14]. Most notably, at a time when gene therapy approaches have presented unanticipated difficulties [15, 16], the potential for alleviating serious consequences of a genetic variant through environmental manipulation has appeared ever more attractive.

A wide range of issues related to gene \times gene and gene \times environment interactions has been addressed by the

authors invited to contribute to this issue, including: strategies for detecting associations with interacting genes [17] or interacting genetic and environmental factors [18] that have minimal apparent influence on their own; pursuit of gene \times environment interactions to detect marginal genetic associations [19]; study designs for identifying environmental modifiers of rare mutations [20]; impact of genotype misclassification on detection of gene \times environment interactions [21]; assumption-free analytic methods for detecting gene \times gene interactions [22]; importance of gene \times gene interactions in failure to replicate associations [23]; and impact of gene \times gene interactions on parametric linkage analysis [24]. These papers highlight the potential pitfalls of ignoring interactions, both in failing to identify genetic and environmental risk factors of potential importance, and in falsely ascribing a role in disease risk to a genetic or environmental factor only because of the company it keeps [25].

The importance of developing such methods cannot be over-emphasized as we face the coming flood of genome-wide association data. Hundreds of thousands of genetic variants are about to be tested in many thousands of well-characterized individuals, each of whom may in turn have many thousands of data points available on environmental exposures, physical and behavioral characteristics, family history, and intra-individual change over time. Among the best-known of these may be NHLBI's Framingham Heart Study, which is finalizing plans for genome-wide association testing in 9,000 study participants in three generations. For five decades, Framingham has been at the forefront of identifying cardiovascular disease risk factors such as hypertension, dyslipidemia, and smoking [26], family history [27], obesity [28], and diabetes [29], that arguably are due to both environmental and genetic influences. Framingham participants have also been characterized for other important non-cardiovascular conditions such as dementia [30], osteoarthritis [31], cancer [32], pulmonary impairment [33], vision [34] and hearing loss [35], and risk factors for these conditions have been identified as well. Yet few true risk factor interactions have been reported from Framingham, despite ample demonstration of the additive nature of many of these risk factors [36] – does this mean that interactions are truly infrequent or that analytic (or measurement) methods to date have been inadequate to detect them?

Despite concerns voiced regarding the impact of genotyping errors on failure to replicate associations [37], it is generally accepted that reliability of genotyping methodology is far greater than that of exposure measurement

[38]. Current high throughput low cost platforms can routinely achieve error rates that are in the range of 0.1% or even lower, a testimony to the unprecedented investment in genotyping technology stimulated by the Human Genome Project and the International HapMap Project. For error rates in this range, Tung et al. [21] document in this issue that there is virtually no loss in power to detect gene \times environment interactions.

Whether similar technological advances are possible with environmental assessment is a question to be addressed in part by the exposure biology component of the proposed Genes and Environment Initiative (GEI), spearheaded by the National Institute of Environmental Health Sciences and several collaborating NIH Institutes (website 8). This program, if approved by Congress, will support development of a wide range of measures of personal exposures, behaviors, and their physical and psychological consequences that contribute to health and disease. Given the influence that misclassification of exposures has been projected to exert on gene-environment interaction research [39], the initiation of this and similar programs should be expected to produce major advances in this field of research.

The Genes and Environment Initiative is also expected to provide a major infusion of funding into genome-wide association studies through its genetic component, spearheaded by the National Human Genome Research Institute in collaboration with several other NIH Institutes [website G]. Up to 15 such studies are projected to be supported through this program alone in the next three years. Of particular relevance to this special issue of *Human Heredity*, GEI is proposed to include substantial support for data analysis and development of analytic methods, with a recently released solicitation on analysis of sequencing data (website 9), in addition to the current ENDGAME project in statistical analysis of genome-wide association studies (website 10). Other NIH Institutes have already begun major genome-wide association initiatives, including the National Cancer Institute's Cancer Genetic Markers of Susceptibility (CGEMS) project in prostate cancer, which recently posted its initial genotype-phenotype association data in publicly accessible form (website 3). More than a dozen genome-wide association studies of common disease will be carried out over the next few months by the multi-component Wellcome Trust Case-Control Consortium (website 1) and the Genetic Association Information Network (website 2), both supported through public-private partnerships. Similar new NIH-supported projects in type I diabetes and diabetic nephropathy (website 5) and in heart, lung,

and blood diseases (website 4) will expand upon newly-reported genome-wide studies in macular degeneration [40, 41], Parkinson's disease [42], and inflammatory bowel disease [43].

Central to many of the NIH initiatives are plans for widespread data accessibility, most likely through controlled access to a database housed and curated by the National Center for Biotechnology Information of the National Library of Medicine. Investigators committing to maintain participant confidentiality, respect informed consent, and allow primacy of publication to the investigators who collected and submitted the data, will receive rapid and unprecedented access to these data for etiologic research, methods development, replication of findings, and likely a host of unforeseen and unforeseeable research applications. Given the large investment in this area and the wide range of potential uses of these data, consistent of course with participants' informed consent, NIH is developing standard policies for data-sharing that are projected to be announced in early 2007 (website 11).

The timing of this special issue could thus not have been better. We hope it is but a first installment for the many intriguing and innovative analytic approaches likely to be stimulated by the widespread availability of

genome-wide association data. More importantly, we hope that application of novel methods and creative minds to the growing mass of such data will yield major advances in identifying the genetic influences on complex disease risks, and on the environmental factors that will permit us to reduce or eliminate those risks.

Websites

1. <http://www.wtccc.org.uk/>
2. http://www.fnih.org/GAIN/GAIN_home.shtml
3. <http://cgems.cancer.gov/>
4. <http://grants.nih.gov/grants/guide/rfa-files/RFA-HL-06-012.html>
5. <http://grants.nih.gov/grants/guide/rfa-files/RFA-DK-06-005.html>
6. <http://www.nhlbi.nih.gov/new/press/06-02-06.htm>
7. <http://www.genome.gov/19518663> (GEI)
8. <http://www.gei.nih.gov/exposurebiology/funding.asp>
9. <http://grants.nih.gov/grants/guide/rfa-files/RFA-HL-07-010.html>
10. <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1513315>
11. http://grants.nih.gov/grants/gwas/fact_sheet.htm

References

- 1 Drumm ML, Konstan MW, Schluchter MD, Handler A, Pace R, Zou F, Zariwala M, Fargo D, Xu A, Dunn JM, Darrach RJ, Dorfman R, Sandford AJ, Corey M, Zielenski J, Durie P, Goddard K, Yankaskas JR, Wright FA, Knowles MR, Gene Modifier Study Group. Genetic modifiers of lung disease in cystic fibrosis. *N Engl J Med* 2005;353:1443-1453.
- 2 Hirschhorn JN, Daly MJ: Genome-wide association studies for common diseases and complex traits. *Nat Rev Genet* 2005;6:95-108.
- 3 Thomas DC: Are we ready for genome-wide association studies? *Cancer Epidemiol Biomarkers Prev* 2006;15:595-598.
- 4 Qian GS, Ross RK, Yu MC, Yuan JM, Gao YT, Henderson BE, Wogan GN, Groopman JD: A follow-up study of urinary markers of aflatoxin exposure and liver cancer risk in Shanghai, People's Republic of China. *Cancer Epidemiol Biomarkers Prev* 1994;3:3-10.
- 5 Selikoff IJ, Hammond EC, Churg J: Asbestos exposure, smoking, and neoplasia. *JAMA* 1968;204:106-112.
- 6 Cambien F, Poirier O, Lecerf L, Evans A, Cambou JP, Arveiler D, Luc G, Bard JM, Bara L, Ricard S, et al: Deletion polymorphism in the gene for angiotensin-converting enzyme is a potent risk factor for myocardial infarction. *Nature* 1992;359:641-644.
- 7 Jacques PF, Bostom AG, Williams RR, Ellison RC, Eckfeldt JH, Rosenberg IH, Selhub J, Rozen R: Relation between folate status, a common mutation in methylenetetrahydrofolate reductase, and plasma homocysteine concentrations. *Circulation* 1996;93:7-9.
- 8 Small KM, Wagoner LE, Levin AM, Kardia SL, Liggett SB: Synergistic polymorphisms of beta1- and alpha2C-adrenergic receptors and the risk of congestive heart failure. *N Engl J Med* 2002;347:1135-1142.
- 9 Howard TD, Koppelman GH, Xu J, Zheng SL, Postma DS, Meyers DA, Bleecker ER: Gene-gene interaction in asthma: IL4RA and IL13 in a Dutch population with asthma. *Am J Hum Genet* 2002;70:230-236.
- 10 Patsch W, Sharrett AR, Chen IY, Lin-Lee YC, Brown SA, Gotto AM Jr, Boerwinkle E: Associations of allelic differences at the A-I/C-III/A-IV gene cluster with carotid artery intima-media thickness and plasma lipid transport in hypercholesterolemic-hypertriglyceridemic humans. *Arterioscler Thromb* 1994;14:874-883.
- 11 Jerrard-Dunne P, Sitzer M, Risley P, Steckel DA, Buehler A, von Kegler S, Markus HS: Carotid Atherosclerosis Progression Study. Interleukin-6 promoter polymorphism modulates the effects of heavy alcohol consumption on early carotid artery atherosclerosis: the Carotid Atherosclerosis Progression Study (CAPS). *Stroke* 2003;34:402-407.
- 12 Feinleib M: Data bases, data banks and data dredging: The agony and the ecstasy. *J Chronic Dis* 1984;37:783-790.
- 13 Furberg CD, Byington RP: What do subgroup analyses reveal about differential response to beta-blocker therapy? The Beta-Blocker Heart Attack Trial experience. *Circulation*. 1983;676(Pt 2):198-101.
- 14 National Institutes of Health Consensus Development Panel. National Institutes of Health Consensus Development Conference Statement: phenylketonuria: screening and management, October 16-18, 2000. *Pediatrics* 2001;108:972-982.
- 15 Hacein-Bey-Abina S, von Kalle C, Schmidt M, Le Deist F, Wulffraat N, McIntyre E, Radford I, Villeval JL, Fraser CC, Cavazzana-Calvo M, Fischer A: A serious adverse event after successful gene therapy for X-linked severe combined immunodeficiency. *N Engl J Med* 2003;348:255-256.

- 16 Raper SE, Chirmule N, Lee FS, Wivel NA, Bagg A, Gao GP, Wilson JM, Batshaw ML: Fatal systemic inflammatory response syndrome in a ornithine transcarbamylase deficient patient following adenoviral gene transfer. *Mol Genet Metab* 2003;80:148–158.
- 17 Kotti S, et al: Strategy for detecting susceptibility genes with weak or no marginal effect. *Human Hered* 2007;2:85–92.
- 18 Culverhouse R, et al: The use of the restricted partition method with case-control data. *Human Hered* 2007;2:93–100.
- 19 Kraft P, et al: Exploiting gene-environment interaction to detect genetic associations. *Human Hered* 2007;2:111–119.
- 20 Whittemore AS: Assessing environmental modifiers of disease risk associated with rare mutations. *Human Hered* 2007;2:134–143.
- 21 Tung L, et al: The impact of genotype misclassification errors on the power to detect a gene-environment interaction using cox proportional hazards modeling. *Human Hered* 2007;2:101–110.
- 22 Moore JH, et al: Symbolic modeling of epistasis. *Human Hered* 2007;2:120–133.
- 23 Musani SK, et al: Detection of gene x gene interactions in genome-wide association studies of human population data. *Human Hered* 2007;2:67–84.
- 24 Sung YJ, et al: Accounting for epistasis in linkage analysis of general pedigrees. *Human Hered* 2007;2:144–152.
- 25 Campbell CD, Ogburn EL, Lunetta KL, Lyon HN, Freedman ML, Groop LC, Altshuler D, Ardlie KG, Hirschhorn JN: Demonstrating stratification in a European American population. *Nat Genet* 2005;37:868–872.
- 26 Kannel WB, Dawber TR, Kagan A, Revotskie N, Stokes JI: Factors of risk in the development of coronary heart disease – six year follow-up experience; the Framingham Study. *Ann Intern Med* 1961;55:33–50.
- 27 Kannel WB, Feinleib M, McNamara PM, Garrison RJ, Castelli WP: An investigation of coronary heart disease in families. The Framingham offspring study. *Am J Epidemiol* 1979;110:281–290.
- 28 Hubert HB, Feinleib M, McNamara PM, Castelli WP: Obesity as an independent risk factor for cardiovascular disease: A 26-year follow-up of participants in the Framingham Heart Study. *Circulation* 1983;67:968–977.
- 29 Kannel WB, McGee DL: Diabetes and glucose tolerance as risk factors for cardiovascular disease: the Framingham Study. *Diabetes Care* 1979;2:120–126.
- 30 Seshadri S, Wolf PA, Beiser A, Au R, McNulty K, White R, D'Agostino RB: Lifetime risk of dementia and Alzheimer's disease. The impact of mortality on risk estimates in the Framingham Study. *Neurology* 1997;49:1498–1504.
- 31 McAlindon TE, Wilson PW, Aliabadi P, Weissman B, Felson DT: Level of physical activity and the risk of radiographic and symptomatic knee osteoarthritis in the elderly: the Framingham study. *Am J Med* 1999;106:151–157.
- 32 Kreger BE, Splansky GL, Schatzkin A: The cancer experience in the Framingham Heart Study cohort. *Cancer* 1991;67:1–6.
- 33 Walter RE, Beiser A, Givelber RJ, O'Connor GT, Gottlieb DJ: Association between glycemic state and lung function: the Framingham Heart Study. *Am J Respir Crit Care Med* 2003;167:911–916.
- 34 Leibowitz HM, Krueger DE, Maunder LR, Milton RC, Kini MM, Kahn HA, Nickerson RJ, Pool J, Colton TL, Ganley JP, Loewenstein JJ, Dawber TR: The Framingham Eye Study monograph: An ophthalmological and epidemiological study of cataract, glaucoma, diabetic retinopathy, macular degeneration, and visual acuity in a general population of 2631 adults, 1973–1975. *Surv Ophthalmol* 1980;24(suppl):335–610.
- 35 Gates GA, Cooper JC Jr, Kannel WB, Miller NJ: Hearing in the elderly: the Framingham cohort, 1983–1985. Part I. Basic audiometric test results. *Ear Hear* 1990;11:247–256.
- 36 Kannel WB, Larson M: Long-term epidemiologic prediction of coronary disease. The Framingham experience. *Cardiology* 1993;82:137–152.
- 37 Clayton DG, Walker NM, Smyth DJ, Pask R, Cooper JD, Maier LM, Smink LJ, Lam AC, Ovington NR, Stevens HE, Nutland S, Howson JM, Faham M, Moorhead M, Jones HB, Falkowski M, Hardenbol P, Willis TD, Todd JA: Population structure, differential bias and genomic control in a large-scale, case-control association study. *Nat Genet* 2005;37:1243–1246.
- 38 Schwartz DA, Weis B, Wilson SH: The need for exposure health sciences. *Environ Health Perspect* 2005;113:A650.
- 39 Garcia-Closas M, Rothman N, Lubin J: Misclassification in case-control studies of gene-environment interactions: Assessment of bias and sample size. *Cancer Epidemiol Biomarkers Prev* 1999;8:1043–1050.
- 40 Klein RJ, Zeiss C, Chew EY, Tsai JY, Sackler RS, Haynes C, Henning AK, SanGiovanni JP, Mane SM, Mayne ST, Bracken MB, Ferris FL, Ott J, Barnstable C, Hoh J: Complement factor H polymorphism in age-related macular degeneration. *Science* 2005;308:385–389.
- 41 DeWan A, Liu M, Hartman S, Zhang S, Liu DTL, Zhao C, Tam POS, Chan WM, Lam DSC, Snyder M, Barnstable C, Pang CP, Hoh J: HTRA1 Promoter Polymorphism in Wet Age-Related Macular Degeneration. Published online October 19 2006;10.1126/science.1133807 (Science Express Reports) www.sciencexpress.org.
- 42 Fung HC, Scholz S, Matarin M, Simon-Sanchez J, Hernandez D, Britton A, Gibbs JR, Langefeld C, Stiebert ML, Schymick J, Okun MS, Mandel RJ, Fernandez HH, Foote KD, Rodriguez RL, Peckham E, De Vrieze FW, Gwinn-Hardy K, Hardy JA, Singleton A: Genome-wide genotyping in Parkinson's disease and neurologically normal controls: first stage analysis and public release of data. *Lancet Neurol* 2006;5:911–916.
- 43 Duerr RH, Taylor KD, Brant SR, Rioux JD, Silverberg MS, Daly MJ, Steinhart AH, Abraham C, Regueiro M, Griffiths A, Dassopoulos T, Bitton A, Yang H, Targan S, Datta LW, Kistner EO, Schumm LP, Lee A, Gregersen PK, Barmada MM, Rotter JI, Nicolae DL, Cho JH: A Genome-Wide Association Study Identifies IL23R as an Inflammatory Bowel Disease Gene. Published online October 26, 2006;10.1126/science.1135245 (Science Express Reports) www.sciencexpress.org.