American Journal of
EPIDEMIOLOGY

Volume 157

Number 8

April 15, 2003

Sponsored by the Society for Epidemiologic Research
Published by Oxford University Press

EDITORIAL

The Human Genome Project Is Complete. How Do We Develop a Handle for the Pump?

Julian Little[1,2], Muin J. Khoury[1], Linda Bradley[1], Mindy Clyne[1], Marta Gwinn[1], Bruce Lin[1], Mary-Lou Lindegren[1], and Paula Yoon[1]

[1] Office of Genomics and Disease Prevention, Centers for Disease Control and Prevention, Atlanta, GA.
[2] Epidemiology Group, Department of Medicine and Therapeutics, University of Aberdeen, Aberdeen, Scotland.

Received for publication February 19, 2003; accepted for publication February 27, 2003.

association; epidemiology; gene frequency; genetics; genomics; meta-analysis; review literature

Abbreviations: HuGE, Human Genome Epidemiology; HuGE Net, Human Genome Epidemiology Network.

"*The sequencing of the human genome offers the greatest opportunity for epidemiology since John Snow discovered the Broad Street Pump*" (1, p. 637).

A half century since the publication of the structure of the DNA molecule (2), the sequence of the human genome is complete. Some expect this achievement to be translated into advances in medicine and public health relatively rapidly (3–6). Proponents of this approach tend to focus on the potential to tailor primary prevention, secondary prevention, or therapy on the basis of genetic information. It is also possible that a better understanding of genetic effects and gene-environment interactions in disease processes will allow us to develop better interventions, such as avoidance of defined exposures and chemoprevention, to apply to the general population (7). Others are more skeptical (8–13). A key element in translation will be the application of epidemiologic studies to evaluate the role of genetic variants in the etiology of human disease (1). There has been a tremendous increase in the number and scope of peer-reviewed articles on human genome epidemiology, which has generated in turn the challenge of integrating these data, and this has led to many reviews of gene-disease associations and to gene-gene and gene-environment interaction in response. The purpose of this article is to provide an overview of the experience gained in integrating evidence in the Human Genome Epidemiology (HuGE) reviews and to suggest changes that may encourage more investigators to contribute reviews and to respond to changes in the character of the evidence.

HuGE reviews were established as a means of integrating evidence from human genome epidemiologic studies, that is, population-based studies of the impact of human genetic variation on health and disease (14). HuGE reviews are systematic, peer-reviewed synopses of epidemiologic aspects of human genes, including prevalence of allelic variants in different populations, population-based information on disease risk, evidence for gene-environment interaction, and quantitative data on genetic tests and services. They are carried out according to specified guidelines (15). As of February 2003, 20 such reviews have been published (table

Correspondence to Dr. Muin J. Khoury, Office of Genomics and Disease Prevention, National Center for Environmental Health, Centers for Disease Control and Prevention, 4770 Buford Highway, MS K-28, Atlanta, GA 30341 (e-mail: muk1@cdc.gov).

TABLE 1. HuGE* reviews, 1999–2002

Topic	Reference no.
Gene variants associated with a high risk of disease	
Medium chain acyl-CoA* dehydrogenase (MCAD) deficiency	16
Sickle hemoglobin (*HbS*) allele and sickle cell disease	17
NF1 gene and neurofibromatosis type 1	18
FMR1 and the fragile X syndrome	19
Mismatch repair genes *hMLH1* and *hMSH2* and colorectal cancer	20
Common complex disorders	
N-Acetyltransferase polymorphisms and colorectal cancer	26
Glutathione *S*-transferase polymorphisms and colorectal cancer	27
GSTM1, *GSTT1*, and risk of squamous cell carcinoma of the head and neck	29
Glutathione *S*-transferase polymorphisms and risk of ovarian cancer	30
Pooled analysis and meta-analysis of *GSTM1* and bladder cancer	32
NAD(P)H*:quinone oxidoreductase (*NQO1*) polymorphism, exposure to benzene, and predisposition to disease	86
5,10-Methylenetetrahydrofolate reductase (*MTHFR*) gene variants and congenital anomalies	25
5,10-Methylenetetrahydrofolate reductase polymorphisms and leukemia risk	33
Molecular epidemiology of vitamin D receptor gene variants	28
Apolipoprotein E polymorphism and cardiovascular disease	31
HLA-DQ and type 1 diabetes mellitus	23
HFE gene and hereditary hemochromatosis	24
δ-Aminolevulinic acid dehydratase (*ALAD*) genotype and lead toxicity	87
GJB2 (connexin 26) variants and nonsyndromic sensorineural hearing loss	88
Androgen receptor CAG repeats and prostate cancer	89

* HuGE, Human Genome Epidemiology; acyl-CoA, acyl coenzyme A; NAD(P)H, nicotinamide adenine dinucleotide phosphate (reduced form).

1), many of which were in the *Journal*. Five of these were concerned with gene variants associated with a high risk of disease (16–20). The other 15 have concerned common complex disorders of childhood or adult life. A further 30 reviews are in preparation (21). We now consider what we have learned from these reviews.

PUBLIC HEALTH APPLICATIONS

In virtually all of the reviews, it was concluded that there was no clear immediate public health application of the data. However, several of the reviews highlighted gene-disease associations for which public health applications are being considered. For example, one review dealt with sickle cell disease, for which an intervention has been established following a randomized trial that showed that oral penicillin could significantly reduce the associated morbidity and mortality (22). The substantial differences in mortality due to sickle cell disease that were demonstrated may reflect differences in the timing of introduction and extent of coverage of newborn screening and differences in medical care, parental education, and penicillin prophylaxis to prevent infections (17). Another review considered *HLA-DQ* and type 1 diabetes (23), as well as the weight of evidence that has led to *HLA-DQ* screening for type 1 diabetes being conducted in high-risk families and the general population for intervention trials and natural history studies. The review also highlighted a critical need to reconsider the risks, benefits, and ethical, legal, and social issues regarding genetic or autoantibody testing for type 1 diabetes, as well as a need to clarify the effects of environmental exposures as independent or interacting with high-risk *HLA* genotypes. In the HuGE review of hereditary hemachromatosis, it was concluded that more information is needed about penetrance of clinical expression among persons with elevated transferrin saturation or *HFE* mutations, about the disease burden associated with hereditary hemachromatosis in the general population, about screening accuracy, and about the diagnostic tests available and the efficacy of early treatment (24). For medium chain acyl coenzyme A dehydrogenase deficiency, the main knowledge gap concerns the natural history of the disease and its clinical outcomes (16). In regard to mismatch repair genes and colorectal cancer, there is no consensus regarding the most efficient approach of identifying mutation carriers (20). Some of the other reviews (25–33) dealt with gene variants that are part of a number of biomarkers included in test kits being marketed commercially (34), and these reviews highlighted important gaps in the evidence base. In future reviews, we encourage authors to emphasize data gaps and make recommendations for research to address these gaps.

VARIATION IN MANIFESTATION

In all of the reviews dealing with gene variants associated with a high risk of disease, variable penetrance or manifestation was noted. This reinforces the point that even for single gene disorders there is wide variation in clinical phenotype (18), and for this reason HuGE reviews of these disorders are valuable. In all of the reviews, there was a lack of data on other factors contributing to variation in manifestation.

METHODOLOGICAL ISSUES

The reviews highlight methodological issues such as selection bias, statistical power, and investigation of interaction or modifying factors, and they uncovered a need for unified guidelines that can be used to synthesize results of the increasing number of such studies. Progress is being made in defining quality standards for genetic-epidemiologic research, but ongoing evaluation is needed to make sure that such guidelines are refined and implemented. In 2001, an expert panel sponsored by the Centers for Disease Control and Prevention and the National Institutes of Health developed guidelines and recommendations for the reporting, evaluation, and integration of data from human genome epidemiology with emphasis on studies of 1) prevalence of gene variants and gene-disease associations, 2) gene-environment and gene-gene interactions, and 3) evaluation of genetic tests. Conclusions and recommendations from this workshop have been published (35, 36). In addition, other groups have proposed guidelines for gene-disease association studies (37–42). Many of the recommendations are similar, and the use of these guidelines in reporting studies should facilitate the integration of evidence in the future. Similarly, there is increasing interest in standardized approaches to the evaluation of genetic tests (43–45).

QUANTITATIVE SYNTHESIS

The use of meta-analysis or pooled analysis as a tool to synthesize evidence has been left to the discretion of the authors of HuGE reviews, in part because of concern about the lack of comparability of study methods and in part because of concern about the validity of meta-analysis of observational studies (46, 47). Meta-analysis was used as a tool for synthesizing evidence in two of the reviews (25, 32). In the future, with the application of guidelines for reporting human genome epidemiology studies, more comparability among published data will make meta-analysis a more feasible option. For the present, as the potential value of using meta-analysis is likely to vary between different gene-disease associations, we prefer to leave this decision to the authors of reviews. In one of the reviews, pooled analysis (which requires data on individual subjects) was used in addition to meta-analysis (32). Interestingly, the results of the pooled and meta-analyses were very similar. Pooled analyses require much greater resources than meta-analyses (48, 49) and would be preferred to meta-analysis only when a high degree of precision of the measures of effect is required. For example, as data on the penetrance of *HFE* mutations accumulate, a pooled analysis might be of considerable value.

REPLICATION

More generally, there has been considerable concern about nonreplication of gene-disease association studies (37, 38, 42, 50–52). Nonreplication has also been an issue in other areas of epidemiologic research, so much so that epidemiology has been occasionally viewed as having reached its limits (53); for example, the results of recent cohort studies are challenging the inverse association between cancer and consumption of vegetables and fruit (54–58). The investigation of gene-disease associations differs from the investigation of exposure-disease associations in two important respects. First, the assessment of genotypes by DNA assays (polymerase chain reaction methods) is generally more accurate than for exposure assessment, and it is less heavily dependent on study design. Second, because of "Mendelian randomization" (59), an association between a disease and a genotype is unlikely to be due to confounding, provided that the study is designed according to the principles of population-based studies (60). Although there has been concern about population stratification (36, 61, 62), empirical studies in non-Hispanic White Americans and modeling suggest that bias from this source may not be substantial when epidemiologic principles of study design, conduct, and analysis are rigorously applied (63, 64). In this context, it is interesting that, in an analysis of 301 published studies covering 25 associations in which the first positive report was excluded, grouping studies by ethnicity generally did not remove heterogeneity (65). In the same meta-analysis, there was an excess of studies replicating the initial report that seemed unlikely to be due to publication bias (65). For eight of the associations, the combined estimate of relative risk was statistically significant; this proportion is similar to the findings of another set of meta-analyses (51). Thus, it is possible that, as an area of investigation matures with a move from small innovative studies which might best be viewed as pilot studies to large well-resourced studies in which potential biases are minimized, more consistent associations will be observed than predicted by the rather bleak commentaries based on early studies.

This raises the challenge of keeping overviews of evidence up-to-date. In the early stages of an area of investigation, publication bias may be of critical importance (51, 52), as suggested for example by the pattern of accrual of evidence regarding the association between the angiotensin-converting-enzyme insertion/deletion polymorphism and myocardial infarction (66). Differences in timing may account for some differences between the results of meta-analyses as evidence accrues. Later, publication bias may be less of an issue as large high-quality studies are likely to be published irrespective of their findings. The best solution to the problem of publication bias appears to be the establishment of a research register for studies of gene-disease associations and of gene-gene and gene-environment interactions, analogous to those for other areas of medicine (67–69). This would help to address the problem of integrating all available evidence (70), taking into account its quality.

TABLE 2. Proposed criteria for prioritizing HuGE* review topics

Public health significance of the disease (in terms of morbidity and mortality)
Availability of effective interventions for genes modulating, or thought to modulate, an exposure
Effect on pathways involved in pathogenesis of multiple diseases of public health significance (e.g., methylation, DNA repair)
Relevance to common disease with evidence of gene-environment or gene-gene interactions
High potential population attributable risk, on the basis of at least two studies

* HuGE, Human Genome Epidemiology.

VOLUME AND TYPE OF EVIDENCE

There is also the challenge of the ever-increasing number of human genome epidemiology studies. For example, in the literature database maintained in the Centers for Disease Control and Prevention Genomics and Disease Prevention Information System (71), 2,436 primary studies of this type were published in 2001, and 2,922 studies were published in 2002. Moreover, as a result of the increasing availability of mapped single nucleotide polymorphism markers (72, 73), this trend is expected to accelerate. Therefore, integration of evidence will become increasingly important as a means of dealing with potentially unmanageable amounts of information. Certainly, the Human Genome Epidemiology Network (HuGE Net) will continue to benefit from the contributions of researchers writing HuGE reviews in their own specialty areas. However, we would also like to suggest some priorities with the hope of encouraging others to invest effort in integrating evidence about the gene-disease associations (and related gene-gene and gene-environment interactions) most likely to expand our knowledge and ability to apply research results. Some proposed criteria for prioritizing HuGE reviews are presented in table 2.

The type of evidence is also relevant. An analysis of abstracts of published human genome epidemiologic papers for 2001–2002 shows that, of the 5,358 published articles, 601 (11.2 percent) reported only on the population prevalence of gene variants, 4,657 (86.9 percent) reported on gene-disease associations, 978 (18.3 percent) reported on gene-gene and gene-environment interactions, and 173 (3.2 percent) dealt with evaluation of genetic tests and population screening (71). Much of the evidence on the population prevalence of genetic variants in HuGE reviews has been derived from data on controls in gene-disease association studies. We recognize that assembling these data for the purposes of a HuGE review may be very labor intensive.

There is increasing emphasis on the importance of biologic data, in particular on gene function and gene expression, in the interpretation of gene-disease associations and gene-gene and gene-environment interactions (74–77). As in the use of biologic data in making causal inference in other contexts, caution is warranted (38). Molecular biologic research has enjoyed explosive development, and it is difficult to identify and organize the information that would be useful in considering the biologic plausibility of an association or putative interaction (78). In regard to the quality of such information, it has been noted, for example, that the lack of standard methodologies or nomenclature for DNA expression studies has made it difficult to compare results (79, 80); recommendations for standardization have now been made (80). We anticipate that development of methods for the synthesis of biologic data will enhance the understanding of the functional effects of gene variants, particularly of multiple genes operating in pathways and networks, and that it will be relevant to consider this evidence in future HuGE reviews.

CALL FOR REVIEWS

In recognition of the increasing volume of evidence and the distribution of types of evidence, we will now propose two additional categories of HuGE review: 1) reviews of gene-disease associations and related gene-environment or gene-gene interactions only and 2) reviews of genotype prevalence only. The relations between these and the existing formats for HuGE reviews to be published by the *Journal* are summarized in table 3; instructions for these are presented on the HuGE Net website (21).

There has been a change in the emphasis of research on genetic susceptibility from single candidate genes to multiple genes operating in pathways and, indeed, networks and systems (81, 82). For example, because the substrates resulting from phase 1 activation may be more reactive and potentially more carcinogenic than the starting xenobiotic compound, coordinated expression of phase 1 and phase 2 genes is likely to be critical in the metabolism of xenobiotic compounds (83). Moreover, many xenobiotic compounds can be metabolized by more than one cytochrome P450 enzyme (84). Similarly, there are genetic polymorphisms of several key proteins involved in folate metabolism (85). Therefore, we encourage authors to submit reviews to the *Journal* involving more than one gene operating in a pathway.

Overall, the task of characterizing the human genome is at the beginning. The concern that the potential value of this exercise to public health has been exaggerated or that the amount of time needed for information relevant to public health to be accrued has been underestimated underlines, more than ever, the need for integration of evidence from carefully conducted population-based studies. We hope that the suggested changes to HuGE reviews will stimulate more investigators to contribute to this task.

TABLE 3. Formats of HuGE* review†

Format	Content
Full review	Information on gene(s), variants of gene(s) (defined, effect on function if known, and variation in genotype frequencies), disease(s), associations with disease(s), interactions, laboratory tests, population testing, other potential public health applications (e.g., setting permissible exposure thresholds), conclusions and recommendations for research,‡ references, Internet sites
Gene-disease association review	Similar to above, except that no information on the variation in genotype frequencies presented
Minireview	This is appropriate when the epidemiologic aspects of specific gene(s) have already been reviewed for HuGE Net,* but the associations between the gene and a different disease are being reviewed. In the section on gene variants, a summary (with reference) of the points covered in the full review(s) relating to this gene should be presented
Prevalence review	Information on gene(s), variants of gene(s) (definition and variation in genotype frequency), laboratory tests, conclusions and recommendations for research, references, Internet sites

* HuGE, Human Genome Epidemiology; HuGE Net, Human Genome Epidemiology Network.
† Detailed instructions are available at http://www.cdc.gov/genomics/hugenet/reviews/guidelines.htm.
‡ This was not requested explicitly in the previous guidelines (15).

REFERENCES

1. Shpilberg O, Dorman JS, Ferrell RE, et al. The next stage: molecular epidemiology. J Clin Epidemiol 1997;50:633–8.
2. Watson JD, Crick FHC. Molecular structure of nucleic acids. A structure for deoxyribose nucleic acid. Nature 1953;171:737–8.
3. Bell J. The new genetics in clinical practice. BMJ 1998;316:618–20.
4. Collins FS. Shattuck Lecture—medical and societal consequences of the Human Genome Project. N Engl J Med 1999;341:28–37.
5. Collins FS, McKusick VA. Implications of the Human Genome Project for medical science. JAMA 2001;285:540–4.
6. Guttmacher AE, Collins FS. Genomic medicine—a primer. N Engl J Med 2002;347:1512–20.
7. Khoury MJ, Little J, Burke W. Human genome epidemiology: scope and strategies. In: Khoury MJ, Little J, Burke W, eds. Human genome epidemiology. Scientific foundation for using genetic information to improve health and prevent disease. New York, NY: Oxford University Press, 2003.
8. Holtzman NA, Marteau TM. Will genetics revolutionize medicine? N Engl J Med 2000;343:141–4.
9. Baird P. The Human Genome Project, genetics, and health. Community Genet 2001;4:77–80.
10. Evans JP, Skrzynia C, Burke W. The complexities of predictive genetic testing. BMJ 2001;322:1052–6.
11. Vineis P, Schulte P, McMichael AJ. Misconceptions about the use of genetic tests in populations. Lancet 2001;357:709–12.
12. Zimmern R, Emery J, Richards T. Putting genetics in perspective. BMJ 2001;322:1005–6.
13. Willett WC. Diet and cancer: one view at the start of the millennium. Cancer Epidemiol Biomarkers Prev 2001;10:3–8.
14. Khoury MJ, Dorman JS. The Human Genome Epidemiology Network. Am J Epidemiol 1998;148:1–3.
15. Revised guidelines for submitting HuGE reviews. Am J Epidemiol 2000;151:4–6.
16. Wang SS, Fernhoff PM, Hannon WH, et al. Medium chain acyl-CoA dehydrogenase deficiency human genome epidemiology review. Genet Med 1999;1:332–9.
17. Ashley-Koch A, Yang Q, Olney RS. Sickle hemoglobin (*HbS*) allele and sickle cell disease: a HuGE review. Am J Epidemiol 2000;151:839–45.
18. Rasmussen SA, Friedman JM. *NF1* gene and neurofibromatosis 1. Am J Epidemiol 2000;151:33–40.
19. Crawford DC, Acuna JM, Sherman SL. *FMR1* and the fragile X syndrome: human genome epidemiology review. Genet Med 2001;3:359–71.
20. Mitchell RJ, Farrington SM, Dunlop MG, et al. Mismatch repair genes *hMLH1* and *hMSH2* and colorectal cancer: a HuGE review. Am J Epidemiol 2002;156:885–902.
21. HuGE Net (Human Genome Epidemiology Network). HuGE reviews. Atlanta, GA: Centers for Disease Control and Prevention, 2003. (http://www.cdc.gov/genomics/hugenet/reviews.htm).
22. Gaston MH, Verter JI, Woods G, et al. Prophylaxis with oral penicillin in children with sickle cell anemia. A randomized trial. N Engl J Med 1986;314:1593–9.
23. Dorman JS, Bunker CH. *HLA-DQ* locus of the human leukocyte antigen complex and type 1 diabetes mellitus: a HuGE review. Epidemiol Rev 2000;22:218–27.
24. Hanson EH, Imperatore G, Burke W. *HFE* gene and hereditary hemochromatosis: a HuGE review. Am J Epidemiol 2001;154:193–206.
25. Botto LD, Yang Q. 5,10-Methylenetetrahydrofolate reductase gene variants and congenital anomalies: a HuGE review. Am J Epidemiol 2000;151:862–77.
26. Brockton N, Little J, Sharp L, et al. *N*-Acetyltransferase polymorphisms and colorectal cancer: a HuGE review. Am J Epidemiol 2000;151:846–61.
27. Cotton SC, Sharp L, Little J, et al. Glutathione *S*-transferase polymorphisms and colorectal cancer: a HuGE review. Am J Epidemiol 2000;151:7–32.
28. Zmuda JM, Cauley JA, Ferrell RE. Molecular epidemiology of vitamin D receptor gene variants. Epidemiol Rev 2000;22:203–17.
29. Geisler SA, Olshan AF. *GSTM1*, *GSTT1*, and the risk of squamous cell carcinoma of the head and neck: a mini-HuGE review. Am J Epidemiol 2001;154:95–105.
30. Coughlin SS, Hall IJ. Glutathione *S*-transferase polymorphisms and risk of ovarian cancer: a HuGE review. Genet Med 2002;4:250–7.
31. Eichner JE, Dunn ST, Perveen G, et al. Apolipoprotein E polymorphism and cardiovascular disease: a HuGE review. Am J Epidemiol 2002;155:487–95.
32. Engel LS, Taioli E, Pfeiffer R, et al. Pooled analysis and meta-analysis of glutathione *S*-transferase M1 and bladder cancer: a HuGE review. Am J Epidemiol 2002;156:95–109.
33. Robien K, Ulrich CM. 5,10-Methylenetetrahydrofolate reductase polymorphisms and leukemia risk: a HuGE minireview. Am J Epidemiol 2003;157:571–82.
34. What's brewing in genetic testing. Nat Genet 2002;32:553–4.

35. Burke W, Atkins D, Gwinn M, et al. Genetic test evaluation: information needs of clinicians, policy makers, and the public. Am J Epidemiol 2002;156:311–18.
36. Little J, Bradley L, Bray MS, et al. Reporting, appraising, and integrating data on genotype prevalence and gene-disease associations. Am J Epidemiol 2002;156:300–10.
37. Freely associating. Nat Genet 1999;22:1–2.
38. Cardon LR, Bell JI. Association study designs for complex diseases. Nat Rev Genet 2001;2:91–9.
39. Weiss ST. Association studies in asthma genetics. Am J Respir Crit Care Med 2001;164:2014–15.
40. Cooper DN, Nussbaum RL, Krawczak M. Proposed guidelines for papers describing DNA polymorphism-disease associations. Hum Genet 2002;110:207–8.
41. Romero R, Kuivaniemi H, Tromp G, et al. The design, execution, and interpretation of genetic association studies to decipher complex diseases. Am J Obstet Gynecol 2002;187:1299–312.
42. Tabor HK, Risch NJ, Myers RM. Opinion: candidate-gene approaches for studying complex genetic traits: practical considerations. Nat Rev Genet 2002;3:391–7.
43. Holtzman NA, Watson MS, eds. Promoting safe and effective genetic testing in the United States. Final report of the Task Force on Genetic Testing. Baltimore, MD: Johns Hopkins University Press, 1997. (http://www.nhgri.nih.gov/ELSI/TFGT_final/).
44. Secretary's Advisory Committee on Genetic Testing. Enhancing the oversight of genetic tests: recommendations of the SACGT. Bethesda, MD: National Institutes of Health, 2000. (http://www4.od.nih.gov/oba/sacgt/reports/oversight_report.pdf).
45. Haddow JE, Palomaki GE. ACCE: a model process for evaluating data on emerging genetic tests. In: Khoury MJ, Little J, Burke W, eds. Human genome epidemiology. Scientific foundation for using genetic information to improve health and prevent disease. New York, NY: Oxford University Press, 2003.
46. Blettner M, Sauerbrei W, Schlehofer B, et al. Traditional reviews, meta-analyses and pooled analyses in epidemiology. Int J Epidemiol 1999;28:1–9.
47. Stroup DF, Berlin JA, Morton SC, et al. Meta-analysis of observational studies in epidemiology: a proposal for reporting. JAMA 2000;283:2008–12.
48. Steinberg KK, Smith SJ, Stroup DF, et al. Comparison of effect estimates from a meta-analysis of summary data from published studies and from a meta-analysis using individual patient data for ovarian cancer studies. Am J Epidemiol 1997;145:917–25.
49. Ioannidis JP, Rosenberg PS, Goedert JJ, et al. Commentary: meta-analysis of individual participants' data in genetic epidemiology. Am J Epidemiol 2002;156:204–10.
50. Gambaro G, Anglani F, D'Angelo A. Association studies of genetic polymorphisms and complex disease. Lancet 2000;355:308–11.
51. Ioannidis JP, Ntzani EE, Trikalinos TA, et al. Replication validity of genetic association studies. Nat Genet 2001;29:306–9.
52. Hirschhorn JN, Lohmueller K, Byrne E, et al. A comprehensive review of genetic association studies. Genet Med 2002;4:45–61.
53. Taubes G. Epidemiology faces its limits. Science 1995;269:164–9.
54. Feskanich D, Ziegler RG, Michaud DS, et al. Prospective study of fruit and vegetable consumption and risk of lung cancer among men and women. J Natl Cancer Inst 2000;92:1812–23.
55. Michels KB, Giovannucci E, Joshipura KJ, et al. Prospective study of fruit and vegetable consumption and incidence of colon and rectal cancers. J Natl Cancer Inst 2000;92:1740–52.
56. Voorrips LE, Goldbohm RA, van Poppel G, et al. Vegetable and fruit consumption and risks of colon and rectal cancer in a prospective cohort study: the Netherlands Cohort Study on Diet and Cancer. Am J Epidemiol 2000;152:1081–92.
57. Smith-Warner SA, Spiegelman D, Yaun SS, et al. Intake of fruits and vegetables and risk of breast cancer: a pooled analysis of cohort studies. JAMA 2001;285:769–76.
58. Flood A, Velie EM, Chaterjee N, et al. Fruit and vegetable intakes and the risk of colorectal cancer in the Breast Cancer Detection Demonstration Project follow-up cohort. Am J Clin Nutr 2002;75:936–43.
59. Youngman LD, Keavney BD, Palmer A, et al. Plasma fibrinogen and fibrinogen genotypes in 4685 cases of myocardial infarction and in 6002 controls: test of causality by "Mendelian randomization." Circulation 2000;102(suppl II):31–2.
60. Clayton D, McKeigue PM. Epidemiological methods for studying genes and environmental factors in complex diseases. Lancet 2001;358:1356–60.
61. Thomas DC, Witte JS. Point: population stratification: a problem for case-control studies of candidate-gene associations? Cancer Epidemiol Biomarkers Prev 2002;11:505–12.
62. Wacholder S, Rothman N, Caporaso N. Counterpoint: bias from population stratification is not a major threat to the validity of conclusions from epidemiological studies of common polymorphisms and cancer. Cancer Epidemiol Biomarkers Prev 2002;11:513–20.
63. Wacholder S, Rothman N, Caporaso N. Population stratification in epidemiologic studies of common genetic variants and cancer: quantification of bias. J Natl Cancer Inst 2000;92:1151–8.
64. Ardlie KG, Lunetta KL, Seielstad M. Testing for population subdivision and association in four case-control studies. Am J Hum Genet 2002;71:304–11.
65. Lohmueller KE, Pearce CL, Pike M, et al. Meta-analysis of genetic association studies supports a contribution of common variants to susceptibility to common disease. Nat Genet 2003;33:177–82.
66. Keavney B, McKenzie C, Parish S, et al. Large-scale test of hypothesised associations between the angiotensin-converting-enzyme insertion/deletion polymorphism and myocardial infarction in about 5000 cases and 6000 controls. International Studies of Infarct Survival (ISIS) Collaborators. Lancet 2000;355:434–42.
67. Chalmers I, Sackett D, Silagg C. The Cochrane collaboration. London, United Kingdom: British Medical Journal Books, 1997:231–49.
68. CRISP. Computer retrieval of information on scientific projects. Bethesda, MD: National Institutes of Health, 2003. (http://crisp.cit.nih.gov).
69. Sankaranarayannan R, Becker N, Démaret E. Directory of ongoing research in cancer prevention. Lyon, France: International Agency for Research on Cancer, 2000. (http://www-dep.iarc.fr/direct/prevent.htm).
70. In search of genetic precision. Lancet 2003;361:357.
71. GDP Info. CDC Genomics and Disease Prevention Information System. Atlanta, GA: Office of Genomics and Disease Prevention, Centers for Disease Control and Prevention, 2003. (http://www2.cdc.gov/nceh/genetics/GDPQueryTool/default.asp).
72. Reich DE, Cargill M, Bolk S, et al. Linkage disequilibrium in the human genome. Nature 2001;411:199–204.
73. Sachidanandam R, Weissman D, Schmidt SC, et al. A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. Nature 2001;409:928–33.
74. Weiss KM, Terwilliger JD. How many diseases does it take to map a gene with SNPs? Nat Genet 2000;26:151–7.
75. Challenges for the 21st century. Nat Genet 2001;29:353–4.
76. Cloninger CR. The discovery of susceptibility genes for mental

disorders. Proc Natl Acad Sci U S A 2002;99:13365–7.
77. Glazier AM, Nadeau JH, Aitman TJ. Finding genes that underlie complex traits. Science 2002;298:2345–9.
78. Friedman C, Kra P, Yu H, et al. GENIES: a natural-language processing system for the extraction of molecular pathways from journal articles. Bioinformatics 2001;17(suppl 1):S74–82.
79. Perou CM. Show me the data! Nat Genet 2001;29:373.
80. Coming to terms with microarrays. Nat Genet 2002;32:333–4.
81. Duyk GM. Sharper tools and simpler methods. Nat Genet 2002; 32(suppl):465–8.
82. Strohman R. Maneuvering in the complex path from genotype to phenotype. Science 2002;296:701–3.
83. Fryer AA, Jones PW. Interactions between detoxifying enzyme polymorphisms and susceptibility to cancer. In: Vineis P, Malats N, Lang M, et al, eds. Metabolic polymorphisms and susceptibility to cancer. Lyon, France: International Agency for Research on Cancer, 1999:303–22.
84. Caporaso NE. Why have we failed to find the low penetrance genetic constituents of common cancers? Cancer Epidemiol Biomarkers Prev 2002;11:1544–9.
85. Ulrich CM, Robien K, Sparks R. Pharmacogenetics and folate metabolism—a promising direction. Pharmacogenetics 2002;3: 299–313.
86. Nebert DW, Roe AL, Vandale SE, et al. NAD(P)H:quinone oxidoreductase (*NQO1*) polymorphism, exposure to benzene, and predisposition to disease: a HuGE review. Genet Med 2002;4:62–70.
87. Kelada SN, Shelton E, Kaufmann RB, et al. Delta-aminolevulinic acid dehydratase genotype and lead toxicity: a HuGE review. Am J Epidemiol 2001;154:1–13.
88. Kenneson A, Van Naarden Braun K, Boyle C. *GJB2* (connexin 26) variants and nonsyndromic sensorineural hearing loss: a HuGE review. Genet Med 2002;4:258–74.
89. Nelson KA, Witte JS. Androgen receptor CAG repeats and prostate cancer. Am J Epidemiol 2002;155:883–90.