OXFORD

Computer Note

XWAS: A Software Toolset for Genetic Data Analysis and Association Studies of the X Chromosome

Feng Gao*, Diana Chang*, Arjun Biddanda*, Li Ma, Yingjie Guo, Zilu Zhou, and Alon Keinan

From the Department of Biological Statistics and Computational Biology, Cornell University, Ithaca, NY 14853 (Gao, Chang, Biddanda, Ma, Guo, Zhou, and Keinan); Program in Computational Biology and Medicine, Cornell University, Ithaca, NY 14853 (Chang and Keinan); Department of Animal and Avian Sciences, University of Maryland, College Park, MD 20740 (Ma); and School of Computer Science and Technology, Harbin Institute of Technology, Harbin, Heilongjiang 150001, China (Guo).

*These authors contributed equally to this work.

Address correspondence to Alon Keinan at the address above, or e-mail: ak735@cornell.edu.

Abstract

XWAS is a new software suite for the analysis of the X chromosome in association studies and similar genetic studies. The X chromosome plays an important role in human disease and traits of many species, especially those with sexually dimorphic characteristics. Special attention needs to be given to its analysis due to the unique inheritance pattern, which leads to analytical complications that have resulted in the majority of genome-wide association studies (GWAS) either not considering X or mishandling it with toolsets that had been designed for non-sex chromosomes. We hence developed XWAS to fill the need for tools that are specially designed for analysis of X. Following extensive, stringent, and X-specific quality control, XWAS offers an array of statistical tests of association, including: 1) the standard test between a SNP (single nucleotide polymorphism) and disease risk, including after first stratifying individuals by sex, 2) a test for a differential effect of a SNP on disease between males and females, 3) motivated by X-inactivation, a test for higher variance of a trait in heterozygous females as compared with homozygous females, and 4) for all tests, a version that allows for combining evidence from all SNPs across a gene. We applied the toolset analysis pipeline to 16 GWAS datasets of immune-related disorders and 7 risk factors of coronary artery disease, and discovered several new X-linked genetic associations. XWAS will provide the tools and incentive for others to incorporate the X chromosome into GWAS and similar studies in any species with an XX/XY system, hence enabling discoveries of novel loci implicated in many diseases and in their sexual dimorphism.

Subject areas: Bioinformatics and computational genetics; Quantitative genetics and Mendelian inheritance
Key words: chromosome X, complex diseases, genetic association study, GWAS, sexual dimorphism, software

Genome-wide association studies (GWAS) have identified thousands of loci underlying complex human diseases and other complex traits (Welter et al. 2014). While successful for the autosomes (nonsex chromosomes), the vast majority of these studies have either incorrectly analyzed or ignored the X chromosome (X) (Wise et al. 2013). In most studies, all variants on the X have been removed

as a consequence of the quality control (QC) procedures (Mailman et al. 2007; Wise et al. 2013; Chang et al. 2014; Tryka et al. 2014). Many other studies that did analyze the X chromosome incorrectly applied methods that have been designed for the autosomes, without accounting for the analytical problems arising from X's unique mode of inheritance and its consequent population genetic and evolutionary patterns (Hammer et al. 2008; Wilson and Makova 2009; Emery et al. 2010; Hammer et al. 2010; Keinan and Reich, 2010; Lambert et al. 2010; Lohmueller et al. 2010; Arbiza et al. 2014). As a result, the role X plays in complex diseases and traits remains largely unknown.

Many human diseases commonly studied in GWAS show sexual dimorphism, including autoimmune diseases (Voskuhl 2011), cardiovascular diseases (Lerner and Kannel 1986) and cancer (Muscat et al. 1996; Matanoski et al. 2006), which suggests a potential contribution of the X chromosome (Carrel and Willard 2005; Ober et al. 2008). Several recent studies have examined this issue and demonstrated the potential value of analyzing X (Chang et al. 2014; Gilks et al. 2014; Tukiainen et al. 2014; Ma et al. 2015; Li YR, unpublished data). However, while association methods, QC and analysis pipelines are well established for the autosomes, respective pipelines for X-linked data are not readily available. Hence, in this article, we introduce the software package XWAS (chromosome X-Wide Analysis toolSet), which is tailored for analysis of genetic variation on X in human and other species with an XX/XY system. It implements extensive functionality that carries out QC specially designed for the X chromosome, statistical tests of single-marker association that account for its unique mode of inheritance, gene-based tests of association, and additional distinct tests only applicable to X that capitalize on its mode of inheritance. In implementing these features, the toolset builds on—and complements—the commonly used PLINK (Purcell et al. 2007) software. It includes many novel features that can facilitate X-wide association studies that are not available in PLINK and, to the best of our knowledge, in any other software. Combined, the XWAS toolset integrates the X chromosome into GWAS as well as into the next generation of sequence-based association studies and into studies of other species.

Features and Functionality

Quality Control Procedures

The XWAS toolset implements a whole pipeline for performing QC on genotype data for the X chromosome. The pipeline first follows standard GWAS QC steps as implemented in PLINK (Purcell et al. 2007) and SMARTPCA (Price et al. 2006) by running these tools. These include the removal of both individual samples and SNPs (single nucleotide polymorphisms) according to multiple criteria. Specifically, samples are removed based on 1) relatedness, 2) high genotype missingness rate, and 3) genetic ancestry differing from the majority of the samples (Price et al. 2006). SNPs are removed based on criteria such as their missingness rate, their minor allele frequency (MAF), and deviation from Hardy–Weinberg equilibrium (HWE). While the toolset is currently focused on case–control GWAS (binary traits), the entire QC pipeline is also applicable to GWAS of quantitative traits. One filter applied only to binary traits is the removal of SNPs for which missingness is correlated with the trait, that is, with case or control status (*--test-missing*).

To consider differences in genotyping between hemizygous males and diploid females, XWAS applies all the aforementioned QC steps

of samples separately for males and females. Consequently, a unified dataset is generated for subsequent analyses that include all SNPs and individuals passing the above filtering criteria in both the male and female QC groups.

The pipeline then applies X-specific QC steps, which are exclusively built into XWAS, to the unified dataset. These include 1) removing SNPs with significantly different MAF between male and female samples in the control group (*--freqdiff-x*), 2) removing SNPs with significantly different missingness rates between male and female controls (*--missdiff-x*), and 3) the removal of SNPs in the pseudoautosomal regions (PARs). The first 2 of these steps capture problems in genotype calling when plates include both males and females (Korn et al. 2008). Further details regarding specific QC procedures can be found in the user manual that is available with the toolset.

Single-Marker Association Testing on the X Chromosome

For an X-linked SNP, while females have 0, 1, or 2 copies of an allele, hemizygous males have at most one copy. Via the process of X-inactivation, 1 of the 2 copies in females is usually transcriptionally silenced. If X-inactivation is complete, it produces monoallelic expression of X-linked protein-coding genes in females. Therefore, when considering loci that undergo complete X-inactivation, it may be apt to consider males as having 0/2 alleles, corresponding to the female homozygotes (the FM_{02} test). The toolset carries out this test for association between a SNP and disease risk by using the *--xchr-model 2* option in PLINK (Purcell et al. 2007). For other scenarios though, including where some genes on the X escape X-inactivation or different genes are inactivated in different cells, it can be more indicative to code males as having 0/1 alleles. Hence, the toolset further carries out such an association test (FM_{01} test) of a SNP by using the following options in PLINK (Purcell et al. 2007): *--logistic* and *--linear* for binary and quantitative traits, respectively.

All tests, including tests described in following sections, allow for covariates such as population structure, sex, and traits that are correlated with the disease, as commonly considered in GWAS. We suggest calculating principal components by using EIGENSTRAT (Price et al. 2006) and include them as covariates to control for population structure. Ten such principal components are considered by default, unless otherwise specified. Any other user-defined covariates can also be incorporated.

Single-Marker Sex-Stratified Analysis on the X Chromosome

The XWAS software further includes new tests that are not included in PLINK. First, we implemented a new sex-stratified test, FM_{comb}, which is particularly relevant for X analyses since SNPs and loci on the sex chromosomes are potentially more likely to exhibit different effects on disease risk between males and females. In such scenarios, as well as in scenarios where the effect is only observed in one sex, a sex-stratified test as described in the following can be better powered. This functionality is accessible by the option *--strat-sex*. The FM_{comb} test first carries out an association test separately in males and females and then combines the results of the 2 tests to obtain a final sex-stratified significance level. The combination of the 2 test statistics is implemented using both Fisher's method (*--fishers*) (Fisher 1925) (in the $FM_{F.comb}$ test) and Stouffer's method (*--stouffers*) (Stouffer 1949) (in the $FM_{S.comb}$ test).

Each of these 2 tests is more powerful in different scenarios (Chang et al. 2014), for example, $FM_{F.comb}$ allows the SNP tested to have different, even an opposite, effect on disease risk in males and females. $FM_{F.comb}$ is also insensitive to whether males are coded as 0/2 (as in the FM_{02} test) or as 0/1 (as in the FM_{01} test), thus making no assumptions regarding X-inactivation status. Alternatively, $FM_{S.comb}$ directly accounts for the potentially differing sample sizes between males and females to maximize power. For this latter test, XWAS weighs by the sample size in males and females in cases and controls following the approach of Willer et al. (2010).

Single-Marker Sex-Differentiated Effect Size Test on the X Chromosome

We described above sex-stratified tests that accommodate associations with different effect size between males in females. In another type of test (FM_{diff}), we directly test whether the effect size is different between the sexes. This test, applied to each SNP, runs a *t*-test to test for difference between the odds ratio (OR) in males alone and the OR in females alone, while accounting for hemizygosity in males. This test is implemented under the *--sex-diff* option and we further described it in Chang et al. (2014). For this test and the sex-stratified test introduced in the previous section, both odds ratios and regression coefficients in each sex can be provided as output for further examination.

Single-Marker Variance-Based Testing Informed by X-Inactivation in Females

During X-inactivation, the expression of one copy of the X chromosome in females is randomly silenced, thereby increasing variation in the expression of X-linked quantitative trait loci (QTL). Specifically, females that are heterozygotes for a QTL might exhibit higher phenotypic variance than homozygous females since one or the other allele might be more dominantly affecting the phenotype in each given female heterozygote, such that for some individuals the QTL expression is more similar to one type of female homozygous, while to the other type in other individuals. We developed a test aimed at capturing this increased variance as a means for detecting X-linked QTLs in females. An overview of the test and its implementation follows, while we refer readers to Ma et al. (2015) for a full description of the test. This test (F_{var}) is implemented under the *--var-het* option. Although this F_{var} test is implemented for quantitative traits, it can be generalized to qualitative traits by applying liability threshold modeling (Zaitlen et al. 2012) to transform disease status to an unobserved continuous liability.

The null hypothesis of the F_{var} test is that phenotypic variances of the 3 genotypic groups of a SNP with 0, 1, or 2 copies of an allele are all equal. The alternative hypothesis is that female heterozygotes show a higher phenotypic variance than others. Hypothesis testing is carried out using a modified Brown-Forsythe test of variances (Brown and Forsythe 1974). We first normalize the phenotypic value and remove the effects of possible covariates by a linear regression as conventionally done, namely $y = \mu + XB + e$, where y is a vector of quantitative trait levels, μ is the population mean, X is the matrix of possible covariates, and e is a vector of residuals. Assume $y_{ilg=j}$ is the phenotypic value of the ith individual in the jth genotypic group and $z_{ilg=j} = |e_{ilg=j}|$ is the absolute residual value of the ith individual in the jth genotypic group ($j = 0, 1$, or 2 copies of an allele of a SNP). A test statistic is derived as

$$T_{var} = \frac{\overline{z_1} - \overline{z_{0/2}}}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_{0/2}^2}{n_0 + n_2}}}$$

where $\overline{z_1}$ is the sample mean of $z_{ilg=1}$ over i, $\overline{z_{0/2}}$ is the sample mean of $z_{ilg=0}$ and $z_{ilg=2}$ combined, s_1^2 and $s_{0/2}^2$ are the sample variances, respectively, and n_j is the sample size of the jth genotypic group. Under the null hypothesis, the statistic follows a *t*-distribution with degrees of freedom given by $df = \dfrac{\left(s_1^2 / n_1 + s_{0/2}^2 / (n_0 + n_2)\right)^2}{(s_1^2 / n_1)^2 / (n_1 - 1) + \left(s_{0/2}^2 / (n_0 + n_2)\right)^2 / (n_0 + n_2 - 1)}$.

This variance-based test captures a novel signal of X-linked associations by directly testing for higher phenotypic variance in heterozygous females than homozygotes. As a test of variance it is generally less powerful than standard tests of association that consider means; however, it provides an independent and complementary test to the standard association test for QTLs on X (Ma et al. 2015).

X-Linked Gene-Based Testing

XWAS also includes unique features for carrying out gene-based association analysis on the X chromosome. Gene-based approaches may be better powered to discover associations than single-marker analyses in cases of a gene with multiple causal variants of small effect size, or of multiple markers that are each in incomplete linkage disequilibrium with underlying causal variant/s. Furthermore, in studying the effect of X on sexual dimorphism in complex disease susceptibility, it may be desirable to analyze whole-genes or all genes of a certain function combined based on their unique function or putatively differential effect between males and females, as we illustrated in Chang et al. (2014).

The XWAS package determines the significance of association between each gene as a whole and disease risk by implementing a gene-level test statistic that combines individual SNP-level test statistics for all SNPs in and around each studied gene. This gene-level approach is applicable to any of the different tests described above. For instance, beyond tests of association, it can be applied to the sex-differentiated tests. In this case the gene-based test captures any scenario whereby SNPs within the gene display different effects in males and females, without restricting such differential effects to be of a similar nature across SNPs. By default, genes are considered from the UCSC browser "knownCanonical" transcript ID track. SNPs are mapped to a gene if they are in the gene or within 15 kb of the gene's start or end positions. The user can also provide a different set of gene definitions or alternate regions of interest and a different window length around them in which SNPs are also to be considered.

Combining SNP statistics across a gene is implemented in the general framework of Liu et al. (2010). Specifically, the significance of an observed gene-based test statistic is assessed from the distribution of test statistics that is expected given the linkage disequilibrium between the SNPs in the gene. In Liu et al. (2010), the test statistics for all SNPs in the gene are summed. Here, we have implemented a slight modification to this procedure, whereby we combined SNP-based *P*-values with either the truncated tail strength (Jiang et al. 2011) or the truncated product (Zaykin et al. 2002) method, which have been suggested to be more powerful in some scenarios (Zaykin et al. 2002; Ma et al. 2013).

To determine significance, XWAS follows the procedure in Liu et al. (2010). The observed statistic is compared to gene-level test statistics obtained when SNP-level statistics are randomly drawn from a multivariate normal distribution with a covariance determined by the empirical linkage disequilibrium between the SNPs in the tested gene. The significance level is then the proportion, out of many such drawings, for which this sampled gene-level statistic is more, or as extreme compared with the empirical one. For computational efficiency, the

number of drawings is determined adaptively as in Liu et al. (2010). By combining the truncated tail measures with this procedure, our new gene-based method combines the test statistics from multiple SNPs that show relatively low *P*-values, while also accounting for the dependency between these *P*-values due to linkage disequilibrium between the SNPs. Such a gene-based *P*-value is estimated for each gene and for each of the X-linked tests described above.

Examples of Use

In this section, we summarize several sets of results obtained using the XWAS software and publicly available GWAS datasets. For many of the results, we include herein a brief description of the main results, with the full description appearing in our separate papers (Chang et al. 2014; Ma et al. 2015). All associations presented herein are significant and details regarding the *P*-values can be found in the respective articles.

Association of X-Linked SNPs with Autoimmune Diseases

We applied the XWAS software described above to 16 GWAS datasets of autoimmune disease and other disorders with a potential autoimmune-related component. These include the following datasets that we obtained from dbGaP (Mailman et al. 2007; Tryka et al. 2014): ALS Finland (Laaksovirta et al. 2010) (phs000344), ALS Irish (Cronin et al. 2008) (phs000127), Celiac disease CIDR (Ahn et al. 2012) (phs000274), MS Case Control (Baranzini et al. 2009) (phs000171), Vitiligo GWAS1 (Jin et al. 2010) (phs000224), CD NIDDK (Duerr et al. 2006) (phs000130), CASP (Nair et al. 2009) (phs000019), and T2D GENEVA (Qi et al. 2010) (phs000091). Similarly, we obtained the following datasets from the Wellcome Trust Case Control Consortium (WT): all WT1 (The Wellcome Trust Case Control Consortium 2007) datasets, WT2 ankylosing spondylitis (Evans et al. 2011), WT2 ulcerative colitis (UK IBD Genetics Consortium et al. 2009) and WT2 multiple sclerosis (International Multiple Sclerosis Genetics Consortium et al. 2011). Finally, we also analyzed data from Vitiligo GWAS2 (Jin et al. 2012). These datasets are described in more detail in Chang et al. (2014).

Following application of the QC pipeline as described above, we applied the SNP-level FM_{02}, $FM_{F.comb}$, and $FM_{S.comb}$ tests to all SNPs in each of the 16 datasets. Based on the Vitiligo GWAS1 datasets, we associated SNPs in a region 17 kilobases (kb) away from the retrotransposed gene retro-*HSPA8* with risk of vitiligo. The parent of this retrotransposed gene, *HSPA8* on chromosome 11, encodes a member of the heat shock protein family, which has been previously associated to vitiligo (Mosenson et al. 2012; Abdou 2013; Mosenson et al. 2013). We discovered another association in WT2 ulcerative colitis of SNPs in an intron of *BCOR* contributing to ulcerative colitis disease risk. BCOR indirectly mediates apoptosis via co-repression of *BCL-6* (Huynh et al. 2000).

Association of Whole X-Linked Genes with Autoimmune Diseases

We next focused on a gene-based analysis of the X chromosome by using the SNP-level results of all the 3 tests in the above results as a basis for gene-based tests in the same 16 datasets. This analysis led to the discovery of the first X-linked gene-based associations with any disease or trait, which supports the utility of the XWAS package in facilitating such analyses. We associated in Vitiligo GWAS1 and replicated in Vitiligo GWAS2 an association between the gene *FOXP3* and vitiligo disease risk, in support of an earlier candidate gene study (Birlea et al. 2011). We also found a novel association of *ARHGEF6* to Crohn's disease and further replicated it in ulcerative colitis, another inflammatory bowel disorder (IBD). ARHGEF6 binds to a surface protein of a gastric bacterium (*Helicobacter pylori*) that has been associated to IBD (Luther et al. 2010; Jin et al. 2013). Finally, we associated *CENPI* as contributing to the risk of 3 different diseases (amyotrophic lateral sclerosis, celiac disease, and vitiligo). Other, autosomal genes in the same family as *CENPI* have previously been associated to amyotrophic lateral sclerosis (Ahmeti et al. 2013) as well as multiple sclerosis (Baranzini et al. 2009), supporting an involvement of *CENPI* with autoimmunity in general.

X-Linked SNPs Showing Sex-Differentiated Effect Size with Autoimmune Disease

As a final analysis on the 16 autoimmune datasets, we applied the FM_{diff} test and its gene-based version. Based on this test, we discovered and replicated the gene *C1GALT1C1* (also known as *Cosmc*) as exhibiting sex-differentiated effect size in risk of IBD. *C1GALT1C1* is necessary for the synthesis of many O-glycan proteins (Ju and Cummings 2005), which are components of antigens. We further found *CENPI*, which we previously associated with several diseases, to show significantly different effects in males and females in the same diseases as in the association analysis.

Increased Variance of Systolic Blood Pressure in Heterozygous Females for an X-Linked SNP

As an example application of the variance-based testing (F_{var}) informed by X-inactivation, we considered data on 7 quantitative traits from the Atherosclerosis Risk in Communities (ARIC) study (Williams 1989) along with Affymetrix 6.0 data from the participating individuals, which included 34 527 X-linked SNPs. First, we applied the entire set of QC procedures implemented in XWAS for quantitative traits. Then, we applied our single-marker variance-based testing and compared with application of standard testing for a QTL. Across the 7 traits, we found 1 SNP with a significant association based on the variance test (Ma et al. 2015). Importantly, the signals of this test are not in the same loci as those of the standard test, in line with them capturing different types of signals. Specifically, the significant SNP, rs4427330, which is associated with systolic blood pressure based on the variance test, is not associated with any trait based on the standard test. It is located upstream of *AFF2*, which might regulate *ATRX*. *ATRX* is associated with alpha-thalassemia, a disease that can cause anemia and has been associated with hypertension (Bowie et al. 1997).

Implementation and Availability

The XWAS software package is implemented in C++ and includes in part functions from open-source PLINK (Purcell et al. 2007). XWAS uses the same input format as PLINK. Beyond C++, additional features are also implemented in scripts, including in shell (for QC), Perl (for converting file formats and using SMARTPCA), and R (for gene-based testing). The entire package is freely available for download from http://keinanlab.cb.bscb.cornell.edu and includes 1) scripts, 2) the binary executable XWAS, 3) all source code with a Makefile, 4) a user manual, and 5) example data and examples of running the different options offered by the package. Additional help is provided via the *--xhelp* option. The XWAS toolset was initially designed and optimized for Linux systems, hence

exhibits best performance in such systems. A Makefile is also provided to facilitate local compilation on Linux environments, and can also be adjusted for Windows and MAC OS by revising a few lines indicated therein.

Conclusions

We have developed XWAS, an extensive toolset that facilitates the inclusion of the X chromosome in GWAS. It offers X-specific QC procedures, a variety of X-adapted tests of association, and an X-specific test of variance testing, available for both single-marker and gene-based statistics. We applied this toolset to successfully discover and replicate a number of genes with autoimmune disease risk and blood pressure.

We are continually developing the software and upcoming versions in the near future will offer additional features, including all features needed to conduct an extensive association study of quantitative traits (many features for quantitative traits are already functional in the current version). Similarly, while imputation of unobserved SNPs is presently performed as a preprocessing step using IMPUTE2 (Howie et al. 2012), we will incorporate X-specific imputation as part of the pipeline. Additional features will include analysis of X-linked data from sequence-based association studies (including burden tests), statistical methods that have been previously designed for the X chromosome (Zheng et al. 2007; Clayton 2008, 2009; Loley et al. 2011; Thornton et al. 2012), additional tests we previously proposed based on the workings of X-inactivation (Ma et al. 2015), and tests for gene–gene interactions. Finally, we will incorporate information regarding whether or how often a gene undergoes or escapes X-inactivation (Carrel and Willard 2005; Cotton et al. 2011; Disteche 2012). For computational efficiency, we will also continually upgrade the functions of PLINK that XWAS uses to the most recent version.

This software, alone and through incorporation of additional features, can be used for other types of studies of the X chromosome beyond association studies, in particular population genetic studies. For instance, allele frequency output and testing for significant differences in allele frequency between males and females as currently implemented, can be utilized to search for signals of selection.

Considering the availability of unutilized data for the X chromosome from thousands of GWAS, and the additional X-linked data that is being generated as part of ongoing GWAS, many researchers will find extensive utility in the XWAS toolset. Furthermore, it is not limited to application to human data, but rather genetic data from all organisms with XX/XY sex determination system, including all mammals. XWAS will facilitate the proper analysis of these data, incorporate X into GWAS and enable discoveries of novel X-linked loci as implicated in many diseases and in their sexual dimorphism.

Funding

Acknowledgments

References

Abdou AG, Maraee AH, Reyad W. 2013. Immunohistochemical expression of heat shock protein 70 in vitiligo. *Ann Diagn Pathol*. 17:245–249.

Ahmeti KB, Ajroud-Driss S, Al-Chalabi A, Andersen PM, Armstrong J, Birve A, Blauw HM, Brown RH, Bruijn L, Chen W, *et al.* 2013. Age of onset of amyotrophic lateral sclerosis is modulated by a locus on 1p34.1. *Neurobiol Aging*. 34:357.e7–357.e19.

Ahn R, Ding YC, Murray J, Fasano A, Green PH, Neuhausen SL, Garner C. 2012. Association analysis of the extended MHC region in celiac disease implicates multiple independent susceptibility loci. *PLoS One*. 7:e36926.

Arbiza L, Gottipati S, Siepel A, Keinan A. 2014. Contrasting X-linked and autosomal diversity across 14 human populations. *Am J Hum Genet*. 94:827–844.

Baranzini SE, Wang J, Gibson RA, Galwey N, Naegelin Y, Barkhof F, Radue EW, Lindberg RL, Uitdehaag BM, Johnson MR, *et al.* 2009. Genome-wide association analysis of susceptibility and clinical phenotype in multiple sclerosis. *Hum Mol Genet*. 18:767–778.

Birlea SA, Jin Y, Bennett DC, Herbstman DM, Wallace MR, McCormack WT, Kemp EH, Gawkrodger DJ, Weetman AP, Picardo M, *et al.* 2011. Comprehensive association analysis of candidate genes for generalized vitiligo supports XBP1, FOXP3, and TSLP. *J Invest Dermatol*. 131:371–381.

Bowie LJ, Reddy PL, Beck KR. 1997. Alpha thalassemia and its impact on other clinical conditions. *Clin Lab Med*. 17:97–108.

Brown MB, Forsythe AB. 1974. Robust tests for equality of variances. *J Am Stat Assoc*. 69:364–367.

Carrel L, Willard HF. 2005. X-inactivation profile reveals extensive variability in X-linked gene expression in females. *Nature*. 434:400–404.

Chang D, Gao F, Slavney A, Ma L, Waldman YY, Sams AJ, Billing-Ross P, Madar A, Spritz R, Keinan A. 2014. Accounting for eXentricities: analysis of the X chromosome in GWAS reveals X-linked genes implicated in autoimmune diseases. *PLoS One*. 9:e113684.

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

Clayton DG. 2009. Sex chromosomes and genetic association studies. *Genome Med*. 1:110.

Cotton AM, Lam L, Affleck JG, Wilson IM, Penaherrera MS, McFadden DE, Kobor MS, Lam WL, Robinson WP, Brown CJ. 2011. Chromosome-wide DNA methylation analysis predicts human tissue-specific X inactivation. *Hum Genet*. 130:187–201.

Cronin S, Berger S, Ding J, Schymick JC, Washecka N, Hernandez DG, Greenway MJ, Bradley DG, Traynor BJ, Hardiman O. 2008. A genome-wide association study of sporadic ALS in a homogenous Irish population. *Hum Mol Genet*. 17:768–774.

Disteche CM. 2012. Dosage compensation of the sex chromosomes. *Annu Rev Genet*. 46:537–560.

Duerr RH, Taylor KD, Brant SR, Rioux JD, Silverberg MS, Daly MJ, Steinhart AH, Abraham C, Regueiro M, Griffiths A, *et al.* 2006. A genome-wide association study identifies IL23R as an inflammatory bowel disease gene. *Science*. 314:1461–1463.

Emery LS, Felsenstein J, Akey JM. 2010. Estimators of the human effective sex ratio detect sex biases on different timescales. *Am J Hum Genet*. 87:848–856.

Evans DM, Spencer CC, Pointon JJ, Su Z, Harvey D, Kochan G, Oppermann U, Dilthey A, Pirinen M, Stone MA, *et al.* 2011. Interaction between ERAP1 and HLA-B27 in ankylosing spondylitis implicates peptide handling in the mechanism for HLA-B27 in disease susceptibility. *Nat Genet*. 43:761–767.

Fisher RA. 1925. *Statistical methods for research workers*. Edinburgh (UK): Oliver and Boyd.

Gilks WP, Abbott JK, Morrow EH. 2014. Sex differences in disease genetics: evidence, evolution, and detection. *Trends Genet*. 30:453–463.

Hammer MF, Mendez FL, Cox MP, Woerner AE, Wall JD. 2008. Sex-biased evolutionary forces shape genomic patterns of human diversity. *PLoS Genet*. 4:e1000202.

Hammer MF, Woerner AE, Mendez FL, Watkins JC, Cox MP, Wall JD. 2010. The ratio of human X chromosome to autosome diversity is positively correlated with genetic distance from genes. *Nat Genet*. 42:830–831.

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.

Huynh KD, Fischle W, Verdin E, Bardwell VJ. 2000. BCoR, a novel corepressor involved in BCL-6 repression. *Genes Dev*. 14:1810–1823.

International Multiple Sclerosis Genetics Consortium, Wellcome Trust Case Control Consortium, Sawcer S, Hellenthal G, Pirinen M, Spencer CC, Patsopoulos NA, Moutsianas L, Dilthey A, Su Z, et al. 2011. Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis. *Nature*. 476:214–219.

Jiang B, Zhang X, Zuo Y, Kang G. 2011. A powerful truncated tail strength method for testing multiple null hypotheses in one dataset. *J Theor Biol*. 277:67–73.

Jin X, Chen YP, Chen SH, Xiang Z. 2013. Association between Helicobacter Pylori infection and ulcerative colitis—a case control study from China. *Int J Med Sci*. 10:1479–1484.

Jin Y, Birlea SA, Fain PR, Ferrara TM, Ben S, Riccardi SL, Cole JB, Gowan K, Holland PJ, Bennett DC, et al. 2012. Genome-wide association analyses identify 13 new susceptibility loci for generalized vitiligo. *Nat Genet*. 44:676–680.

Jin Y, Birlea SA, Fain PR, Gowan K, Riccardi SL, Holland PJ, Mailloux CM, Sufit AJ, Hutton SM, Amadi-Myers A, et al. 2010. Variant of TYR and autoimmunity susceptibility loci in generalized vitiligo. *N Engl J Med*. 362:1686–1697.

Ju T, Cummings RD. 2005. Protein glycosylation: chaperone mutation in Tn syndrome. *Nature*. 437:1252.

Keinan A, Reich D. 2010. Can a sex-biased human demography account for the reduced effective population size of chromosome X in non-Africans? *Mol Biol Evol*. 27:2312–2321.

Korn JM, Kuruvilla FG, McCarroll SA, Wysoker A, Nemesh J, Cawley S, Hubbell E, Veitch J, Collins PJ, Darvishi K, et al. 2008. Integrated genotype calling and association analysis of SNPs, common copy number polymorphisms and rare CNVs. *Nat Genet*. 40:1253–1260.

Laaksovirta H, Peuralinna T, Schymick JC, Scholz SW, Lai SL, Myllykangas L, Sulkava R, Jansson L, Hernandez DG, Gibbs JR, et al. 2010. Chromosome 9p21 in amyotrophic lateral sclerosis in Finland: a genome-wide association study. *Lancet Neurol*. 9:978–985.

Lambert CA, Connelly CF, Madeoy J, Qiu R, Olson MV, Akey JM. 2010. Highly punctuated patterns of population structure on the X chromosome and implications for African evolutionary history. *Am J Hum Genet*. 86:34–44.

Lerner DJ, Kannel WB. 1986. Patterns of coronary heart disease morbidity and mortality in the sexes: a 26-year follow-up of the Framingham population. *Am Heart J*. 111:383–390.

Liu JZ, McRae AF, Nyholt DR, Medland SE, Wray NR, Brown KM, AMFS Investigators, Hayward NK, Montgomery GW, Visscher PM, et al. 2010. A versatile gene-based test for genome-wide association studies. *Am J Hum Genet*. 87:139–145.

Lohmueller KE, Degenhardt JD, Keinan A. 2010. Sex-averaged recombination and mutation rates on the X chromosome: a comment on Labuda et al. *Am J Hum Genet*. 86:978–980; author reply 980–981.

Loley C, Ziegler A, Konig IR. 2011. Association tests for X-chromosomal markers—a comparison of different test statistics. *Hum Hered*. 71:23–36.

Luther J, Dave M, Higgins PD, Kao JY. 2010. Association between *Helicobacter pylori* infection and inflammatory bowel disease: a meta-analysis and systematic review of the literature. *Inflamm Bowel Dis*. 16:1077–1084.

Ma L, Clark AG, Keinan A. 2013. Gene-based testing of interactions in association studies of quantitative traits. *PLoS Genet*. 9:e1003321.

Ma L, Hoffman G, Keinan A. 2015. X-inactivation informs variance-based testing for X-linked association of a quantitative trait. *BMC Genomics*. 16:241.

Mailman MD, Feolo M, Jin Y, Kimura M, Tryka K, Bagoutdinov R, Hao L, Kiang A, Paschall J, Phan L, et al. 2007. The NCBI dbGaP database of genotypes and phenotypes. *Nat Genet*. 39:1181–1186.

Matanoski G, Tao X, Almon L, Adade AA, Davies-Cole JO. 2006. Demographics and tumor characteristics of colorectal cancers in the United States, 1998–2001. *Cancer*. 107(5 Suppl):1112–1120.

Mosenson JA, Eby JM, Hernandez C, Le Poole IC. 2013. A central role for inducible heat-shock protein 70 in autoimmune vitiligo. *Exp Dermatol*. 22:566–569.

Mosenson JA, Zloza A, Klarquist J, Barfuss AJ, Guevara-Patino JA, Poole IC. 2012. HSP70i is a critical component of the immune response leading to vitiligo. *Pigment Cell Melanoma Res*. 25:88–98.

Muscat JE, Richie JP Jr, Thompson S, Wynder EL. 1996. Gender differences in smoking and risk for oral cancer. *Cancer Res*. 56:5192–5197.

Nair RP, Duffin KC, Helms C, Ding J, Stuart PE, Goldgar D, Gudjonsson JE, Li Y, Tejasvi T, Feng BJ, et al. 2009. Genome-wide scan reveals association of psoriasis with IL-23 and NF-kappaB pathways. *Nat Genet*. 41:199–204.

Ober C, Loisel DA, Gilad Y. 2008. Sex-specific genetic architecture of human disease. *Nat Rev Genet*. 9:911–922.

Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. 2006. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet*. 38:904–909.

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, De Bakker PI, Daly MJ, et al. 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 81:559–575.

Qi L, Cornelis MC, Kraft P, Stanya KJ, Linda Kao WH, Pankow JS, Dupuis J, Florez JC, Fox CS, Pare G, et al. 2010. Genetic variants at 2q24 are associated with susceptibility to type 2 diabetes. *Hum Mol Genet*. 19:2706–2715.

Stouffer SA. 1949. *The American soldier*. Princeton (NJ): Princeton University Press.

The Wellcome Trust Case Control Consortium. 2007. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature*. 447:661–678.

Thornton T, Zhang Q, Cai X, Ober C, McPeek MS. 2012. XM: association testing on the X-chromosome in case-control samples with related individuals. *Genet Epidemiol*. 36:438–450.

Tryka KA, Hao L, Sturcke A, Jin Y, Wang ZY, Ziyabari L, Lee M, Popova N, Sharopova N, Kimura M, et al. 2014. NCBI's database of genotypes and phenotypes: dbGaP. *Nucleic Acids Res*. 42(Database issue):D975–D979.

Tukiainen T, Pirinen M, Sarin AP, Ladenvall C, Kettunen J, Lehtimaki T, Lokki ML, Perola M, Sinisalo J, Vlachopoulou E, et al. 2014. Chromosome X-wide association study identifies Loci for fasting insulin and height and evidence for incomplete dosage compensation. *PLoS Genet*. 10:e1004127.

UK IBD Genetics Consortium, Barrett JC, Lee JC, Lees CW, Prescott NJ, Anderson CA, Phillips A, Wesley E, Parnell K, et al. 2009. Genome-wide association study of ulcerative colitis identifies three new susceptibility loci, including the HNF4A region. *Nat Genet*. 41:1330–1334.

Voskuhl R. 2011. Sex differences in autoimmune diseases. *Biol Sex Differ*. 2:1.

Welter D, Macarthur J, Morales J, Burdett T, Hall P, Junkins H, Klemm A, Flicek P, Manolio T, Hindorff L, et al. 2014. The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Res*. 42(Database issue):D1001–D1006.

Willer CJ, Li Y, Abecasis GR. 2010. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics*. 26:2190–2191.

Williams OD. 1989. The Atherosclerosis Risk in Communities (Aric) Study—design and objectives. *Am J Epidemiol*. 129:687–702.

Wilson MA, Makova KD. 2009. Genomic analyses of sex chromosome evolution. *Annu Rev Genomics Hum Genet*. 10:333–354.

Wise AL, Gyi L, Manolio TA. 2013. eXclusion: toward integrating the X chromosome in genome-wide association analyses. *Am J Hum Genet*. 92:643–647.

Zaitlen N, Pasaniuc B, Patterson N, Pollack S, Voight B, Groop L, Altshuler D, Henderson BE, Kolonel LN, Le Marchand L, et al. 2012. Analysis of case-control association studies with known risk variants. *Bioinformatics*. 28:1729–1737.

Zaykin DV, Zhivotovsky LA, Westfall PH, Weir BS. 2002. Truncated product method for combining *P*-values. *Genet Epidemiol*. 22:170–185.

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