A Genome-wide Association Study Primer for Clinicians

Tzu-Hao Wang[1,2,3], Hsin-Shih Wang[1,4]*

[1]*Department of Obstetrics and Gynecology, Chang Gung Memorial Hospital and Chang Gung University,* [2]*Genomic Medicine Research Core Laboratory, Chang Gung Memorial Hospital,* [3]*Graduate Institute of Basic Medical Sciences, and* [4]*Graduate Institute of Clinical Medical Sciences, Chang Gung University, Tao-Yuan, Taiwan.*

SUMMARY

Genome-wide association studies (GWAS) use high-throughput genotyping technology to relate hundreds of thousands of genetic markers (genotypes) to clinical conditions and measurable traits (phenotypes). This review is intended to serve as an introduction to GWAS for clinicians, to allow them to better appreciate the value and limitations of GWAS for genotype-disease association studies. The input of clinicians is vital for GWAS, since disease heterogeneity is frequently a confounding factor that can only really be solved by clinicians. For diseases that are difficult to diagnose, clinicians should ensure that the cases do indeed have the disease; for common diseases, clinicians should ensure that the controls are truly disease-free. [*Taiwan J Obstet Gynecol* 2009;48(2):89–95]

Key Words: copy number variation, genome-wide association studies, genotype, linkage disequilibrium, phenotype, single nucleotide polymorphisms

Introduction

Genome-wide association studies (GWAS) are based on the "common disease, common variant" assumption, which means that the genetic risk for a common disease is mostly attributable to a relatively small number of common genetic variants [1–4]. GWAS have only recently become practical after three key tools have been made available: the discovery of more than 10 million single nucleotide polymorphisms (SNPs) in the human genome (http://www.ncbi.nlm.nih.gov/SNP), the HapMap (http://www.hapmap.org) that measures the degree of association between the alleles of neighboring SNPs, and high-throughput genotyping platforms [5].

GWAS use high-throughput genotyping technology to relate hundreds of thousands of genetic markers to clinical conditions and measurable traits [1], with the expectation that systematic study of DNA variations throughout the genome in relation to a disease may lead to localization of the causal genes [3]. GWAS rely on a nonrandom association called linkage disequilibrium (Table 1) [6]. Contrary to the traditional hypothesis-driven research, GWAS are unconstrained by prior hypotheses regarding genetic associations with disease. Hence, GWAS represent an important step beyond candidate gene studies [1].

Case-control association studies, the most common type of GWAS, detect nonrandom co-occurrences between alleles (genotypes) and traits (phenotypes) [7]. Genotypes can be defined by many types of genetic markers, including SNPs, microsatellites, minisatellites and copy number variations (Table 1) [1,2,8–10]. SNPs are the preferred genetic markers for GWAS because of their abundance; about 12 million unique human SNPs have been assigned a reference SNP number in the National Center for Biotechnology Information's SNP database [10]. Recently, copy number variation polymorphisms have attracted researchers' attention because of their ubiquity and potential dosage effects on gene expression (Table 1) [11].

MacArthur praised GWAS, saying that "for the first time in human history, we have the power to identify the precise genetic differences between human beings that contribute to variation in disease susceptibility" [4].

Correspondence to: Dr Hsin-Shih Wang, Department of Obstetrics and Gynecology, Chang Gung Memorial Hospital and Chang Gung University, 5, Fu-Shin Road, Gwei-Shan, Tao-Yuan 333, Taiwan.
E-mail: hswang@cgmh.org.tw
Accepted: April 29, 2009

Table 1. Glossary of published genome-wide association studies

Term	Definition	Reference
Alleles	Alternate forms of a gene or chromosomal locus that differ in DNA sequence.	1
Copy number polymorphism	A copy number variation that occurs in more than 1% of the population.	8
Copy number variation	A DNA segment of at least 1,000 bp (1 kb) in size, for which copy number differences are observed in the comparison of two or more genomes.	8
Genome-wide association study	A type of phenotype–genotype association study conducted at the scale of the genome, usually using hundreds of thousands of markers of genetic variation.	1
Genotype	Any type of genetic variation. The most useful markers are single nucleotide polymorphisms, microsatellites, and copy number polymorphism markers.	2
Linkage disequilibrium	Association between two alleles located so near each other that they are inherited together more frequently than expected by chance.	1
Microsatellites	The length of the repeating unit is less than 10 bp.	9
Minisatellites	The length of the repeating units is from 10 to 100 bp.	9
Phenotype	A trait, including the disease to be studied.	2
Single nucleotide polymorphism	The smallest genomic variation is the change in a single nucleotide. When the single nucleotide variation occurs in more than 1% of the population, it is called a single nucleotide polymorphism.	10
Variable number tandem repeats	Also called short tandem repeats, including microsatellites and minisatellites.	9

Above the public excitement from the media and professional enthusiasm from academia, we need sufficient genomic knowledge to understand the results of GWAS. This review is not intended to exhaustively cover all reports of GWAS; instead, it aims to provide an introduction for clinicians to enable them to appreciate the values and limitations of GWAS as genotype–phenotype association studies.

GWAS in Complex Diseases

Several GWAS of complex human diseases (Tables 2 and 3) have recently been performed using the newly completed sequencing of the human genome [12] and the haplotype mapping of human SNPs [13,14]. In 2007 alone, GWAS were used to identify the risk loci for type 1 diabetes [15], type 2 diabetes [16–19], breast cancer [20–22], prostate cancer [23–25], myocardial infarction [26], atrial fibrillation [27], and autoimmunity [28] (Table 2). All of these papers were published in *Science* and *Nature* series journals.

Other GWAS results have been published in the *New England Journal of Medicine* from 2007 to 2008. The *New England Journal of Medicine* is the most rigorously reviewed and well-esteemed clinical journal, and their publication of GWAS results is indicative that such studies are considered to be of tremendous relevance to clinical medicine today. Risk loci in the human genome have been identified for coronary artery disease [29], sporadic amyotrophic lateral sclerosis [30], multiple sclerosis [31], rheumatoid arthritis and systemic lupus erythematosus [32], rheumatoid arthritis [33], autism [34], prostate cancer [35], asthma [36], neuroblastomas [37], and statin-induced myopathies [38] (Table 3).

The discovery of the gene regions associated with type 2 diabetes best demonstrates the strength of GWAS [39]. The identification of the disease genes for type 2 diabetes was regarded as "the geneticist's nightmare" [39]. This is because for more than a decade, geneticists devoted enormous efforts to candidate gene studies and international linkage consortia in attempts to find type 2 diabetes genes, but only obtained evidence for common variants in four genes, *KCNJ11* [40,41], *PPARG* [42], *TCF2* [24], and *WFS1* [43]. In 2007, however, six GWAS reported evidence for six new "gene regions" that may be involved in type 2 diabetes (Table 2) [16–19,44,45]. The closest genes to these regions are *HHEX-IDE*, *SLC30A8*, *CDKAL1*, *CDKN2A-2B*, *IGF2BP2*, and *FTO*. The fast pace of GWAS suggests that researchers will soon be able to simply look up case-control results for any common disease on the Web [39].

Table 2. Genome-wide association study reports in the *Nature* series and *Science* in 2007

Disease	Authors	Journal	Analytical platforms	Independent replication studies
Atrial fibrillation	Gudbjartsson et al [27]	*Nature*	Illumina	Yes
Autoimmunity	Burton et al [28]	*Nature Genetics*	Illumina	Yes
Breast cancer	Stacey et al [20]	*Nature Genetics*	Illumina	Yes
	Hunter et al [21]	*Nature Genetics*	Illumina	Yes
	Easton et al [22]	*Nature*	Affymetrix	Yes
Diabetes, type 2	Saxena et al [16]	*Science*	Affymetrix	Yes
	Zeggini et al [17]	*Science*	Affymetrix	Yes
	Scott et al [19]	*Science*	Illumina	Yes
	Sladek et al [18]	*Nature*	Illumina and Sequenom	No
	Steinthorsdottir et al [44]	*Nature Genetics*	Illumina	Yes
Myocardial infarction	Helgadottir et al [26]	*Science*	Illumina	Yes
Prostate cancer	Yeager et al [23]	*Nature Genetics*	Illumina	Yes
	Gudmundsson et al [24]	*Nature Genetics*	Illumina	Yes
Seven diseases*	Wellcome Trust Case Control Consortium [45]	*Nature*	Affymetrix	No

**Bipolar disorder, coronary disease, Crohn's disease, hypertension, rheumatoid arthritis, and type 1 and type 2 diabetes.*

Table 3. Genome-wide association study reports in the *New England Journal of Medicine* (2007–2008)

Disease	Authors (yr)	Analytical platforms	Independent replication studies
Asthma and lung function	Ober et al (2008) [36]	Affymetrix	Yes
Autism	Weiss et al (2008) [34]	Affymetrix	Yes
Coronary artery disease	Samani et al (2007) [29]	Affymetrix	Yes
Multiple sclerosis	Hafler et al (2007) [31]	Affymetrix	Yes
Neuroblastoma	Maris et al (2008) [37]	Illumina	Yes
Prostate cancer	Zheng et al (2008) [35]	Sequenom MassArray	This study is a replication itself
Rheumatoid arthritis	Plenge et al (2007) [33]	Illumina	Yes
Rheumatoid arthritis and systemic lupus erythematosus	Remmers et al (2007) [32]	Illumina	Yes
Sporadic amyotrophic lateral sclerosis	Dunckley et al (2007) [30]	Affymetrix and Illumina	Yes
Statin-induced myopathy	Link et al (2008) [38]	Illumina	Yes

Study Designs and Strategies for GWAS

Detailed instructions for conducting a sound GWAS are beyond the scope of this introductory review, and readers who plan to perform their own GWAS are encouraged to read the comprehensive reviews that address the rationale [46], data analysis [47], biostatistical aspects [48], and interpretation [1] of GWAS.

Furthermore, a working group organized by the National Cancer Institute and the National Human Genome Research Institute of the United States [2] has recently addressed three topics: (1) assessment of the validity and limitations of any genetic association study, (2) criteria for establishing replication in genetic association studies, and (3) points to consider for publication of sound genotype–phenotype association reports. Chanock et al [2] summarized criteria both for evaluating the soundness of an initial association report and for establishing positive replication in a checklist format.

GWAS can follow case-control, cohort or trio designs [1]. Case-control studies are by far the most commonly used for GWAS, and include all the studies listed in Tables 2 and 3. The most obvious advantage of the case-control study design is that large numbers of case and control participants can be recruited in a short time frame. However, this design is prone to biases that are mainly caused by population stratification (discussed later in the "Limitations of GWAS" section). In cohort studies, such as the Women's Genome Health Study [49] or Framingham Heart Study [50], a large number of participants with extensive baseline information are followed up to detect the incidence of disease development throughout the study period, which is usually decades. Some participants develop disease during the observation period, and thus the disease risk can be estimated [36]. The cohort design is free of survival bias, but its apparent disadvantage is the expense of a lengthy follow-up period. The trio design enrolls the affected participant and both of his/her parents, in order to study linkage disequilibrium of the disease and genetic markers [51]. The trio design is not affected by genetic differences between case and control participants (unrelated to disease or population stratification), but it can be difficult to recruit both parents and the diseased offspring in disorders with older ages of onset [1]. Ober et al [36] applied both case-control and cohort designs to identify potential serum biomarkers for asthma.

The two-stage approach for GWAS is a common strategy that maximizes statistical power while still maintaining reasonable costs [1,52]. In the first stage, complete panels of genetic markers (e.g. 500,000 SNPs or more) are screened for a fraction of samples (e.g. 500 cases and 500 controls) to identify groups of markers that reach a threshold of statistical significance, e.g. $p < 10^{-7}$. In the second stage, only the groups of genetic markers shown to be potentially associated with the disease in the first stage are further genotyped across all the recruited case and control participants (e.g. 3,000–10,000 cases and the same numbers of controls). Furthermore, a three-stage strategy has been proposed: (1) in stage one, 400 cases and 400 controls are genotyped for 500,000 SNPs; (2) in stage two, 4,000 cases and 4,000 controls are genotyped for 25,000 SNPs; and (3) in stage three, 20,000 cases and 20,000 controls are genotyped for 25 SNPs [1]. However, thanks to the advent of new versions of microarrays, such as Affymetrix SNP array 6.0 (Affymetrix Inc., Santa Clara, CA, USA) with 1.8 million genetic markers on a single microarray but with a lower cost per microarray than previous chips, we have a better platform to use for genotyping the whole sample population using the same microarray platform, i.e. the one-stage approach.

Genotyping Platforms for GWAS

Exponential improvements in analytic platforms, in terms of both hardware and software, are among the driving factors behind the dramatic discovery of genome-wide gene-disease associations. For GWAS with SNP and copy number polymorphism markers [34,53], both Affymetrix (http://www.affymetrix.com) and Illumina (http://www.illumina.com) microarray systems have been used successfully (Tables 2 and 3). In 2006, Barrett and Cardon [54] evaluated the first generation of these high-throughput platforms and concluded that both offered similar levels of genome coverage.

If the goal of GWAS is to identify the SNPs that account for amino acid changes in disease-related genes, i.e. nonsynonymous SNPs, then both Illumina [28] and Affymetrix coding SNP 10K can be used. To perform replication studies on a limited number of preselected genetic markers, mass spectrometry genotyping (e.g. Sequenom MassArrays [Sequenom, San Diego, CA, USA]) [35] and real-time quantitative polymerase chain reaction (PCR) genotyping (e.g. TaqMan technology [Applied Biosystems, Foster City, CA, USA]) may be less expensive than entire-genome microarrays. Another PCR-based platform, Dynal RELI SSO assay (Dynal Biotech Ltd., Bromborough, UK) has been used to analyze 1,729 SNPs and to localize type 1 diabetes susceptibility to the major histocompatibility complex class I genes, HLA-B and HLA-A [15].

Limitations of GWAS

Potential limitations of GWAS include: (1) false-positive and false-negative results, (2) insensitivity to rare variants and structural variants, (3) requirement for large sample sizes, (4) genotyping errors, (5) lack of information on gene function, and (6) possible biases due

to inappropriate selection of cases and controls [1]. Possible solutions for these limitations are suggested below.

A larger sample size will overcome the first three limitations. Confirmation of gene-disease association in replication studies is currently mandatory [2]. The sample size for the discovery of gene regions in a common disease such as type 2 diabetes is proposed to be >10,000 cases and controls [39]. Regarding the limitation of genotyping errors, dramatic advancements in analytical technology will eventually resolve genotyping errors. Currently, confirmation of critical SNPs using other technology is advised, such as real-time PCR or mass spectrometry genotyping. For the fifth limitation, GWAS are intended to identify associations between candidate genomic loci, or gene regions (rather than actual genes), and a disease [39]. After a disease-associated region has been confirmed in replication studies, whether or not the genes within the region really cause the disease should be examined using cell culture systems and/or animal models that have knock-out or knock-in genes [55].

Two groups of biases can confound GWAS results: population stratification and disease heterogeneity. Population stratification, also called population substructure, is caused by differences in allele frequency between cases and controls due to systematic ancestral differences that are defined by ethnicity or geographic origin [56]. Population stratification-related variations in allele frequencies may cause differences in disease risk, which, in turn, falsely identify the subgroup-associated genes as being disease-related [1]. To avoid this bias, attention should be paid to ensure that the distribution of risk factors of interest in controls is the same as that in the diseased cases [57]. Disease heterogeneity, on the other hand, frequently makes the classification of case participants less straightforward than expected. Technology and geneticists have to rely on clinicians to weed out false-positives and false-negatives. For diseases that are difficult to diagnose, clinicians should ensure that the cases do indeed have the disease; whereas for common diseases, clinicians should ensure that the controls are truly disease-free [1].

Concluding Remarks

Given the abundance of genetic association studies (Tables 2 and 3), we doubt the need to plan new initial genetic association or replication studies. To address this, the first step is to carry out a comprehensive literature review of the phenotype. If the selected phenotype has not previously been convincingly studied, and we have sufficient numbers of subjects with an unambiguous and relatively homogenous phenotype, we are in a good position to perform an initial genetic association study. In considering the need for replication studies, it is important to remember that evaluations of an association in populations with different ancestries from that of the initial report are acutely needed, considering that genomic variation is greater across populations [2]. Most genetic association studies have been conducted in Caucasians, rather than in individuals of Asian or Chinese ethnicities. Therefore, replication studies in Chinese populations, which validate the initial genetic association for the phenotype, would greatly increase our confidence regarding the initial genotype–phenotype association.

When an independent replication study is planned, the selection criteria for the SNPs to be replicated from the initial study include an estimated large effect of the reported loci (>2 is the optimum suggested by Chanock et al [2]), a low *p* value in linkage-disequilibrium calculations, and that supporting results of those SNP have been reported. Additionally, SNPs to be further verified should be selected after considering their putative functional correlations. Validation of such SNPs not only confirms the significance of previously reported phenotype–genotype associations (despite genomic variations among different ethnicities), but also provides insight into our understanding of diseases, so helping us to develop new testable experiments to discover the biologic and molecular mechanisms behind the diseases.

Acknowledgments

The authors would like to thank Dr Yun-Shien Lee (Department of Biotechnology, Ming Chuan University) for his longstanding contribution to the Genomic Medicine Research Core Laboratory of Chang Gung Memorial Hospital (Tao-Yuan, Taiwan) and Shih-Yee Mimi Wang (University of Illinois, Chicago College of Medicine) for editing the manuscript. This study was supported by grants NSC 96-2314-B-182A-124 (to T.H. Wang) and NSC 95-2314-B-182-061 (to H.S. Wang) from the National Science Council, and CMRPG340463 (to T.H. Wang) and CMRPG340713 (to H.S. Wang) from Chang Gung Memorial Hospital, Taiwan.

References

1. Pearson TA, Manolio TA. How to interpret a genome-wide association study. *JAMA* 2008;299:1335–44.

2. Chanock SJ, Manolio T, Boehnke M, et al; for NCI-NHGRI Working Group on Replication in Association Studies. Replicating genotype-phenotype associations. *Nature* 2007; 447:655–60.
3. Hunter DJ, Altshuler D, Rader DJ. From Darwin's finches to canaries in the coal mine—mining the genome for new biology. *N Engl J Med* 2008;358:2760–3.
4. MacArthur D. Why do genome-wide scans fail? Available at: http://www.genetic-future.com/2008/03/why-do-genome-wide-scans-fail.html
5. Mathew CG. New links to the pathogenesis of Crohn disease provided by genome-wide association scans. *Nat Rev Genet* 2008;9:9–14.
6. Gabriel SB, Schaffner SF, Nguyen H, et al. The structure of haplotype blocks in the human genome. *Science* 2002;296: 2225–9.
7. Pasternak JJ. *An Introduction to Human Molecular Genetics*, 2nd edition. Hoboken, NJ: Wiley-Liss, 2005.
8. Scherer SW, Lee C, Birney E, et al. Challenges and standards in integrating surveys of structural variation. *Nat Genet* 2007;39:7–15.
9. Lewin B. *Genes IX*. Sudbury, MA: Jones & Bartlett Publishers, 2008.
10. Wheeler DL, Barrett T, Benson DA, et al. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res* 2008;36:D13–21.
11. McCarroll SA, Altshuler DM. Copy-number variation and association studies of human disease. *Nat Genet* 2007; 39(7 Suppl):S37–42.
12. International Human Genome Sequencing Consortium. Finishing the euchromatic sequence of the human genome. *Nature* 2004;431:931–45.
13. Hinds DA, Stuve LL, Nilsen GB, et al. Whole-genome patterns of common DNA variation in three human populations. *Science* 2005;307:1072–9.
14. International HapMap Consortium. A haplotype map of the human genome. *Nature* 2005;437:1299–320.
15. Nejentsev S, Howson JM, Walker NM, et al. Localization of type 1 diabetes susceptibility to the MHC class I genes HLA-B and HLA-A. *Nature* 2007;450:887–92.
16. Saxena R, Voight BF, Lyssenko V, et al; Diabetes Genetics Initiative of Broad Institute of Harvard and MIT, Lund University, and Novartis Institutes of BioMedical Research. Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. *Science* 2007;316:1331–6.
17. Zeggini E, Weedon MN, Lindgren CM, et al. Replication of genome-wide association signals in UK samples reveals risk loci for type 2 diabetes. *Science* 2007;316:1336–41.
18. Sladek R, Rocheleau G, Rung J, et al. A genome-wide association study identifies novel risk loci for type 2 diabetes. *Nature* 2007;445:881–5.
19. Scott LJ, Mohlke KL, Bonnycastle LL, et al. A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. *Science* 2007;316:1341–5.
20. Stacey SN, Manolescu A, Sulem P, et al. Common variants on chromosomes 2q35 and 16q12 confer susceptibility to estrogen receptor-positive breast cancer. *Nat Genet* 2007; 39:865–9.
21. Hunter DJ, Kraft P, Jacobs KB, et al. A genome-wide association study identifies alleles in FGFR2 associated with risk of sporadic postmenopausal breast cancer. *Nat Genet* 2007; 39:870–4.
22. Easton DF, Pooley KA, Dunning AM, et al. Genome-wide association study identifies novel breast cancer susceptibility loci. *Nature* 2007;447:1087–93.
23. Yeager M, Orr N, Hayes RB, et al. Genome-wide association study of prostate cancer identifies a second risk locus at 8q24. *Nat Genet* 2007;39:645–9.
24. Gudmundsson J, Sulem P, Steinthorsdottir V, et al. Two variants on chromosome 17 confer prostate cancer risk, and the one in *TCF2* protects against type 2 diabetes. *Nat Genet* 2007;39:977–83.
25. Haiman CA, Patterson N, Freedman ML, et al. Multiple regions within 8q24 independently affect risk for prostate cancer. *Nat Genet* 2007;39:638–44.
26. Helgadottir A, Thorleifsson G, Manolescu A, et al. A common variant on chromosome 9p21 affects the risk of myocardial infarction. *Science* 2007;316:1491–3.
27. Gudbjartsson DF, Arnar DO, Helgadottir A, et al. Variants conferring risk of atrial fibrillation on chromosome 4q25. *Nature* 2007;448:353–7.
28. Burton PR, Clayton DG, Cardon LR, et al; Wellcome Trust Case Control Consortium, and Australo-Anglo-American Spondylitis Consortium (TASC). Association scan of 14,500 nonsynonymous SNPs in four diseases identifies autoimmunity variants. *Nat Genet* 2007;39:1329–37.
29. Samani NJ, Erdmann J, Hall AS, et al. Genomewide association analysis of coronary artery disease. *N Engl J Med* 2007; 357:443–53.
30. Dunckley T, Huentelman MJ, Craig DW, et al. Whole-genome analysis of sporadic amyotrophic lateral sclerosis. *N Engl J Med* 2007;357:775–88.
31. Hafler DA, Compston A, Sawcer S, et al; International Multiple Sclerosis Genetics Consortium. Risk alleles for multiple sclerosis identified by a genomewide study. *N Engl J Med* 2007;357:851–62.
32. Remmers EF, Plenge RM, Lee AT, et al. *STAT4* and the risk of rheumatoid arthritis and systemic lupus erythematosus. *N Engl J Med* 2007;357:977–86.
33. Plenge RM, Seielstad M, Padyukov L, et al. TRAF1-C5 as a risk locus for rheumatoid arthritis—a genomewide study. *N Engl J Med* 2007;357:1199–209.
34. Weiss LA, Shen Y, Korn JM, et al. Association between microdeletion and microduplication at 16p11.2 and autism. *N Engl J Med* 2008;358:667–75.
35. Zheng SL, Sun J, Wiklund F, et al. Cumulative association of five genetic variants with prostate cancer. *N Engl J Med* 2008;358:910–9.
36. Ober C, Tan Z, Sun Y, et al. Effect of variation in *CHI3L1* on serum YKL-40 level, risk of asthma, and lung function. *N Engl J Med* 2008;358:1682–91.
37. Maris JM, Mosse YP, Bradfield JP, et al. Chromosome 6p22 locus associated with clinically aggressive neuroblastoma. *N Engl J Med* 2008;358:2585–93.
38. Link E, Parish S, Armitage J, et al; SEARCH Collaborative Group. *SLCO1B1* variants and statin-induced myopathy—a genomewide study. *N Engl J Med* 2008;359:789–99.
39. Frayling TM. Genome-wide association studies provide new insights into type 2 diabetes aetiology. *Nat Rev Genet* 2007; 8:657–62.

40. Nielsen EM, Hansen L, Carstensen B, et al. The E23K variant of Kir6.2 associates with impaired post-OGTT serum insulin response and increased risk of type 2 diabetes. *Diabetes* 2003;52:573–7.
41. Gloyn AL, Weedon MN, Owen KR, et al. Large-scale association studies of variants in genes encoding the pancreatic β-cell K_{ATP} channel subunits Kir6.2 (*KCNJ11*) and SUR1 (*ABCC8*) confirm that the *KCNJ11* E23K variant is associated with type 2 diabetes. *Diabetes* 2003;52:568–72.
42. Altshuler D, Hirschhorn JN, Klannemark M, et al. The common PPARγ Pro12Ala polymorphism is associated with decreased risk of type 2 diabetes. *Nat Genet* 2000;26:76–80.
43. Sandhu MS, Weedon MN, Fawcett KA, et al. Common variants in *WFS1* confer risk of type 2 diabetes. *Nat Genet* 2007; 39:951–3.
44. Steinthorsdottir V, Thorleifsson G, Reynisdottir I, et al. A variant in *CDKAL1* influences insulin response and risk of type 2 diabetes. *Nat Genet* 2007;39:770–5.
45. Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* 2007;447:661–78.
46. Hirschhorn JN, Daly MJ. Genome-wide association studies for common diseases and complex traits. *Nat Rev Genet* 2005;6:95–108.
47. Zhang H, Liu L, Wang X, Gruen JR. Guideline for data analysis of genomewide association studies. *Cancer Genomics Proteomics* 2007;4:27–34.
48. Ziegler A, Konig IR, Thompson JR. Biostatistical aspects of genome-wide association studies. *Biom J* 2008;50:8–28.
49. Ridker PM, Chasman DI, Zee RY, et al. Rationale, design, and methodology of the Women's Genome Health Study: a genome-wide association study of more than 25,000 initially healthy American women. *Clin Chem* 2008;54:249–55.
50. Jaquish CE. The Framingham Heart Study, on its way to becoming the gold standard for Cardiovascular Genetic Epidemiology? *BMC Med Genet* 2007;8:63.
51. Spielman RS, McGinnis RE, Ewens WJ. Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). *Am J Hum Genet* 1993; 52:506–16.
52. Kronenberg F. Genome-wide association studies in aging-related processes such as diabetes mellitus, atherosclerosis and cancer. *Exp Gerontol* 2008;43:39–43.
53. Xu B, Roos JL, Levy S, van Rensburg EJ, Gogos JA, Karayiorgou M. Strong association of de novo copy number mutations with sporadic schizophrenia. *Nat Genet* 2008; 40:880–5.
54. Barrett JC, Cardon LR. Evaluating coverage of genome-wide association studies. *Nat Genet* 2006;38:659–62.
55. Frayling TM, McCarthy MI. Genetic studies of diabetes following the advent of the genome-wide association study: where do we go from here? *Diabetologia* 2007;50:2229–33.
56. Yu K, Wang Z, Li Q, et al. Population substructure and control selection in genome-wide association studies. *PLoS ONE* 2008;3:e2551.
57. Wacholder S, McLaughlin JK, Silverman DT, Mandel JS. Selection of controls in case-control studies, I: principles. *Am J Epidemiol* 1992;135:1019–28.