Rethinking Our Public Health Genetics Research Paradigm

Since the sequencing of the human genome, tremendous resources have been dedicated to understanding how genetic determinants may drive the production of disease. Despite some successes, the promise of genetics research in these areas remains largely unrealized.

The focus on isolating individual (or clusters of) genes that may be associated with narrowly defined phenotypes in large part explains this discrepancy. In particular, efforts to identify genotypes associated with narrow phenotypes force the field to use study designs that capitalize on homogeneous samples to minimize the potential for competing influences or confounders, which imposes important limitations on understanding the role of genes in human health.

We argue that a population health genetics that incorporates genetics into large, multiwave, multilevel cohorts has the best potential to clarify how genes, in combination and with the environment, jointly influence population health. (*Am J Public Health.* 2013;103: S14–S18. doi:10.2105/AJPH. 2012.301127)

Abdulrahman M. El-Sayed, DPhil, Karestan C. Koenen, PhD, and Sandro Galea, MD, DrPH

IT IS NEARLY IMPOSSIBLE TO overestimate the influence that genetics has had on the health research enterprise in the United States and throughout the world. The National Institutes of Health, the world's dominant funding source for health research, spends nearly a sixth of its overall operating budget on research in genetics, and that proportion continues to grow.[1] Concomitant with this investment, scientific output in genetics has exploded: the number of studies published exploring genetic associations in human health has nearly quadrupled, from just more than 2500 published in 2001 to just fewer than 10 000 in 2010.[2]

GENETICS IN PUBLIC HEALTH: THE FIRST 10 YEARS

The first decade of the "genomics revolution" in biomedical research has been dominated by clinical medicine. Guided by a concern about how genetics can improve clinical diagnosis and treatment, research has focused on identifying causal genetic variants for particular disease phenotypes. The hope is that such research will inform drug development, improve the capacity to target therapies to both intercellular genetic pathways and particular patient populations, and improve clinical outcomes. These approaches have inspired a rush of new technologies—hundreds of new drugs and thousands of clinical genetic tests have become available.[3,4] The field has been marked by some notable successes. For example, genetic information has been used to improve the diagnostic accuracy of breast cancer screening,[5] and major improvements have occurred in screening neonates after birth.[6]

However, two assumptions are implicit in this approach. First, the clinical focus of genetic inquiry that characterized its first 10 years suggests, first and foremost, a concern with isolating mechanisms that can improve individual health, in line with the notion that the clinical encounter is and should remain the dominant paradigm in health promotion. For example, scientists and clinicians alike are heralding the future of a "personalized medicine" in which genetic tests could quickly, comfortably, and cheaply characterize an individual patient's risk profile and likely responses to treatment.[7]

Second, this focus on an individual-based genetics research paradigm is ultimately informed by the notion that particular genotypes are simple risk markers for narrowly defined phenotypes—that someday researchers might characterize the gene "for" any particular disease. This assumption is dissonant with the complexity characterizing the influence of genetics on human health. Although this complexity and its implications for research are well recognized,[8–12] researchers' approaches have generally remained inured to the realities of polygeny or pleiotropy, gene–gene and gene–environment interactions, or epigenetic modifications that complicate their understanding of the role of genetics in human health.

The evidence that these assumptions are limited, and limiting, is abundant. First, even a strict adherence to an individual-medicine paradigm often acknowledges that, at their core, therapies that focus on individuals must also take into account the health needs of populations—the concept of public health genomics has often been invoked to suggest the import of genetics being applicable to population health.[13–15] However, this notion of public health genomics seldom advocates a departure from the focus on individuals, arguing, in effect, that the population health influence of genomics will be the mere aggregate of improvements in diagnosis and care at the individual level.[13,14] Second, a decade of human genome research has confirmed that the role of genes in human health is supremely complex and that the input of any particular locus is best understood in concert with other loci and environmental factors operating dynamically throughout the life course.[11]

These two limitations account in large part for the gap between the promise of genetics research and the outcomes. For example, although this genetics research paradigm has produced thousands of marketed clinical genetic tests,[3,4] little is known about how efficacious these tests may be at the population level.[13] What is more, because these clinical tools largely screen for rare diseases, they are unlikely to make a real or lasting dent in broader disease metrics, such as life expectancy or mortality. Beyond this, relegating genetics research to the clinical arena precludes any insight that genetics may provide into

preventing disease earlier in the causal pathway, well before patients present to their providers. Finally, health disparities are of crucial interest in population health.[16] Important racial, ethnic, and socioeconomic predictors of access to health care technologies exist,[17,18] suggesting that confining genetics research to the clinical space exacerbates the likelihood that people who are poor and marginalized will have the least access to breakthroughs informed by genetics research.

BROADENING THE LENS OF GENETIC INQUIRY

Although much of the narrow lens that has dominated genetic inquiry over the past decade is attributable to conceptual focus, it has been further reinforced by the study designs that have been adopted. These studies have had two dominant characteristics. First, in keeping with the focus of genetics research on identifying genotype–phenotype associations, researchers have used an array of ever-larger case–control studies among homogeneous populations. Case–control studies are designed to maximize power when studying relatively rare outcomes, and so large sample sizes are usually unnecessary when isolating reasonable effect sizes.[19] However, as it has become clear that relatively few large effect sizes are identified when the classic genotype–phenotype paradigm is applied to common, complex disorders, the field has moved ever further in the direction of large-scale case–control studies designed to isolate particular genotypes associated with phenotypes.[20] Second, these studies have made use of, as much as possible, homogeneous populations, which allows investigators to match cases and controls on third variables that

could be alternate explanations for the observed genotype–phenotype associations.

These studies have played a critical role in identifying loci involved in the etiology of complex diseases, but their potential is largely exhausted, and a transition in thinking is in order. Whereas previous studies were intended for gene discovery, future studies must situate genetic information in the context of the complex social and environmental etiology of these outcomes. Persisting with current study designs reinforces a broader disciplinary focus on an individualized genetics by implicitly diverting attention from the broader social and environmental determinants of health and taking the field further away from a useful public health genomics.

Current dominant genetic epidemiological study designs are not equipped to move thinking forward. A focus on homogeneous samples restricts the capacity to understand how genetic factors

influence the health of human populations through complex pathways that involve genes acting in concert with other genes and with features of an individual's environment. By fixing populations, these studies effectively limit variability in third variables such as environmental characteristics and other genetic and behavioral characteristics of study participants that contribute to disease causation and that, if not properly accounted for, introduce unmeasured confounding.

We illustrate two circumstances in which current study designs may pose a problem with respect to causal inference regarding the genotype–phenotype relationship. First, as shown in Figure 1, these designs are not equipped to characterize the complexity that undergirds genetic influences on health today. For example, epigenetic modification plays an important role in gene expression.[8–10] However, epigenetic involvement usually implies environmentally

induced modifications at gene-promoter regions that in turn influence gene expression. Properly measuring these phenomena requires measurement at three distinct time points, at minimum; investigators must measure the environmental antecedent at one point, the epigenetic modification at the second, and the phenotype of interest at the third. In this way, limitations in study design have challenged the ability to understand time-dependent genetic processes, such as epigenetic modifications, because they do not incorporate multiple measurements or information about environmental antecedents. A recent systematic literature review about epigenetic modification in the etiology of mood and anxiety disorders and suicide, for example, found 21 case-control studies, of which only three included data about common environmental stressors in relation to epigenetic modifications and the phenotypes of interest.[21]
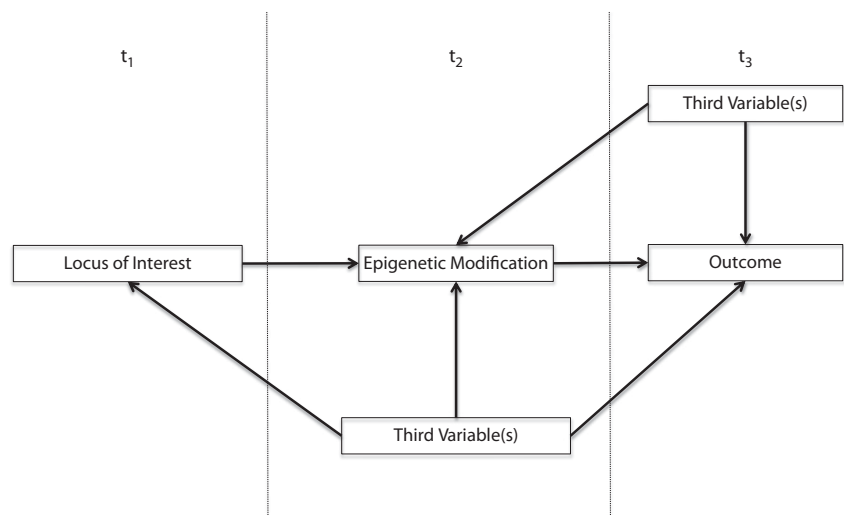
Note. t_1 = time 1; t_2 = time 2; t_3 = time 3. Support of epigenetic hypotheses requires measurement at three time points, and third variables that might confound relationships among loci, epigenetic modification, and outcomes should be measured at the second and third time points and adjusted for appropriately in analyses.

FIGURE 1—Epigenetic modification as a mediator between genetic factors and outcomes and associated confounders.
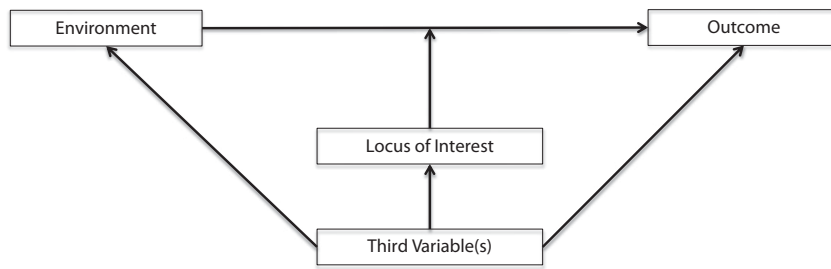
Note. Third variables can predict all three factors, as can occur in selected study samples, and can challenge population-level inference regarding the role of genes as effect modifiers of relationships of interest.

FIGURE 2—Effect modification of the relationship between an environment (genetic factors) and an outcome and associated confounders by a locus of interest.

Second, as shown in Figure 2, studies with highly selected samples limit investigators' ability to assess effect modification across any number of other genetic, demographic, behavioral, or environmental factors. The influence of genes in human health is complex and varies across a host of other factors, and homogeneous studies limit researchers' capacity to test hypotheses regarding these variations. For example, one seminal study considered effect modification of the relationship between the fat mass and obesity-associated (*FTO*) gene and body mass index by a third variable, physical activity.[22] The authors demonstrated that the relationship between an *FTO* polymorphism and body mass index was dependent on physical activity level.[22] In participants with low physical activity, *FTO* was found to predict higher body mass index, and in those with high physical activity, no relationship was found between *FTO* and body mass index.[22]

Similar findings have also been noted across population groups; gene–environment interactions have been observed in some racial groups, but not others. For example, in a study of the role of four single-nucleotide polymorphisms at the *FKBP5* locus as modifiers of

the association between childhood maltreatment and risk of post-traumatic stress disorder, Xie et al. demonstrated gene–environment interaction among African Americans, but not among European Americans.[23] Another study assessed the role of *MAOA* in modifying the association between childhood abuse and dysthymia.[24] The authors found significant three-way interactions among childhood abuse, *MAOA,* and race, as well as among childhood abuse, *MAOA,* and sex. In this way, the homogeneous samples that are ubiquitously used in extant studies limit inference regarding the mechanisms that may underlie the effects observed.

Similarly, one can consider the human life course as a third variable across which the influence of genetics on health may be modified. From this perspective, present genotype–phenotype studies, largely retrospective or cross-sectional, limit the understanding of the dynamics of the genotype–phenotype relationship throughout the life course—a crucial limitation because longitudinal studies have demonstrated variation in the relationship between genotypes and phenotypes with age.[11] Helpful examples come from the literature on genetic involvement in the etiology of obesity. A study by North et al.

with a longitudinal cohort of siblings demonstrated that the influence of genetics on change in body mass index were variable throughout childhood development.[12] Similarly, a study by Hardy et al. demonstrated that the relationship between polymorphisms at two loci associated with obesity, *FTO* and melanocortin 4 receptor (*MC4R*), changed between ages 2 to 20 years and ages 20 to 53 years in a longitudinal birth cohort in the United Kingdom.[11] These studies suggest that the role of genes in the etiology of obesity, and likely in that of other common, complex disorders, may be dynamic throughout the life course.

Given all of these issues, it is unsurprising that replicability has been a serious challenge for many genotype–phenotype associations in the literature.[25–27] Along with the important limitations discussed earlier, generalizability is an inherent limitation of highly selected samples, and differences in allelic frequency in populations selected on ethnicity or region can influence the likelihood of replication.[26,28]

RETHINKING STUDY DESIGNS

Although several authors have noted the persistent limitations of

genetic study designs,[25–27,29] a substantial gap remains between the observation that current study design techniques are poorly suited to the task at hand and the development and implementation of more robust population health study designs. Hence, despite the recognition of the limitations of current study designs in genetic epidemiology, genetic studies have largely had homogeneous samples and have not collected data about other factors external to the gene–outcome pathway.

As the complexity of genes' role in human health becomes more apparent, researchers clearly need to take a different perspective; a broader lens is needed to allow researchers to better understand the role of genetics in the production of human health. We suggest, then, that to fulfill the promise of a population health genetics, researchers need to take a different approach that highlights and clarifies the role of genetics in concert with social and environmental processes in the production of population patterns of health and disease. In this way, the end of population health genetics is not the production of diagnostic tests with moderate influence on the clinical decisions affecting individual care. Rather, it is on characterizing the causes of population health to improve the understanding of the heterogeneous influences of environmental exposures on population health and to target high-impact population interventions more effectively. The technological advances made as a result of the large investment in genetics research over the past decade have made DNA collection and genotyping widely available, standardized, and relatively inexpensive, which has enabled studies of cohorts with rich environmental data such as the

Great Smoky Mountain Study of Youth,[30] the Dunedin Longitudinal Study,[31] and the Atherosclerosis Risk in Communities study[32] to add genetic information offering the promise of insights beyond those provided by the extant approaches. These studies are promising early examples of those that can push the future of population health genetics.

Yet to continue to move beyond the limitations imposed by current trends in genetic study designs toward a paradigm whereby genetics can educate an understanding of health at the population level, the field must continue down this early path. Study designs must move beyond the restricted case–control studies that currently dominate the field toward studies that allow investigators to characterize the heterogeneity in genotype–phenotype associations, promote a life course perspective on the interplay between genes and health, and generalize beyond small subsections of the population. To accomplish this goal, genetics studies must continue to be folded into large, multiwave, heterogeneous population-representative samples that robustly measure other exposures of import to population health, in particular the social environment.

Several ongoing studies have begun to demonstrate the potential of large cohort studies to inform understanding of the role of genetics in population health. These studies include, for example, those mentioned earlier,[30–32] as well as the Coronary Artery Risk Development in Young Adults study,[33] the Multiethnic Study of Atherosclerosis,[34] and the Detroit Neighborhood Health Study (DNHS).[35] In addition, several studies have been able to combine population-level data

with biobanks that have also been used to further inference that combines genetic and environmental data, such as the Research Program on Genes, Environment, and Health, a collaboration between the University of California, San Francisco, and the Kaiser Permanente health system, which has collected genetic data for nearly 100 000 respondents.[36]

One of the key contributions of population health genetics going forward will be to identify effect modification of the relationship between alleles and phenotypes of interest by other factors of population health import. The role of the social environment will be central. In that regard, the DNHS is a particular exemplar of the potential contribution of large, multiwave studies in population health genetics. The DNHS is a multidisciplinary effort to bring together genetics, epigenetics, and relevant features of the social environment to better formulate a multifactorial model of the production of behavioral disorders. A multiwave study of a population-representative sample of 1547 adults in Detroit, Michigan, the DNHS uses systematic neighborhood assessments of each Detroit neighborhood to characterize each participant's social environment. Assessments consider signs of blight and deprivation, including the presence of graffiti, presence of vacant buildings, and street noise. In addition, data have been collected on baseline demographic characteristics, health status, social support, and traumatic exposure, and detailed information has been obtained regarding posttraumatic stress, depression, anxiety symptomatology, and substance abuse behavior. Whole blood samples have been collected and analyzed at 210 single nucleotide polymorphisms

involved in variation at 22 loci of interest in the etiology of psychopathology. To date, four waves of data have been collected.

Although not a gene discovery study, the DNHS typifies the incorporation of genetic data in the context of a large, multiwave cohort study with the capacity to assess the contributions of individual and contextual exposures alongside genetic information in the production of population health. These data allow investigators to consider heterogeneity of genetic influence across demographic, behavioral, and socioenvironmental exposures rather than simple genotype–phenotype associations. For example, a recent analysis from the study assessed whether serotonin transporter (SLC6A4) genotype or methylation status modified the relationship between the frequency of traumatic exposure and risk of posttraumatic stress disorder risk in a subset of the study population.[37] The investigators found that although neither genotype nor methylation level predicted posttraumatic stress disorder risk in main effects models, methylation level did, in fact, modify the relationship between the frequency of traumatic event exposure and posttraumatic stress disorder risk.

CONCLUSIONS

Although genetics research has throughout its first decade certainly managed to produce thousands of clinical genetic tests hailing a potential new dawn—in many minds, of a new personalized era in medicine[3,4]—this industry's growth has largely outstripped its ability to assess the efficacy of these implements at the population level,[13] where, at best, this overall approach has had only moderate success.[38–40]

Adopting a population-based genetics approach can lead to studies that collect data about multiple exposures in diverse samples, ultimately allowing researchers to characterize the pathways by which genetic factors influence population health. As we move through the second decade of genomic research, the opportunity to promote a population health genetics will rely on incorporating genetic components into these types of large, multiwave cohorts. Clearly, if genetics is to have its greatest impact on human health, this shift in methodology will be critical. Public health investigators interested in incorporating a population health genetics approach into their work might consider adding genetic components to existing cohorts or constructing study samples around understanding the role of genetics in population health. Beyond the production of clinical tests of unclear influence, this approach is primed to have far-reaching implications, educating researchers about society's most morbid and costly diseases and improving the field's capacity for population health promotion and disease prevention. Ultimately, we focused in this article on the necessity of a shift in study design. However, several other issues remain that require a reexamination of the current focus on an individual genomics. For example, the effect sizes of many genes thus far identified have been small,[41] raising important questions about the relative import of efforts that aim to improve public health that target genes versus other, perhaps more tractable features of the social environment. Also, population-based study design has limited capacity to help researchers discriminate individual-level risk,[42] somewhat obviating

one of the central goals of an individual genetics approach. Both of these points merit further consideration in future work. ∎

About the Authors

Abdulrahman M. El-Sayed, Karestan C. Koenen, and Sandro Galea are with the Department of Epidemiology, Mailman School of Public Health, Columbia University, New York, NY. Abdulrahman El-Sayed is also with the College of Physicians and Surgeons, Columbia University.

Correspondence should be sent to Sandro Galea, Gelman Professor and Chair, Columbia University, Mailman School of Public Health, Department of Epidemiology, 722 West 168th Street, 1508, New York, NY 10032 (e-mail: sgalea@columbia.edu). Reprints can be ordered at http://www.ajph. org by clicking the "Reprints" link.

This commentary was accepted October 22, 2012.

Contributors

A. M. El-Sayed drafted the article. S. Galea and K. C. Koenen conceptualized this article. S. Galea and K. C. Koenen critically edited the article for intellectual content. All authors take full responsibility for the article's integrity and accuracy.

References

1. Pohlhaus JR, Cook-Deegan RM. Genomics research: world survey of public funding. *BMC Genomics.* 2008;9(1):472.

2. HuGE Navigator. Available at: http://www.hugenavigator.net. Accessed March 20, 2012.

3. Gwinn M, Grossniklaus DA, Yu W, et al. Horizon scanning for new genomic tests. *Genet Med.* 2011;13(2):161–165.

4. National Center for Biotechnology Information. GeneTests. Available at: http://www.genetests.org. Accessed July 8, 2013.

5. van de Vijver MJ, He YD, Van't Veer LJ, et al. A gene-expression signature as a predictor of survival in breast cancer. *N Engl J Med.* 2002;347(25):1999–2009.

6. Lloyd-Puryear MA, Brower A. Long-term follow-up in newborn screening: a systems approach for improving health outcomes. *Genet Med.* 2010;12(12 suppl):S256–S260.

7. Collins FS. *The Language of Life: DNA and the Revolution in Personalized Medicine.* New York, NY: HarperCollins; 2010.

8. McGowan PO, Sasaki A, D'Alessio AC, et al. Epigenetic regulation of the glucocorticoid receptor in human brain associates with childhood abuse. *Nat Neurosci.* 2009;12(3):342–348.

9. Meaney MJ, Szyf M, Seckl JR. Epigenetic mechanisms of perinatal programming of hypothalamic-pituitary-adrenal function and health. *Trends Mol Med.* 2007;13(7):269–277.

10. Szyf M, Weaver I, Meaney M. Maternal care, the epigenome and phenotypic differences in behavior. *Reprod Toxicol.* 2007;24(1):9–19.

11. Hardy R, Wills AK, Wong A, et al. Life course variations in the associations between *FTO* and *MC4R* gene variants and body size. *Hum Mol Genet.* 2010;19(3):545–552.

12. North KE, Graff M, Adair LS, et al. Genetic epidemiology of BMI and body mass change from adolescence to young adulthood. *Obesity (Silver Spring).* 2010;18(7):1474–1476.

13. Khoury MJ, Gwinn M, Bowen MS, Dotson WD. Beyond base pairs to bedside: a population perspective on how genomics can improve health. *Am J Public Health.* 2012;102(1):34–37.

14. Khoury MJ, Gwinn M, Burke W, Bowen S, Zimmern R. Will genomics widen or help heal the schism between medicine and public health? *Am J Prev Med.* 2007;33(4):310–317.

15. Khoury MJ, McCabe LL, McCabe ERB. Population screening in the age of genomic medicine. *N Engl J Med.* 2003;348(1):50–58.

16. Krieger N. *Epidemiology and the People's Health: Theory and Context.* New York, NY: Oxford University Press; 2011.

17. Nelson A. Unequal treatment: confronting racial and ethnic disparities in health care. *J Natl Med Assoc.* 2002;94(8):666–668.

18. Kelley E, Moy E, Stryer D, Burstin H, Clancy C. The national healthcare quality and disparities reports: an overview. *Med Care.* 2005;43(3):I3–I8.

19. Schlesselman JJ, Stolley PD. *Case-Control Studies: Design, Conduct, Analysis.* Oxford, UK: Oxford University Press; 1982.

20. Burton PR, Hansell AL, Fortier I, et al. Size matters: just how big is BIG?: quantifying realistic sample size requirements for human genome epidemiology. *Int J Epidemiol.* 2009;38(1):263–273.

21. El-Sayed AM, Haloossim MR, Galea S, Koenen KC. Epigenetic modifications associated with suicide and common mood and anxiety disorders: a systematic review of the literature. *Biol Mood Anxiety Disord.* 2012;2(1):10.

22. Rampersaud E, Mitchell BD, Pollin TI, et al. Physical activity and the association of common FTO gene variants with body mass index and obesity. *Arch Intern Med.* 2008;168(16):1791–1797.

23. Xie P, Kranzler HR, Poling J, et al. Interaction of FKBP5 with childhood adversity on risk for post-traumatic stress disorder. *Neuropsychopharmacology.* 2010;35(8):1684–1692.

24. Nikulina V, Widom CS, Brzustowicz LM. Child abuse and neglect, MAOA, and mental health outcomes: a prospective examination. *Biol Psychiatry.* 2012;71(4):350–357.

25. Ioannidis JP, Ntzani EE, Trikalinos TA, Contopoulos-Ioannidis DG. Replication validity of genetic association studies. *Nat Genet.* 2001;29(3):306–309.

26. Colhoun HM, McKeigue PM, Davey Smith G. Problems of reporting genetic associations with complex outcomes. *Lancet.* 2003;361(9360):865–872.

27. Ryan SG. Regression to the truth: replication of association in pharmacogenetic studies. *Pharmacogenomics.* 2003;4(2):201–207.

28. Gambaro G, Anglani F, D'Angelo A. Association studies of genetic polymorphisms and complex disease. *Lancet.* 2000;355(9200):308–311.

29. Taneri B, Ambrosino E, van Os J, Brand A. A new public health genomics model for common complex diseases, with an application to common behavioral disorders. *Personalized Med.* 2012;9(1):29–38.

30. The Great Smoky Mountains Study. Available at: http://devepi.duhs.duke.edu/gsms.html. Accessed March 27, 2012.

31. Dunedin Study. Available at: http://dunedinstudy.otago.ac.nz/studies/main-study/description. Accessed May 27, 2012.

32. Atherosclerosis Risk in Communities Study. Available at: http://www.cscc.unc.edu/aric. Accessed August 1, 2012.

33. CARDIA: Coronary Artery Risk Development in Young Adults. Available at: http://www.cardia.dopm.uab.edu. Accessed August 1, 2012.

34. Multi-Ethnic Study of Atherosclerosis. 2012. Available at: http://www.mesa-nhlbi.org. Accessed August 1, 2012.

35. Detroit Neighborhood Health Study. 2012. Available at: http://detroitneighborhoodhealthstudy.org. Accessed March 27, 2012.

36. Kaiser Permanente. The research program on genes, environment, and health. 2012. Available at: http://www.dor.kaiser.org/external/DORExternal/rpgeh/index.aspx. Accessed September 4, 2012.

37. Koenen KC, Uddin M, Chang SC, et al. SLC6A4 methylation modifies the effect of the number of traumatic events on risk for posttraumatic stress disorder. *Depress Anxiety.* 2011;28(8):639–647.

38. McDermott U, Downing JR, Stratton MR. Genomics and the continuum of cancer care. *N Engl J Med.* 2011;364(4):340–350.

39. Wideroff L, Phillips KA, Randhawa G, et al. A health services research agenda for cellular, molecular and genomic technologies in cancer care. *Public Health Genomics.* 2009;12(4):233–244.

40. Armstrong K. Can genomics bend the cost curve? *JAMA.* 2012;307(10):1031–1032.

41. Ioannidis JP, Trikalinos TA, Khoury MJ. Implications of small effect sizes of individual genetic variants on the design and interpretation of genetic association studies of complex diseases. *Am J Epidemiol.* 2006;164(7):609–614.

42. Rockhill B. Theorizing about causes at the individual level while estimating effects at the population level: implications for prevention. *Epidemiology.* 2005;16(1):124–129.