

Chapter 2

Concepts of Genetic Epidemiology

Kathleen Ries Merikangas

Abstract The major aim of this chapter is to provide an overview of the field of genetic epidemiology and its relevance to the identification of the causes and risk factors for human diseases. The most important goal of the methods of genetic epidemiology is to elucidate the joint contribution of genes and environmental exposures to the etiology of complex diseases. The key study designs used to achieve this goal including family, twin, adoption, and migration studies are summarized. The field of genetic epidemiology is expected to have increasing importance with advances in molecular genetics.

Keywords Genetics • Epidemiology • Family studies • Twin studies • Adoption studies • Migration studies

2.1 Introduction: Genetic Epidemiology

Genetic epidemiology is defined as the study of the distribution of and risk factors for diseases and genetic and environmental causes of familial resemblance. Genetic epidemiology focuses on how genetic factors and their interactions with other risk factors increase vulnerability to, or protection against, disease (Beaty 1997). Genetic epidemiology employs traditional epidemiologic study designs to explain aggregation in groups as closely related as twins or as loosely related as migrant cohorts. Epidemiology has developed sophisticated designs and analytic methods for identifying

K.R. Merikangas, Ph.D. (✉)

Genetic Epidemiology Research Branch, Intramural Research Program,
National Institute of Mental Health, National Institutes of Health,
35 Convent Drive, MSC#3720, Bethesda, MD 20892, USA
e-mail: Kathleen.merikangas@nih.gov

disease risk factors. With increasing progress in gene identification, these methods have been extended to include both genetic and environmental factors (MacMahon and Trichopoulos 1996; Kuller 1979). In general, study designs in genetic epidemiology either control for genetic background while letting the environment vary (e.g., migrant studies, half siblings, separated twins) or control for the environment while allowing variance in the genetic background (e.g., siblings, twins, adoptees/nonbiological siblings). Investigations in genetic epidemiology are typically based on a combination of study designs including family, twin, and adoption studies.

2.1.1 Family Studies

Familial aggregation is generally the first source of evidence that genetic factors may play a role in a disorder. The most common indicator of familial aggregation is the relative risk ratio, computed as the rate of a disorder in families of affected persons divided by the corresponding rate in families of controls. The patterns of genetic factors underlying a disorder can be inferred from the extent to which patterns of familial resemblance adhere to the expectations of Mendelian laws of inheritance. The degree of genetic relatedness among relatives is based on the proportion of shared genes between a particular relative and an index family member or proband. First-degree relatives share 50% of their genes in common, second-degree relatives share 25% of their genes in common, and third-degree relatives share 12.5% of their genes in common. If familial resemblance is wholly attributable to genes, there should be a 50% decrease in disease risk with each successive increase in degree of relatedness, from first to second to third and so forth. This information can be used to derive estimates of familial recurrence risk within and across generations as a function of population prevalence (λ) (Risch 1990b). Whereas λ tends to exceed 20 for most autosomal dominant diseases, values of λ derived from family studies of many complex disorders tend to range from 2 to 5. Diseases with strong genetic contributions tend to be characterized by 50% decrease in risk across successive generations. Decrease in risk according to the degree of genetic relatedness can also be examined to detect interactions between several loci. If the risk to second- and third-degree relatives decreases by more than 50%, this implies that more than a single locus must contribute to disease risk and that no single locus can largely predominate.

The major advantage of studying diseases within families is that disease manifestations are more likely to result within families than they are across families from the same underlying etiologic factors. Family studies are therefore more effective than between family designs in examining the validity of diagnostic categories because they more accurately assess the specificity of transmission of symptom patterns and disorders. Data from family studies can also provide evidence regarding etiologic or phenotypic heterogeneity. Phenotypic heterogeneity is suggested by variable expressivity of symptoms of the same underlying risk factors, whereas etiologic heterogeneity is demonstrated by common manifestations of expression of different etiologic factors between families. Moreover, the family study method

permits assessment of associations between disorders by evaluating specific patterns of co-segregation of two or more disorders within families (Merikangas 1990).

2.1.2 Twin Studies

Twin studies that compare concordance rates for monozygotic twins (who share the same genotype) with those of dizygotic twins (who share an average of 50% of their genes) provide estimates of the degree to which genetic factors contribute to the etiology of a disease phenotype. A crude estimate of the genetic contribution to risk for a disorder is calculated by doubling the difference between the concordance rates for monozygous and dizygous twin pairs. Modern genetic studies employ path analytic models to estimate the proportion of variance attributable to additive genes, common environment, and unique environment. There are several other applications of the twin study design that may inform our understanding of the roles of genetic and environmental risk factors for disease. First, twin studies provide information on the genetic and environmental sources of sex differences in a disease. Second, environmental exposures may be identified through comparison of discordant monozygotic twins. Third, twin studies can be used to identify the genetic mode of transmission of a disease by inspection of the degree of adherence of the difference in risk between monozygotic and dizygotic twins to the Mendelian ratio of 50%. Fourth, twin studies may contribute to enhancing the validity of a disease through inspection of the components of the phenotypes that are most heritable. The twin family design is one of the most powerful study designs in genetic epidemiology because it yields estimates of heritability but also permits evaluation of multigenerational patterns of expression of genetic and environmental risk factors.

2.1.3 Adoption Studies

Adoption studies have been the major source of evidence regarding the joint contribution of genetic and environmental factors to disease etiology. Adoption studies either compare the similarity between an adoptee and his or her biological versus adoptive relatives or the similarity between biological relatives of affected adoptees with those of unaffected or control adoptees. The latter approach is more powerful because it eliminates the potentially confounding effect of environmental factors. Similar to the familial recurrence risk, the genetic contribution in adoption studies is estimated by comparing the risk of disease to biological versus adoptive relatives or the risk of disease in biological relatives of affected versus control adoptees. These estimates of risk are often adjusted for sex, age, ethnicity, and other factors that may confound the links between adoption status and an index disease.

With the recent trends toward selective adoption and the diminishing frequency of adoptions in the USA, adoption studies are becoming less feasible methods for identifying genetic and environmental sources of disease etiology (National

Adoption Information Clearinghouse 2007). However, the increased rate of reconstituted families (families comprised of both siblings and half siblings) may offer a new way to evaluate the role of genetic factors in the transmission of complex disorders. Genetic models predict that half siblings should have a 50% reduction in disease risk compared to that of full siblings. Deviations from this risk provide evidence for either polygenic transmission, gene-environment interaction, or other complex modes of transmission.

2.1.4 Migration Studies

Migrant studies are perhaps the most powerful study design to identify environmental and cultural risk factors. When used to study Asian immigrants to the USA, this study design demonstrated the significant contribution of the environment to the development of many forms of cancer and heart disease (Kolonel et al. 2004). One of the earliest controlled migrant studies evaluated rates of psychosis among Norwegian immigrants to Minnesota compared to native Minnesotans and native Norwegians (Ödegaard 1932). A higher rate of psychosis was found among the immigrants compared to both the native Minnesotans and Norwegians and was attributed to increased susceptibility to psychosis among the migrants who left Norway. It was found that migration selection bias was the major explanatory factor, rather than environmental exposure in the new culture. The application of migration studies to the identification of environmental factors is only valid if potential bias attributed to selection is considered. Selection bias has been tested through comparisons of factors that may influence a particular disease of interest in a migrant sample and a similar sample that did not migrate.

2.2 Applications of Genetic Epidemiology to Gene Identification

There is a widespread consensus among geneticists and epidemiologists on the importance of epidemiology to the future of genetics and on the conclusion that the best strategy for susceptibility risk factor identification for common and complex disorders will ultimately involve large epidemiologic studies from diverse populations (Peltonen and McKusick 2001; Khoury and Little 2000; Yang and Khoury 1997; Merikangas 2003; Merikangas and Risch 2003; Risch 1990a). It is likely that population-based association studies will assume increasing importance in translating the products of genomics to public health. There are several reasons that population-based studies are critical to current studies seeking to identify genes underlying complex disorders. First, the frequency of newly identified polymorphisms, whether SNPs or other variants such as copy number variations (CNVs),

especially in particular population subgroups, is not known. Second, current knowledge of genes as risk factors is based nearly exclusively on clinical and nonsystematic samples. Hence, the significance of the susceptibility alleles that have been identified for cancer, heart disease, diabetes, and other common disorders is unknown in the population at large. In order to provide accurate risk estimates, the next stage of research needs to move beyond samples identified through affected individuals to the population as a whole. Third, identification of risk profiles will require large samples to assess the significance of vulnerability genes with relatively low expected population frequencies. Fourth, similar to the role of epidemiology in quantifying risk associated with traditional disease risk factors, applications of human genome epidemiology can provide information on the specificity, sensitivity, and impact of genetic tests to inform science and individuals (Khoury and Little 2000).

2.2.1 Samples

The shift from systematic large-scale family studies to linkage studies has led to the collection of families according to very specific sampling strategies (e.g., many affected relatives, affected sibling pairs, affected relatives on one side of the family only, and availability of parents for study) in order to maximize the power of detecting genes according to the assumed model of familial transmission. Despite the increase in power for detecting genes, these sampling approaches have diminished the generalizability of the study findings and contribute little else to the knowledge base if genes are not discovered. Future studies will attempt to collect both families and controls from representative samples of the population so that results can be used to estimate population risk parameters and to examine the specificity of endophenotypic transmission and so results can be generalized to whole populations.

2.2.2 Selection of Controls

The most serious problem in the design of association studies is the difficulty of selecting controls that are comparable to the cases on all factors except the disease of interest (Wacholder et al. 2000; Ott 2004). Controls should be drawn from the same population as cases and must have the same probability of exposure (i.e., genes) as cases. Controls should be selected to ensure the validity rather than the representativeness of a study. Failure to equate cases and controls may lead to confounding (i.e., a spurious association due to an unmeasured factor that is associated with both the candidate gene and the disease). In genetic case-control studies, the most likely source of confounding is ethnicity because of differential gene and disease frequencies in different ethnic subgroups. The matching of controls to cases on ethnic background is largely based on self-report; several methods are used to screen for and exclude subjects with substantial differences in ancestry.

2.2.3 Risk Estimation

Because genetic polymorphisms involved in complex diseases are likely to be nondeterministic (i.e., the marker neither predicts disease nor non-disease with certainty), traditional epidemiologic risk factor designs can be used to estimate the impact of these genetic polymorphisms. Increased attention to alleles as a part of risk equations in epidemiology will likely resolve the contradictory findings from studies that have generally employed solely environmental risk factors, such as diet, smoking, and alcohol use. Likewise, the studies that seek solely to identify small risk alleles will continue to be inconsistent because they do not consider the effects of nongenetic biological parameters or environmental factors that contribute to the diseases of interest.

There are several types of risk estimates that are used in public health. The most common is *relative risk*, defined as the magnitude of the association between an exposure and disease. It is independent of the prevalence of the exposure. The *absolute risk* is the overall probability of developing a disease in an individual or in a particular population (Gordis 2000). The *attributable risk* is the difference in the risk of the disease in those exposed to a particular risk factor compared to the background risk of a disease in a population (i.e., in the unexposed). The *population attributable risk* relates to the risk of a disease in a total population (exposed and unexposed) and indicates the amount the disease can be reduced in a population if an exposure is eliminated. The population attributable risk depends on the prevalence of the exposure or, in the case of risk alleles, the allele frequency. Genetic attributable risk would indicate the proportion of a particular disease that would be eliminated if a particular gene or genes were not involved in the disease. For example, the two vulnerability alleles for Alzheimer's disease include the very rare but *deterministic alleles* in the β -amyloid precursor, presenilin-1, and presenilin-2 genes, which signal a very high probability of the development of Alzheimer's disease, particularly at a young age, and the *susceptibility* allele $\epsilon 4$ in the apolipoprotein-E gene (APOE $\epsilon 4$) (Tol et al. 1999). The apolipoprotein-E $\epsilon 4$ (APOE $\epsilon 4$) allele has been shown to increase the risk of Alzheimer's disease in a dose-dependent fashion. Using data from a large multiethnic sample collected by more than 40 research teams, Farrer (Farrer et al. 1997) reported a 2.6–3.2 greater odds of Alzheimer's disease among those with one copy and 14.9 odds of Alzheimer's disease among those with two copies of the APOE $\epsilon 4$ allele. Moreover, there was a significant protective effect among those with the $\epsilon 2/\epsilon 3$ genotype. As opposed to the deterministic mutations, the APOE $\epsilon 4$ allele has a very high population attributable risk because of its high frequency in the population. The APOE $\epsilon 4$ allele is likely to interact with environmental risk and protective factors (Kivipelto et al. 2001; Kivipelto et al. 2002). The population risk attributable to these mutations is quite low because of the very low population prevalence of disease associated with these alleles. This model of combination of several rare deterministic alleles in a small subset of families and common alleles with lower relative risk to individuals but high population attributable risk is likely to apply to many other complex diseases as well. Genome-wide

association studies have now identified genes for more than 300 diseases and traits, such as coronary artery disease, Crohn's disease, rheumatoid arthritis, and type 1 and type 2 diabetes (Wellcome Trust Case Control Consortium 2007) with 1,291 publications by the end of 2011 (www.genome.gov/gwastudies). Those genetic variants appear to confer only modest increases in disease risk (ORs between 1.2 and 1.5) compared to other established risk factors for common chronic diseases.

2.2.4 Identification of Environmental Factors

In parallel with the identification of susceptibility alleles, it is important to identify environmental factors that operate either specifically or nonspecifically on those with susceptibility to complex disorders in order to develop effective prevention and intervention efforts. Langholz et al. (1999) describe some of the world's prospective cohort studies that may serve as a basis for studies of gene-disease associations or gene-environment interactions. There is increasing evidence that gene-environment interaction will underlie many of the complex human diseases. Some examples include inborn errors of metabolism, individual variation in response to drugs (Nebert 1999), substance use disorders (Heath et al. 2001; Rose et al. 2000), and the protective influence of a deletion in the CCR5 gene on exposure to HIV (Michael 1999). In prospective studies, however, few environmental exposures have been shown to have an etiologic role in complex disorders (Eaton 2004). Over the next decades, it will be important to identify and evaluate the effects of specific environmental factors on disease outcomes and to refine measurement of environmental exposures to evaluate the specificity of effects. Study designs and statistical methods should focus increasingly on the nature of the relationships between genetic and environmental factors, particularly epistasis and gene-environment interaction (Yang and Khoury 1997; Ottman 1990; Beaty and Khoury 2000). For example, recent breakthroughs in identifying the mechanisms for hypocretin deficiency as the causal mechanism in narcolepsy occurred through a convergence of epidemiologic studies that documented a recent surge in incidence among those exposed to H1N1 virus or vaccine, successful application of genome-wide association studies that implicated specific autoimmune mechanisms (i.e., the T-cell receptor α polymorphism), and specificity of the findings for the phenotype of narcolepsy with cataplexy rather than narcolepsy alone (Kornum et al. 2011).

2.3 Applications, Impact, and Future Directions

The advances in bioinformatics and statistical methods described in the following chapters will be critical to translation of progress in molecular genetics to human diseases. Genetic epidemiologic approaches, particularly the family study design, will have renewed importance in facilitating integration between methodological

developments and human diseases. Despite the long history of information provided by family studies regarding the genetic architecture of Mendelian diseases as well as heterogeneity of complex diseases such as breast cancer (Claus et al. 1993) and diabetes (Hawa et al. 2002), the family study approach has largely been abandoned in psychiatry in favor of very large case-control studies of diagnosed patients from clinical samples or registries. Yet, family studies still have an essential role in identifying cross-generational transmission of phenotypes and genotypes. Family-based studies will be even more valuable with application of advances in molecular biology to inform interpretation of sequencing data and to distinguish *de novo* from heritable structural variants. Based on increasing awareness of the neglect of family studies for risk prediction, even in the absence of specification of disease genetic architecture, the US surgeon general has launched a national public health campaign to encourage all American families to learn more about their family health history (<http://www.hhs.gov/familyhistory/>). A positive family history remains a more potent predictor of disease vulnerability than nearly all other risk factors combined (Meigs et al. 2008). Moreover, since genetic factors, common environmental exposure, and sociocultural factors have been shown to jointly contribute to disease etiology, family history may ultimately have greater explanatory power than genes in predicting risk, particularly if genetic influences are weak.

Progress in genomics has far outstripped advances in our understanding of many of the complex multifactorial human disorders and their etiologies. Technical advances and availability of rapidly expanding genetic databases provide extraordinary opportunities for understanding disease pathogenesis. Over the next decade, increasing understanding of the complex mechanisms through which genetic risk factors influence disease should enhance the clinical utility of genetics. The above issues regarding sampling, complexity of the links between genes and environmental factors in multifactorially determined complex diseases, and phenotypic heterogeneity also highlight the complexity of etiology of complex human diseases. This work demonstrates that predictions that human genomics would lead to a radical transformation of medical practice were overly optimistic. In fact, Varmus (Varmus 2002) concluded that despite the journalistic hyperbole, the sequencing of the human genome is unlikely to lead either to a radical transformation of medical practice or even to an information-based science that can predict with certainty future diseases and effective treatment interventions. Therefore, despite the extraordinary opportunity for understanding disease pathogenesis afforded by the technical advances and availability of rapidly expanding genetic databases, it is unlikely that we will soon experience the light speed progress of genomics in understanding, treating, or preventing many of the multifactorial complex human diseases.

The chasm between genetic information and clinical utility should gradually close as we develop new methods and tools in human genetic and clinical research to maximize the knowledge afforded by the exciting advances in genomics. Increased integration of advances in basic sciences and genomics along with information from population-based studies and longitudinal cohorts; innovations in our conceptualizations of the disease etiology, particularly the role of infectious agents; and the identification of specific risk and protective factors will lead to more informed

intervention strategies. As we learn more about the role of genes as risk factors, rather than as the chief causes of common human diseases, it will be essential to provide accurate risk estimation and to inform the public of the need for population-based integrated data on genetic, biological, and environmental risk factors.

The goal of genomics research is ultimately prevention, the cornerstone of public health. An understanding of the significance of genetic risk factors and proper interpretations of their meaning for patients and their families will ultimately become part of clinical practice. Clinicians will become increasingly involved in helping patients to comprehend the meaning and potential impact of genetic risk for complex disorders. As our knowledge of the role of genetic risk factors in advances, it will be incumbent upon clinicians to become familiar with knowledge gleaned from genetic epidemiologic and genomics research. In the meanwhile, use of recurrence risk estimates from family studies best predicts the risk of the development of complex disorders.

References

- Beaty TH. Evolving methods in genetic epidemiology. I. Analysis of genetic and environmental factors in family studies. *Epidemiol Rev.* 1997;19(1):14–23.
- Beaty TH, Khoury MJ. Interface of genetics and epidemiology. *Epidemiol Rev.* 2000;22(1):120–5.
- Claus EB, Risch N, Thompson WD. The calculation of breast cancer risk for women with a first degree family history of ovarian cancer. *Breast Cancer Res Treat.* 1993;28(2):115–20.
- Eaton WW. Risk factors for mental health disorders (Unpublished Report). National Institute of Mental Health; 2004.
- Farrer LA, Cupples LA, Haines JL, Hyman B, Kukull WA, Mayeux R, Myers RH, Pericak-Vance MA, Risch N, van Duijn CM. Effects of age, sex, and ethnicity on the association between apolipoprotein E genotype and Alzheimer disease. A meta-analysis. APOE and Alzheimer disease meta analysis consortium. *JAMA.* 1997;278(16):1349–56.
- Gordis L, editor. *Epidemiology*. 2nd ed. Philadelphia: WB Saunders; 2000.
- Hawa MI, Beyan H, Buckley LR, Leslie RD. Impact of genetic and non-genetic factors in type 1 diabetes. *Am J Med Genet.* 2002;115(1):8–17. doi:10.1002/ajmg.10339.
- Heath AC, Whitfield JB, Madden PA, Bucholz KK, Dinwiddie SH, Slutske WS, Bierut LJ, Statham DB, Martin NG. Towards a molecular epidemiology of alcohol dependence: analysing the interplay of genetic and environmental risk factors. *Br J Psychiatry Suppl.* 2001;40:s33–40.
- Khoury MJ, Little J. Human genome epidemiologic reviews: the beginning of something HuGE. *Am J Epidemiol.* 2000;151(1):2–3.
- Kivipelto M, Helkala EL, Hanninen T, Laakso MP, Hallikainen M, Alhainen K, Soininen H, Tuomilehto J, Nissinen A. Midlife vascular risk factors and late-life mild cognitive impairment: a population-based study. *Neurology.* 2001;56(12):1683–9.
- Kivipelto M, Helkala EL, Laakso MP, Hanninen T, Hallikainen M, Alhainen K, Iivonen S, Mannerman A, Tuomilehto J, Nissinen A, Soininen H. Apolipoprotein E epsilon4 allele, elevated midlife total cholesterol level, and high midlife systolic blood pressure are independent risk factors for late-life Alzheimer disease. *Ann Intern Med.* 2002;137(3):149–55.
- Kolonel LN, Altshuler D, Henderson BE. The multiethnic cohort study: exploring genes, lifestyle and cancer risk. *Nat Rev Cancer.* 2004;4(7):519–27. doi:10.1038/nrc1389.
- Kornum BR, Faraco J, Mignot E. Narcolepsy with hypocretin/orexin deficiency, infections and autoimmunity of the brain. *Curr Opin Neurobiol.* 2011;21(6):897–903. doi:S0959-4388(11)00147-4.
- Kuller LH. The role of population genetics in the study of the epidemiology of cardiovascular risk factors. *Prog Clin Biol Res.* 1979;32:489–95.

- Langholz B, Rothman N, Wacholder S, Thomas DC. Cohort studies for characterizing measured genes. *J Natl Cancer Inst Monogr*. 1999;26:39–42.
- MacMahon B, Trichopoulos D, editors. *Epidemiology: principles and methods*. Boston: Little Brown and Company; 1996.
- Meigs JB, Shrader P, Sullivan LM, McAteer JB, Fox CS, Dupuis J, Manning AK, Florez JC, Wilson PW, D'Agostino Sr RB, Cupples LA. Genotype score in addition to common risk factors for prediction of type 2 diabetes. *N Engl J Med*. 2008;359(21):2208–19. doi:[359/21/2208](https://doi.org/10.1056/NEJM020119).
- Merikangas KR, editor. *Comorbidity of mood and anxiety disorders*. Washington, DC: American Psychiatric Press Inc; 1990.
- Merikangas KR, editor. *Genetic epidemiology of substance-use disorders, in biological psychiatry*. Chichester: Wiley; 2003.
- Merikangas KR, Risch N. Genomic priorities and public health. *Science*. 2003;302(5645):599–601. doi:[10.1126/science.1091468](https://doi.org/10.1126/science.1091468).
- Michael NL. Host genetic influences on HIV-1 pathogenesis. *Curr Opin Immunol*. 1999;11(4):466–74. doi:[10.1016/S0952-7915\(99\)80078-8](https://doi.org/10.1016/S0952-7915(99)80078-8).
- National Adoption Information Clearinghouse. The adoption home study process, 2007, available online at http://naic.acf.hhs.gov/pubs/f_homstu.cfm
- Nebert DW. Pharmacogenetics and pharmacogenomics: why is this relevant to the clinical geneticist? *Clin Genet*. 1999;56(4):247–58.
- Ödegaard Ö, editor. *Emigration and insanity: a study of mental disease among the Norwegian born population of Minnesota*. Copenhagen: Levin & Munksgaards; 1932.
- Ott J. Association of genetic loci: replication or not, that is the question. *Neurology*. 2004;63(6):955–8.
- Ottman R. An epidemiologic approach to gene-environment interaction. *Genet Epidemiol*. 1990;7(3):177–85. doi:[10.1002/gepi.1370070302](https://doi.org/10.1002/gepi.1370070302).
- Peltonen L, McKusick VA. Genomics and medicine. Dissecting human disease in the postgenomic era. *Science*. 2001;291(5507):1224–9.
- Risch N. Genetic linkage and complex diseases, with special reference to psychiatric disorders. *Genet Epidemiol*. 1990a;7(1):3–16; discussion 17–45. doi:[10.1002/gepi.1370070103](https://doi.org/10.1002/gepi.1370070103).
- Risch N. Linkage strategies for genetically complex traits. I. Multilocus models. *Am J Hum Genet*. 1990b;46(2):222–8.
- Rose RJ, Dick DM, Viken RJ, Kaprio J. Gene-environment interaction in patterns of adolescent drinking: regional residency moderates longitudinal influences of alcohol use. *Alcohol Clin Exp Res*. 2000;25:637.
- Tol J, Roks G, Slooter AJ, van Duijn CM. Genetic and environmental factors in Alzheimer's disease. *Rev Neurol (Paris)*. 1999;155 Suppl 4:S10–16.
- Varmus H. Getting ready for gene-based medicine. *N Engl J Med*. 2002;347(19):1526–7. doi:[10.1056/NEJMe020119](https://doi.org/10.1056/NEJMe020119).
- 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(14):1151–8.
- Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature*. 2007;447(7145):661–78. doi:[10.1038/nature05911](https://doi.org/10.1038/nature05911).
- Yang Q, Khoury MJ. Evolving methods in genetic epidemiology. III. Gene-environment interaction in epidemiologic research. *Epidemiol Rev*. 1997;19(1):33–43.



<http://www.springer.com/978-94-007-5557-4>

Applied Computational Genomics

Yao, Y. (Ed.)

2012, XII, 184 p., Hardcover

ISBN: 978-94-007-5557-4