How to Include Chromosome X in Your Genome-Wide Association Study

Inke R. König,1,2,* Christina Loley,1,3 Jeanette Erdmann,2,3 and Andreas Ziegler1,2

1Institut für Medizinische Biometrie und Statistik, Universität zu Lübeck, Universitätsklinikum Schleswig-Holstein, Campus Lübeck, Lübeck, Germany; 2DZHK (German Centre for Cardiovascular Research), Lübeck, Germany; 3Institut für Integrative und Experimentelle Genomik, Universitätsklinikum Schleswig-Holstein, Campus Lübeck, Lübeck, Germany; 4Zentrum für Klinische Studien Lübeck, Lübeck, Germany

Received 21 August 2013; Revised 29 October 2013; accepted revised manuscript 2 December 2013.
Published online 9 January 2014 in Wiley Online Library (wileyonlinelibrary.com). DOI 10.1002/gepi.21782

ABSTRACT: In current genome-wide association studies (GWAS), the analysis is usually focused on autosomal variants only, and the sex chromosomes are often neglected. Recently, a number of technical hurdles have been described that add to a reluctance of including chromosome X in a GWAS, including complications in genotype calling, imputation, and selection of test statistics. To overcome this, we provide a “how to” guide for analyzing X chromosomal data within a standard GWAS. Following a general pipeline for GWAS, we highlight the steps in which the X chromosome requires specific attention, and we give tentative advice for each of these. Through this, we show that by selection of sensible algorithms and parameter settings, the inclusion of chromosome X in GWAS is manageable. Closing this gap is expected to further elucidate the genetic background of complex diseases, especially of those with sex-specific features.

KEY WORDS: case-control association; sex chromosomes

Introduction

Genome-wide association studies (GWAS) have been a successful approach in the past years to identify candidate regions playing a role in complex diseases. However, the X chromosome was excluded from most GWAS, and only 242 out of all 743 GWAS conducted from 2005 to 2011 considered chromosome X in their analyses according to the recent report of Wise et al. [2013]. When considering only family-based GWAS, this proportion is the same: We identified 92 family-based GWAS until May 2013 in the GWAS catalog [Hindorff et al., 2013]; 67 of them provided information on whether chromosome X was analyzed. Of these, only 22 (33%) had considered chromosome X.

Wise et al. [2013] discussed several reasons for this neglect of chromosome X, and these include a lower proportion of genes on chromosome X and a lower coverage of chromosome X on current genotyping platforms compared with autosomal coverage among others. These authors also described a number of technical hurdles that might add to the reluctance of including chromosome X in a GWAS, including complications in genotype calling, imputation, selection of test statistics, and the lack of readily available implementations. However, no “how to” guide for analyzing X chromosomal data has been provided yet. The aim of this paper therefore is to give recommendations on how to overcome these issues. For this, we build on our experience from a number of GWAS and meta-analyses of GWAS [CARDIoGRAMplusC4D Consortium et al., 2013; Erdmann et al., 2009; Samani et al., 2007; Schunkert et al., 2011; Thye et al., 2010; Timmann et al., 2012].

Before describing the technical issues that need to be considered in the analysis of chromosome X, we recall the specific characteristics that make chromosome X different from the autosomes. The first is the special genetic make-up of chromosome X with females having two copies and males having one copy, apart from the pseudo-autosomal regions. Thus, as soon as males are included in the analysis, variants on X chromosomal loci need special treatment. For example, as males carry only one allele, the signal intensities obtained from standard array genotyping platforms are lower than for females who carry two alleles. This needs to be adequately addressed in the genotype-calling step and has further consequences for genotype imputation and association analysis.

The second special feature of the X chromosome is the process of inactivation. Early in embryonic development, large parts of one of the two female X chromosomes are silenced, which is hypothesized to be a mechanism of dosage compensation, resulting in equal effects for one copy of the X chromosome in males and two copies in females. This inactivation is not complete, and it is estimated that about three quarters of X chromosomal genes are silenced on one of the female X chromosomes in some of the individuals. This is of
importance when deciding how to test for association with X chromosomal variants as described below.

Methods and Results

The general pipeline of a GWAS has been detailed previously [Ziegler, 2009; Ziegler and König, 2010]. Where applicable, we will follow this pathway in the following, and the steps that merit special attention when working on the X chromosome are highlighted in Fig. 1. As before [Ziegler et al., 2008], we here focus on the most common setting for GWAS, which is the analysis of complex traits in case-control or cohort studies with independent individuals.

Design Stage

The starting point of a GWAS is the biological hypothesis. Thus, the first question is why one should consider X chromosomal variants in the analysis in the first place, and there might be a number of reasons for this. First, according to the 1000 Genomes Project [The 1000 Genomes Project Consortium, 2012], the X chromosome carries more than 2300 coding genes, noncoding genes and pseudogenes. It has more than two million variations, and there is no a priori reason why these should be neglected in the analyses. Indeed, a number of these are known to be responsible for classical Mendelian diseases such as Coffin-Lowry syndrome [OMIM #303600, Marques Pereira et al., 2010], Charcot-Marie-Tooth neuropathy [OMIM #302800, Ionascu et al., 1991], Duchenne muscular dystrophy [OMIM #310200, Bushby, 1992], or Hemophilia A [OMIM +306700, Mannucci and Tuddenham, 2001], so these regions are also candidates for complex diseases. Second, as evidenced in those diseases, X chromosomal inheritance leads to sex differences in prevalence and specific appearance of the disease. From the opposite perspective, if there are sex differences in the disease, these might be due to the genetic predisposition being X linked. For example, for coronary artery disease and myocardial infarction, sex differences are pronounced with different risk profiles, ages at onset, symptoms, and prognosis for males and females [Solimene, 2010]. Thus, it might be hypothesized that X chromosomal regions are involved in the predisposition to coronary artery disease.

Moving on to the actual design for a GWAS, the adequate sample is selected in the first stage. In addition to usual considerations, it is especially important to consider the male to female ratio for X chromosomal analyses. For example, in some algorithms for genotype calling and genotype imputation, males and females are called separately so that a sufficient sample size in each subgroup is required [Affymetrix, 2006]. The power of the association tests also depends on the male to female ratio in cases and controls [Loley et al., 2011]. For example, most available tests have a decreased power in unbalanced designs, but some tests that are described below have increased power if there are more males in the cases but fewer males in the controls. Nevertheless, a balanced sampling is to be recommended because different proportions of females in cases and controls will cause a severe increase in type I error frequencies of some test statistics as soon as sex-specific allele frequencies occur, which is typically observed for a sensible number of variants. For instance, in our own study that is described below [Erdmann et al., 2009], we observed that in about 16% of the X chromosomal SNPs there
were allele frequency differences of at least 2%, which already led to visibly increased error levels [Loley et al., 2011].

An additional issue in the study design is the selection of the genotyping chip. In this context, the coverage of chromosome X on different chips is important, which can be estimated by the number of available single-nucleotide polymorphisms (SNPs) related to the physical length of the chromosome [2013]. Unexpectedly, although the length of the X chromosome is about 5% of the entire human genome, typical genotyping platforms have a range of 2.0% (GeneChip Human Mapping 500K Array Set by Affymetrix), 2.6% (HumanOmní-Quad BeadChip by Illumina) to 3.5% (Genome-Wide Human SNP Array 6.0 by Affymetrix) of X chromosomal SNPs [Wise et al., 2013]. In a similar way, we can relate the proportion of available X chromosomal variants to the proportion of X chromosomal variants identified in the 1000 Genomes Project [The 1000 Genomes Project Consortium, 2012], which is roughly 4%. This shows that X chromosomal variants are underrepresented in current genotyping technologies compared with autosomal regions. Important for chip selection, this imbalance is greater for older genotyping chips.

**Low-Level Analysis: Genotype Calling**

After having undergone the laboratory stages and early low-level analysis steps as described elsewhere [Ziegler, 2009; Ziegler and König, 2010], genotypes need to be called. Software packages differ with regard to how chromosome X is handled and, consequently, how well chromosome X genotypes can be called. As shown by Ritchie et al. [2011], genotype calling algorithms which utilize the sex information in the genotype calling, such as Illuminus [Teo et al., 2007] and CRLMM [Carvalho et al., 2007] perform better than those using the same model for both sexes, such as GenCall [Kermani, 2008] and GenoSNP [Giannoulatou et al., 2008]. Birdseed [Korn et al., 2008] similarly clusters males and females separately, and the same advantage is likely to apply.

**Quality Control**

The first round of quality control after genotype calling merits specific attention. On the subject level, the same criteria are used as defined by the Tramvände filter criteria [Ziegler, 2009; Ziegler et al., 2008]. In addition to that, the availability of X chromosomal data allows comparing genotypic sex with phenotypic sex, and individuals with divergent sex information are excluded from further analysis. This can be accompanied by a more specific method such as plotting the mean signal intensities of the X and Y chromosomes. Here, females should have low-Y and high-X intensities, whereas males should have similar intensity levels for X and Y [Laurie et al., 2010; Turner et al., 2011]. On the SNP level, the filter criteria are modified for X chromosomal SNPs and shown in Table 1.

It should be noted that there is some redundancy in the criteria concerning missing frequency (lines 4–7). For practical purposes, we recommend using either the separate missing frequencies in males and females (lines 4–5) or the absolute difference in missing frequencies between males and females together with the differential missingness (lines 6–7) as suggested by Ling et al. [2009] and Ziegler [2009]. Furthermore, we stress that if genotype calling is performed as described above, that is, separately in males and females, no heterozygote calls in males are allowed. If this is not the case, the proportion of heterozygote calls in males should be considered as an additional criterion as described by Ziegler [2009] and Ling [2009]. Although the described filter criteria are helpful in identifying SNPs of low quality, the visual inspection of signal intensity plots is still the ultimate quality control approach after an association has been identified [Ziegler, 2009]. Due to the different intensity levels expected for X chromosomal data, these plots should be generated separately for males and females.

To compare the genotyping quality of X chromosomal variants with that on the autosome, Table 2 shows the results of the quality control on the SNP level for chromosome X based on the data from the German Myocardial Infarction Family Study 2 (GerMIFS-2) [Erdmann et al., 2009]. In this study, we genotyped 248 female cases, 974 male cases, 622 female controls, and 676 male controls on the Affymetrix Genome-Wide Human SNP Array 6.0 called with Birdseed. Because chromosome 7 is of about equal size physically, quality control results are shown for comparison.

First, we note that using the missing frequencies as shown in the table leads to a slightly higher exclusion fraction than using the absolute difference in missing frequencies and the test for differential missingness (exclusion of 7,071 SNPs vs. 6,615 SNPs). Second, comparing the results between chromosomes shows that in almost every step, a greater proportion of SNPs are lost on chromosome X than on 7. Consequently, almost half of the genotyped SNPs failed QC. This coincides with our further experience that, in general, quality is lower on chromosome X than on the autosomes. Furthermore, the sex-imbalanced sample design of this study might add to this problem. Because there are only 248 female cases, the calling of this group might have been less reliable and thereby explain some extent of the lower quality of these data.

### Table 1. Filters for quality control on the SNP level for X chromosomal SNPs

| SNP filter criterion | Standard value for filter |
|----------------------|---------------------------|
| MAF in males         | ≥ 1%                      |
| MAF in females       | ≥ 1%                      |
| MAFF in males        | ≤ 2% in any study group   |
| MAFF in females      | ≤ 2% in any study group   |
| MAFF (males) – MAFF (females) | ≤ 2% in any study group |
| Differential missingness | \( P \geq 10^{-4} \) in \( \chi^2 \) test for independence between males and females |
| Departure from HWE in females | \( P \geq 10^{-4} \) |
| Differences between control groups | \( P \geq 10^{-4} \) in Cochran-Armitage trend test between control groups |
| Sex differences among controls | \( P \geq 10^{-4} \) in Cochran-Armitage trend test between males and females |
| Heterozygosity in males | Set to missing, considered in MAF in males |

MAF = minor allele frequency; MAFF = missing frequency; HWE = Hardy–Weinberg equilibrium.
Table 2. Results from the SNP level quality control for genotyped SNPs on chromosomes X and 7 based on data from the German Myocardial Infarction Family Study 2. Shown are absolute and relative numbers of SNPs available in each step and percent of SNPs lost in each step compared with the previous one.

|                        | Chromosome X |          | Chromosome 7 |          |
|------------------------|--------------|----------|--------------|----------|
|                        | Absolute number (%) | % Lost | Absolute number (%) | % Lost |
| Number of SNPs before QC | 36,310 (100.0) |        | 47,056 (100.0) |       |
| Not monomorphic        | 33,991 (93.6) | 6.4     | 46,061 (97.9) | 2.1     |
| MAF ≥ 1%               |              |         |              |         |
| Females                | 27,242 (75.0) | 19.9    | 40,577 (86.2) | 11.9    |
| Males                  | 27,101 (74.6) | 0.5     |              |         |
| MiF cases ≤ 2%         |              |         |              |         |
| Females                | 22,218 (61.2) | 18.0    | 34,536 (73.4) | 14.9    |
| Males                  | 22,202 (61.2) | 0.1     |              |         |
| MiF controls ≤ 2%      |              |         |              |         |
| Females                | 20,039 (55.2) | 9.7     | 31,300 (66.5) | 9.4     |
| Males                  | 20,036 (55.2) | 0.0     |              |         |
| Departure from HWE in (female) controls P ≤ 10⁻⁴ | 20,003 (55.1) | 0.1 | 31,216 (66.3) | 0.3 |

QC = quality control; MAF = minor allele frequency; MiF = missing frequency; HWE = Hardy–Weinberg equilibrium.

High-Level Analysis: Imputation

Given the generally low coverage and high loss of X chromosomal SNPs in QC as described above, it is especially important for meta-analyses to fill these gaps by imputation of SNPs that are not available. For example, when combining preliminary data from 10 GWAS on coronary artery disease, in principle, about 35,000 SNPs were available on chromosome X. However, after exclusion of low quality SNPs, less than 10,000 were available in six or more of these studies, making imputation crucial for combining data [Loley et al., 2012].

It needs to be noted that the imputation of variations on chromosome X is handled differently in common algorithms. An overview by Marchini and Howie [2010] shows that some algorithms, such as IMPUTE v1 and v2.2, Howie et al., 2009, can directly impute on chromosome X. For this, additional parameters need to be set with “chrX” for non-pseudo-autosomal regions. But the effort then is similar to imputation on the autosome. Other algorithms, such as MACH (v1.0.16) or Minimac [Howie et al., 2012], fastPHASE [v1.4.0, Scheet and Stephens, 2006] and BIMBAM [v0.99, Servin and Stephens, 2007], cannot impute the X chromosome directly. Instead, the imputation needs to be run separately for males and females. Imputation can thus not be performed together with the autosome, and even not jointly for males and females.

Evaluating the quality of the genotype imputation is the next step. In this, again, the criteria are used as before, although one might be more liberal with the criterion of missing genotypes; we excluded SNPs with a missing frequency greater than 25% in male cases, male controls, female cases, or female controls. Additionally, a threshold for the quality of the imputation is used with INFO = 0.5. Again, we display results from comparing chromosomes X and 7 based on data from the GerMIFS-2 [Erdmann et al., 2009] in Table 3 after imputation with IMPUTE v2.2.2 [Howie et al., 2009]. This was based on a reference panel of the 1000 Genomes Project including 567 female and 525 male subjects [The 1000 Genomes Project Consortium, 2012].

Strikingly, almost two-thirds of the imputed SNPs were lost on chromosome X because of being monomorphic, whereas this was less than 50% on chromosome 7. For the other filter criteria, no clear differences were visible. As a result, 16% of the imputed SNPs were available after QC on chromosome X but about 25% on chromosome 7. As a side note, our experiences have shown that imputation of SNPs in the pseudo-autosomal region is worse rendering mostly too few SNPs to be considered for analysis at all.

High-Level Analysis: Statistical Analysis

Entering the stage of association analysis with high-quality genotypes on both the SNP and the subject level, the first aspect to note is that the analysis is not different from the standard analysis on autosomes if either only the pseudo-autosomal area or only female subjects are considered. Otherwise, a number of different test statistics are available that differ from those for the autosomal loci. These are described briefly in Table 4, more details are found in the literature [Clayton, 2008; Loley et al., 2011; Zheng et al., 2007]. Which of these statistics are valid and powerful depends on the answers to the following questions [Hickey and Bahlo, 2011; Loley et al., 2011]:

1. Is there departure from Hardy–Weinberg equilibrium in females? If yes, some of the test statistics have inflated type I error frequencies, such as $Z_A^2$ and $Z_{A0}^2$, by Zheng et al. [2007].
2. Are the allele frequencies different in males and females? If yes, some of the test statistics have inflated type I error frequencies, such as $T_A$ and $T_{AD}$ by Clayton [2008].
3. Is the ratio of males to females different in cases and controls? If yes, this affects the relative power of different tests. If combined with sex-specific allele frequencies, tests $Z_A^2$, $T_A$ and $T_{AD}$ suffer from increased type I errors.
4. Is X chromosomal inactivation assumed? If yes, the statistic $T_A$ suggested by Clayton [2007] is likely to have more power. If no, the test $Z_A^2$ by Zheng et al. [2007] usually fares better.
5. Is the genetic model known? As in analyses for autosomal loci, if one specific genetic model is likely, the greatest power is achieved by choosing the respective test statistic.
In the previous sections, we have outlined the pipeline for the analysis of chromosome X in the context of a typical GWAS. Based on our own experience, we described the technical issues to consider together with tentative recommendations. Importantly, the overall coverage and genotyping quality seems to be lower on chromosome X, but the special genetic situation can be handled with sensible selection of calling, imputation, and association algorithms. Although it needs to be acknowledged that there is no single best way for the analysis, including X chromosomal variants in the analysis is thus manageable. We hope that our descriptions are an encouragement for further analysts to include chromosome X in the GWAS.

An important hindrance in the endeavor of including chromosome X in GWAS by default is that it cannot, at the moment, simply run with the autosomal analyses. For genotype calling, different models should be considered for males and females. Similarly, for imputation on chromosome X, additional parameters need to be set. Finally, although the statistical techniques are straightforward, analysis tools for association on chromosome X are rarely implemented in typical packages such as PLINK [Purcell et al., 2007] and SNPTEST [Marchini et al., 2007]. In PLINK, the test for the additive model without inactivation is implemented, in SNPTEST some tests are implemented in the new beta-version of the program but there is no description how these are calculated.
Recommendations for X chromosomal analysis

**Sampling:**
- Balanced if possible
- Sufficiently large subgroup (of male/female case and controls)

**Genotyping:**
- Check number of SNPs available on SNP chip array

**Genotype calling:**
- Take sex information into account

**Quality control on genotyped SNPs:**
- Sex-specific filters for minor allele frequency
- Sex-specific missingness filters
- See Table 1

**Imputation:**
- IMPUTE2 (v2.3.0):
  - Use additional flag `-chrX`
  - For pseudo-autosomal regions use flags `-chrX` and `-Xpar`
- Other imputation software:
  - Impute males and females separately

**Association analysis:**
- Check for sex-specific allele frequencies:
  - Equal:
    - Inactivation:
      - Use $T_A$
    - No Inactivation:
      - Use $Z_A^n$ if HWE holds
      - Use $Z_{mFG}^n$ if departure from HWE is observed
  - Sex-specific:
    - Use $Z_{mFG}^n$

**Meta-analysis:**
- Inactivation:
  - Logistic regression with covariate sex
    - Males coded 0 and 2 (for 0 and 1 risk allele, respectively)
  - No Inactivation:
    - Logistic regression with covariate sex
    - Males coded 0 and 1 (for 0 and 1 risk allele, respectively)

---

In conclusion, we have shown that by selection of sensible algorithms and parameter settings, the inclusion of chromosome X in GWAS is manageable. Including these settings by default in standard software solutions would greatly facilitate the analysis of chromosome X. Closing this gap is expected to further elucidate the genetic background of complex diseases, especially of those with sex-specific features.

**Acknowledgments**

I.R.K. receives funding from the German Science Foundation (grant number: KO 2250/4-1). J.E. received funding from the German Science Foundation and the German Federal Ministry of Education and Research (BMBF) in the context of the German National Genome Research Network (NGFN-2 (01GS0417) and NGFN-plus (01GS0832)), the FP6 and FP7 EU-funded integrated projects Cardiogenics (LSHM-CT-2006-037593) and ENGAGE (201413), the binational BMBF/ANR funded project CARDomics (01KU0908A), and the local focus programs “Kardiovaskuläre
Genomforschung" and "Medizinische Genetik" of the Universität zu Lübeck. A.Z. acknowledges funding from the German Science Foundation Excellence Cluster "Inflammation at Interfaces," the European Union (BiomarCare, grant number: HEALTH-2011-278913), and the German Ministry of Education and Research (CARDomics, grant numbers: 01KU9080A and 01KU9080B).

The authors have declared no conflict of interest.

References

Affymetrix. 2006. BRLMM: An Improved Genotype Calling Method for the GeneChip Human Mapping 500K Array Set. http://media.affymetrix.com/support/technical/whitepapers/brlmm_whitepaper.pdf.

Bushby KM. 1992. Genetic and clinical correlations of Xp21 muscular dystrophy. J Inherit Metab Dis 15:551–564.

CARDioGRAMplusC4D, Consortium Deloukas P, Kanoni S, Willenborg C, Farrall M, Assimes TL, Thompson JR, Ingelsson E, Saleheen D, Erdmann J and others. 2013. Large-scale association analysis identifies new risk loci for coronary artery disease. Nat Genet 45:25–33.

Carvalho B, Bengtsson H, Speed TP, Irizarry RA. 2007. Exploration, normalization, and genotype calls of high-density oligonucleotide SNP array data. Biostatistics 8:485–499.

Clayton D. 2008. Testing for association on the X chromosome. Biostatistics 9:593–600.

Erdmann J, Grosshennig A, Braund PS, König IR, Hengstenberg C, Hall AS, Linsel-Nitschke P, Kathiresan S, Wright B, Tregouet DA and others. 2009. New susceptibility locus for coronary artery disease on chromosome 3q22.3. Nat Genet 41:280–282.

Giannoulatou E, Yau C, Colella S, Ragoussis J, Holmes CC. 2008. GenoSNP: a variational Bayes within-sample SNP genotyping algorithm that does not require a reference population. Bioinformatics 24:2209–2214.

Hickey PF, Bahlo M. 2011. X chromosome association testing in genome wide association studies. Genet Epidemiol 35:664–670.

Hindorff LA, MacArthur JD, Morales J, Junkins H, Hall P, Klemm A, Manolio TA. 2013. A Catalog of Published Genome-Wide Association Studies. www.genome.gov/gwastudies.

Howie B, Fuchsberger C, Stephens M, Marchini J, Abecasis GR. 2012. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. Nat Genet 44:955–959.

Howie GR, Donnelly P, Marchini J. 2009. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. PLoS Genet 5:e1000529.

Ionesescu VV, Trofatter J, Haines JL, Summers AM, Ionescu-S倒塌, Searby C. 1991. Heterogeneity in X-linked recessive Charcot-Marie-Tooth neuropathy. Am J Hum Genet 48:1075–1083.

Kermari B. 2008. Artificial intelligence and global normalization methods for genotyping. http://www.patentstorm.us/patents/7467117/fulltext.html.

Korn J, Myers S, McVean G, Donnelly P. 2007. A new multipoint method for genome-wide association studies by imputation of genotypes. Nat Genet 39:906–913.

Marques Pereira P, Schneider A, Pannieret S, Heron D, Hanauer A. 2010. Coffin-Lowry syndrome. Eur J Hum Genet 18:627–633.

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ and others. 2007. PLINK: a toolset for whole-genome association and population-based linkage analysis. Am J Hum Genet 81: 559–575.

R Core Team for the R Foundation for Statistical Computing. 2013. R: A Language and Environment for Statistical Computing. Computing RFIS. Vienna, Austria. http://www.R-project.org.

Ritchie ME, Liu R, Carvalho BS, Irizarry RA. 2011. Comparing genotyping algorithms for Illumina’s Infinium whole-genome SNP BeadChips. BMC Bioinformatics 12:68.

Samani NJ, Erdmann J, Hall AS, Hengstenberg C, Mangino M, Mayer B, Dixon RI, Meitinger T, Braund P, Wichmann H-E and others. 2007. Genome-wide association analysis of coronary artery disease. N Engl J Med 357:443–453.

Scheet P, Stephens M. 2006. A fast and flexible statistical model for large-scale population genotype data: applications to inferring missing genotypes and haplotypic phase. Am J Hum Genet 78:629–644.

Schunkert H, König IR, Kathiresan S, Reilly MP, Assimes TL, Holm H, Preuss M, Stewart AFR, Barbalic M, Gieger C and others. 2011. Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease. Nat Genet 43:333–338.

Servin B, Stephens M. 2007. Imputation–based analysis of association studies: candidate regions and quantitative traits. PLoS Genet 3:e114.

Solimene MC. 2010. Coronary heart disease in women: a challenge for the 21st century. Clinics (Sao Paulo) 65:99–106.

Teo YY, Jonuie M, Small KS, Gwiliam R, Deloukas P, Kwiatkowski DP, Clark TG. 2007. A genotype calling algorithm for the Illumina BeadArray platform. Bioinformatics 23:2741–2746.

The 1000 Genomes Project Consortium. 2012. An integrated map of genetic variation from 1,092 human genomes. Nature 491:56–65.

Thye T, Vannberg FO, Wong SH, Owusu-Dabo E, Osei I, Gyapong J, Sirugo G, Sisay-Joof F, Enulil A, Chimbah MA and others. 2010. Genome-wide association analyses identifies a susceptibility locus for tuberculosis on chromosome 18q11.2. Nat Genet 42:739–741.

Timminn C, Thye T, Vens M, Evans J, May J, Ehmen C, Sievertsen J, Muntab B, Ruge G, Loag W and others. 2012. Genome-wide association study indicates two novel resistance loci for severe malaria. Nature 489:443–446.

Turner S, Armstrong LL, Bradford Y, Carlson CS, Crawford DC, Crenshaw AP, de Andrade M, Dohery KR, Haines J, Hayes G and others. 2011. Quality control procedures for genome-wide association studies. Curr Protoc Hum Genet 68:1.9.1–1.19.18.

Wise AL, Gy, L, Manolio TA. 2013. eXclusion: toward integrating the X chromosome in genome-wide association analyses. Am J Hum Genet 92:843–847.

Zheng G, Prendin B, Li Z, Gastwirth JL. 2003. Choice of scores in trend tests for case-control studies of candidate-gene associations. Biometrics 6:335–348.

Zheng G, Joo J, Zhang C, Geller NL. 2007. Testing association for markers on the X chromosome. Genet Epidemiol 31:834–843.

Ziegler A. 2009. Genome-wide association studies: quality control and population-based measures. Genet Epidemiol 33(Suppl 1):S45–S50.

Ziegler A, König IR. 2010. A Statistical Approach to Genetic Epidemiology. Concepts and Applications. Weinheim: Wiley-VCH.

Ziegler A, König IR, Thompson JR. 2008. Biostatistical aspects of genome-wide association studies. Biomed J 5:508–528.