bioRxiv preprint doi: https://doi.org/10.1101/009795; this version posted April 1, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

XWAS: a software toolset for genetic data analysis and association studies of the X chromosome

Feng Gao[1,5], Diana Chang[1,2,5], Arjun Biddanda[1,5], Li Ma[1,3], Yingjie Guo[1,4], Zilu Zhou[1], and Alon Keinan[1,2,*]

[1]Department of Biological Statistics and Computational Biology, Cornell University, Ithaca, NY 14853, USA

[2]Program in Computational Biology and Medicine, Cornell University, Ithaca, NY 14853, USA

[3]Department of Animal and Avian Sciences, University of Maryland, College Park, MD 20740, USA

[4]School of Computer Science and Technology, Harbin Institute of Technology, Harbin, Heilongjiang 150001, China

[5]These authors contributed equally to this work

*To whom correspondence should be addressed: ak735@cornell.edu

bioRxiv preprint doi: https://doi.org/10.1101/009795; this version posted April 1, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Abstract

XWAS is a new software for the analysis of the X chromosome in association studies and similar studies. The X chromosome plays an important role in human disease, especially those with sexually dimorphic characteristics. Special attention needs to be given to its analysis due to the unique inheritance pattern, leading to analytical complications that have resulted in the majority of genome-wide association studies (GWAS) either not considering X or mishandling it with GWAS toolsets that have been designed for non-sex chromosomes.. Hence, XWAS fills 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 to homozygous females, and (4) for all tests, a version that allows for combining evidence across all SNPs in a whole gene. We applied the toolset analysis pipeline to 16 GWAS datasets of immune-related disorders and to 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, hence enabling discoveries of novel loci implicated in many diseases and in their sexual dimorphism.

Key words: GWAS, software, genetic association study, sexual dimorphism, complex diseases

bioRxiv preprint doi: https://doi.org/10.1101/009795; this version posted April 1, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Introduction

Genome-wide association studies (GWAS) have identified thousands of loci underlying complex human diseases and other complex traits (Welter, *et al.*, 2014). While very successful for the autosomes (non-sex chromosomes), the vast majority of these studies have either incorrectly analyzed or completely ignored the X chromosome (X) (Wise, Gyi and Manolio, 2013). In most studies, all variants on the X chromosomes have been removed as a consequence of the quality control (QC) procedures (Chang, *et al.*, 2014; Mailman, *et al.*, 2007; Tryka, *et al.*, 2014; Wise, *et al.*, 2013). 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 (i.e. males only bear one copy of X, while one of the two copies in females is transcriptionally silenced via X-inactivation) and consequentially from X's population genetics patterns and evolution forces that differs from that of autosome's (Emery, Felsenstein and Akey, 2010; Lambert, *et al.*, 2010; Wilson and Makova, 2009). 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 (Matanoski, *et al.*, 2006; Muscat, *et al.*, 1996), which suggests a potential contribution of X (Carrel and Willard, 2005; Ober, Loisel and Gilad, 2008). Several recent studies have examined this issue and further demonstrated the value of analyzing X (Chang, *et al.*, 2014; Gilks, Abbott and Morrow, 2014; Li, *et al.*; Ma, Hoffman and Keinan, 2015; Tukiainen, *et al.*, 2014). However, while association methods, QC and analysis pipelines are well established for analysis of autosomes, respective pipelines for X-linked data are not widely available. Hence, in

bioRxiv preprint doi: https://doi.org/10.1101/009795; this version posted April 1, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

this paper, we introduce the software package XWAS (chromosome X-Wide Analysis toolSet), which is tailored for analysis of genetic variation on X. It implements extensive functionality that carries out QC specially-designed for the X chromosome, statistical tests of single-marker association that account for the unique mode of inheritance, gene-based tests of association and additional distinct tests that capitalize on the mode of inheritance that are only applicable to X. In implementing these features, the toolset builds on--and complements--the commonly-used PLINK (Purcell, *et al.*, 2007) software. Combined, the XWAS toolset integrates X into GWAS as well as into the next generation of sequence-based association studies.

bioRxiv preprint doi: https://doi.org/10.1101/009795; this version posted April 1, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

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 (i) relatedness, (ii) high genotype missingness rate, and (iii) 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 for binary traits is the removal of SNPs which missingness correlates with trait, i.e. 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 (i) removing SNPs with significantly different MAF between male and female samples in the control group (*--freqdiff-x*), (ii) removing SNPs with significantly different missingness rates between male and female controls (*--missdiff-x*), and (iii) removal of

bioRxiv preprint doi: https://doi.org/10.1101/009795; this version posted April 1, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

SNPs in the pseudoautosomal regions (PARs). The first two 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 1 copy. 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 EIGENSTART (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.

bioRxiv preprint doi: https://doi.org/10.1101/009795; this version posted April 1, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Single-Marker Sex-stratified Analysis on the X chromosome

The XWAS software further includes new tests that are not a part of PLINK. First, we implemented a new sex-stratified test FM_{comb} that 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, a sex-stratified test as described in the following can be better powered, as well as in scenarios where the effect is only observed in one sex. This functionality is accessible by the option *--strat-sex*. The FM_{comb} tests first carryout an association test separately in males and females and then combines the results of the two tests to obtain a final sex-stratified significance level. The combination of the two 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 two tests is more powerful in different scenarios (Chang, *et al.*, 2014), e.g. $FM_{F.comb}$ 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, Li and Abecasis, 2010).

Single-Marker Sex-differentiated Effect Size Test on the X chromosome

We described above sex-stratified tests that accommodate associations that have different effect size between males in females. In another type of test (FM_{diff}), we directly test whether the effect

bioRxiv preprint doi: https://doi.org/10.1101/009795; this version posted April 1, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

size is different between the sexes, which we find especially plausible for the X chromosome. This test, applied for 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 is further described 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 QTL 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 currently implemented under the --*var-het* option.

The null hypothesis of the F_{var} test is that phenotypic variances of the three genotypic groups of a SNP with 0, 1, or 2 copies of a reference 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

bioRxiv preprint doi: https://doi.org/10.1101/009795; this version posted April 1, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

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 $y_{i|g=j}$ is the phenotypic value of the *i*th individual in the *j*th genotypic group and $z_{i|g=j} = |e_{i|g=j}|$ is the absolute residual value of the *i*th individual in the *j*th genotypic group (j = 0, 1, or 2 copies of an allele of a SNP). A test statistic is derived as

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

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

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 Analysis

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

bioRxiv preprint doi: https://doi.org/10.1101/009795; this version posted April 1, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

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 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 were mapped to a gene if they are in the gene or within15 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

bioRxiv preprint doi: https://doi.org/10.1101/009795; this version posted April 1, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

(Zaykin, *et al.*, 2002) method, which have been suggested to be more powerful in some scenarios (Ma, Clark and Keinan, 2013; Zaykin, *et al.*, 2002).

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 to 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 p-value is estimated for each gene and for each of the X-linked tests described above.

bioRxiv preprint doi: https://doi.org/10.1101/009795; this version posted April 1, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

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 separate papers (Chang, *et al.*, 2014; Ma, *et al.*, 2015). All associations presented herein are significant and details of the variety of p-values can be found in the respective papers.

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

bioRxiv preprint doi: https://doi.org/10.1101/009795; this version posted April 1, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

datasets, we associated SNPs in a region 17 kilobases (kb) away from the retrotransposed gene retro-*HSPA8* with risk of viligo. 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 (Abdou, Maraee and Reyad, 2013; Mosenson, *et al.*, 2013; Mosenson, *et al.*, 2012). 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 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 three 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 (Jin, *et al.*, 2013; Luther, *et al.*, 2010). Finally, we associated *CENPI* as contributing to the risk of three 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.

bioRxiv preprint doi: https://doi.org/10.1101/009795; this version posted April 1, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

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 of application of the variance-based testing 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. Note that for an example use of this test we turned to a different dataset since this test is currently only applicable to quantitative traits. 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 one 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, Reddy and Beck, 1997).

bioRxiv preprint doi: https://doi.org/10.1101/009795; this version posted April 1, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Implementation and Availability

The XWAS software package is implemented in C++ and includes functions from PLINK (Purcell, *et al.*, 2007) . Additional features are also implemented in scripts, including in shell (for QC), Perl (for converting file formats and using SMARTPCA), and R (gene-based testing), as appropriate. 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, (4) a user manual, and (5) example data and run. Additional help is provided via the --*xhelp* option. The XWAS has been initially designed and optimized for Linux systems, hence exhibits best performance in such systems. However, the binary executable (but not scripts) is also available for Windows and MAC OS. The Makefile is written for Linux systems but it can be easily adjusted for Windows and MAC OS by changing several indicated lines therein.

bioRxiv preprint doi: https://doi.org/10.1101/009795; this version posted April 1, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Conclusions

We have developed an extensive toolset that facilitate the inclusion of the X chromosome in genome-wide association studies. 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 (Clayton, 2008; Clayton, 2009; Loley, Ziegler and Konig, 2011; Thornton, *et al.*, 2012; Zheng, *et al.*, 2007), tests for gene-gene interactions, and additional tests we previously proposed based on the workings of X-inactivation (Ma, *et al.*, 2015). Finally, we will directly account for whether or how often any gene undergo or escape X-inactivation (Carrel and Willard, 2005; Cotton, *et al.*, 2011; Disteche, 2012; Slavney, *et al.*). For computational efficiency, we will also upgrade to the most recent version of PLINK.

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 a part of ongoing GWAS, many researchers will find extensive utility in the XWAS toolset. It will facilitate the proper analysis of

bioRxiv preprint doi: https://doi.org/10.1101/009795; this version posted April 1, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

these data, incorporate X into GWAS and enable discoveries of novel X-linked loci as implicated in many diseases and in their sexual dimorphism.

bioRxiv preprint doi: https://doi.org/10.1101/009795; this version posted April 1, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Funding

This work was supported by a NIH grants to AK (R01HG006849 and R01GM108805), as well as by an award from The Ellison Medical Foundation to AK, and an award by The Edward Mallinckrodt, Jr. Foundation to AK. FG is a Howard Hughes Medical Institute (HHMI) International Student Research fellow.

Acknowledgements

We thank Paul Billing-Ross, Aviv Madar, Aaron Sams, Andrea Slavney, Richard Spritz, Yedael Y. Waldman and Liang Zhang for helpful comments on the software and previous versions of this manuscript.

bioRxiv preprint doi: https://doi.org/10.1101/009795; this version posted April 1, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

References

ABDOU AG, MARAEE AH, REYAD W. 2013 Immunohistochemical expression of heat shock protein 70 in vitiligo. Ann Diagn Pathol. 17(3):245-249.

AHMETI KB, AJROUD-DRISS S, AL-CHALABI A, ANDERSEN PM, ARMSTRONG J, BIRVE A, BLAUW HM, BROWN RH, BRUIJN L, CHEN W, CHIO A, COMEAU MC, CRONIN S, DIEKSTRA FP, SORAYA GKAZI A, GLASS JD, GRAB JD, GROEN EJ, HAINES JL, HARDIMAN O, HELLER S, HUANG J, HUNG WY, CONSORTIUM I, JAWORSKI JM, JONES A, KHAN H, LANDERS JE, LANGEFELD CD, LEIGH PN, MARION MC, MCLAUGHLIN RL, MEININGER V, MELKI J, MILLER JW, MORA G, PERICAK-VANCE MA, RAMPERSAUD E, ROBBERECHT W, RUSSELL LP, SALACHAS F, SARIS CG, SHATUNOV A, SHAW CE, SIDDIQUE N, SIDDIQUE T, SMITH BN, SUFIT R, TOPP S, TRAYNOR BJ, VANCE C, VAN DAMME P, VAN DEN BERG LH, VAN ES MA, VAN VUGHT PW, VELDINK JH, YANG Y, ZHENG JG, CONSORTIUM A. 2013 Age of onset of amyotrophic lateral sclerosis is modulated by a locus on 1p34.1. Neurobiol Aging. 34(1):357 e357-319.

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(5):e36926.

BARANZINI SE, WANG J, GIBSON RA, GALWEY N, NAEGELIN Y, BARKHOF F, RADUE EW, LINDBERG RL, UITDEHAAG BM, JOHNSON MR, ANGELAKOPOULOU A, HALL L, RICHARDSON JC, PRINJHA RK, GASS A, GEURTS JJ, KRAGT J, SOMBEKKE M, VRENKEN H, QUALLEY P, LINCOLN RR, GOMEZ R, CAILLIER SJ, GEORGE MF, MOUSAVI H, GUERRERO R, OKUDA DT, CREE BA, GREEN AJ, WAUBANT E, GOODIN DS, PELLETIER D, MATTHEWS PM, HAUSER SL, KAPPOS L, POLMAN CH, OKSENBERG JR. 2009 Genome-wide association analysis of susceptibility and clinical phenotype in multiple sclerosis. Hum Mol Genet. 18(4):767-778.

BIRLEA SA, JIN Y, BENNETT DC, HERBSTMAN DM, WALLACE MR, MCCORMACK WT, KEMP EH, GAWKRODGER DJ, WEETMAN AP, PICARDO M, LEONE G, TAIEB A, JOUARY T, EZZEDINE K, VAN GEEL N, LAMBERT J, OVERBECK A, FAIN PR, SPRITZ RA. 2011 Comprehensive association analysis of candidate genes for generalized vitiligo supports XBP1, FOXP3, and TSLP. J Invest Dermatol. 131(2):371-381.

BOWIE LJ, REDDY PL, BECK KR. 1997 Alpha thalassemia and its impact on other clinical conditions. Clin Lab Med. 17(1):97-108.

BROWN MB, FORSYTHE AB. 1974 Robust Tests for Equality of Variances. Journal of the American Statistical Association. 69(346):364-367.

CARREL L, WILLARD HF. 2005 X-inactivation profile reveals extensive variability in X-linked gene expression in females. Nature. 434(7031):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(12):e113684.

CLAYTON D. 2008 Testing for association on the X chromosome. Biostatistics. 9(4):593-600.

CLAYTON DG. 2009 Sex chromosomes and genetic association studies. Genome Med. 1(11):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(2):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(5):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, DASSOPOULOS T, BITTON A, YANG H, TARGAN S, DATTA LW, KISTNER EO, SCHUMM LP, LEE AT, GREGERSEN PK, BARMADA MM, ROTTER JI, NICOLAE DL, CHO JH. 2006 A genome-wide association study identifies IL23R as an inflammatory bowel disease gene. Science. 314(5804):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(6):848-856.

EVANS DM, SPENCER CC, POINTON JJ, SU Z, HARVEY D, KOCHAN G, OPPERMANN U, DILTHEY A, PIRINEN M, STONE MA, APPLETON L, MOUTSIANAS L, LESLIE S, WORDSWORTH T, KENNA TJ, KARADERI T, THOMAS GP, WARD MM, WEISMAN MH, FARRAR C, BRADBURY LA, DANOY P, INMAN RD, MAKSYMOWYCH W, GLADMAN D, RAHMAN P, SPONDYLOARTHRITIS RESEARCH CONSORTIUM OF C, MORGAN A, MARZO-ORTEGA H, BOWNESS P, GAFFNEY K, GASTON JS, SMITH M, BRUGES-ARMAS J, COUTO AR, SORRENTINO R, PALADINI F, FERREIRA MA, XU H, LIU Y, JIANG

bioRxiv preprint doi: https://doi.org/10.1101/009795; this version posted April 1, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

L, LOPEZ-LARREA C, DIAZ-PENA R, LOPEZ-VAZQUEZ A, ZAYATS T, BAND G, BELLENGUEZ C, BLACKBURN H, BLACKWELL JM, BRAMON E, BUMPSTEAD SJ, CASAS JP, CORVIN A, CRADDOCK N, DELOUKAS P, DRONOV S, DUNCANSON A, EDKINS S, FREEMAN C, GILLMAN M, GRAY E, GWILLIAM R, HAMMOND N, HUNT SE, JANKOWSKI J, JAYAKUMAR A, LANGFORD C, LIDDLE J, MARKUS HS, MATHEW CG, MCCANN OT, MCCARTHY MI, PALMER CN, PELTONEN L, PLOMIN R, POTTER SC, RAUTANEN A, RAVINDRARAJAH R, RICKETTS M, SAMANI N, SAWCER SJ, STRANGE A, TREMBATH RC, VISWANATHAN AC, WALLER M, WESTON P, WHITTAKER P, WIDAA S, WOOD NW, MCVEAN G, REVEILLE JD, WORDSWORTH BP, BROWN MA, DONNELLY P, AUSTRALO-ANGLO-AMERICAN SPONDYLOARTHRITIS CONSORTIUM, WELLCOME TRUST CASE CONTROL CONSORTIUM. 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(8):761-767.

FISHER RA. 1925 *Statistical methods for research workers*. Oliver and Boyd, Edinburgh, London,.

GILKS WP, ABBOTT JK, MORROW EH. 2014 Sex differences in disease genetics: evidence, evolution, and detection. Trends Genet. 30(10):453-463.

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(8):955-959.

HUYNH KD, FISCHLE W, VERDIN E, BARDWELL VJ. 2000 BCoR, a novel corepressor involved in BCL-6 repression. Genes Dev. 14(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, FREEMAN C, HUNT SE, EDKINS S, GRAY E, BOOTH DR, POTTER SC, GORIS A, BAND G, OTURAI AB, STRANGE A, SAARELA J, BELLENGUEZ C, FONTAINE B, GILLMAN M, HEMMER B, GWILLIAM R, ZIPP F, JAYAKUMAR A, MARTIN R, LESLIE S, HAWKINS S, GIANNOULATOU E, D'ALFONSO S, BLACKBURN H, MARTINELLI BONESCHI F, LIDDLE J, HARBO HF, PEREZ ML, SPURKLAND A, WALLER MJ, MYCKO MP, RICKETTS M, COMABELLA M, HAMMOND N, KOCKUM I, MCCANN OT, BAN M, WHITTAKER P, KEMPPINEN A, WESTON P, HAWKINS C, WIDAA S, ZAJICEK J, DRONOV S, ROBERTSON N, BUMPSTEAD SJ, BARCELLOS LF, RAVINDRARAJAH R, ABRAHAM R, ALFREDSSON L, ARDLIE K, AUBIN C, BAKER A, BAKER K, BARANZINI SE, BERGAMASCHI L, BERGAMASCHI R, BERNSTEIN A, BERTHELE A, BOGGILD M, BRADFIELD JP, BRASSAT D, BROADLEY SA, BUCK D, BUTZKUEVEN H, CAPRA R, CARROLL WM, CAVALLA P, CELIUS EG, CEPOK S, CHIAVACCI R, CLERGET-DARPOUX F, CLYSTERS K, COMI G, COSSBURN M, COURNU-REBEIX I, COX MB, COZEN W, CREE BA, CROSS AH, CUSI D, DALY MJ, DAVIS E, DE BAKKER PI, DEBOUVERIE M, D'HOOGHE M B, DIXON K, DOBOSI R, DUBOIS B, ELLINGHAUS D, ELOVAARA I, ESPOSITO F, FONTENILLE C, FOOTE S, FRANKE A, GALIMBERTI D, GHEZZI A, GLESSNER J, GOMEZ R, GOUT O, GRAHAM C, GRANT SF, GUERINI FR, HAKONARSON H, HALL P, HAMSTEN A, HARTUNG HP, HEARD RN, HEATH S, HOBART J, HOSHI M, INFANTE-DUARTE C, INGRAM G, INGRAM W, ISLAM T, JAGODIC M, KABESCH M, KERMODE AG, KILPATRICK TJ, KIM C, KLOPP N, KOIVISTO K, LARSSON M, LATHROP M, LECHNER-SCOTT JS, LEONE MA, LEPPA V, LILJEDAHL U, BOMFIM IL, LINCOLN RR, LINK J, LIU J, LORENTZEN AR, LUPOLI S, MACCIARDI F, MACK T, MARRIOTT M, MARTINELLI V, MASON D, MCCAULEY JL, MENTCH F, MERO IL, MIHALOVA T, MONTALBAN X, MOTTERSHEAD J, MYHR KM, NALDI P, OLLIER W, PAGE A, PALOTIE A, PELLETIER J, PICCIO L, PICKERSGILL T, PIEHL F, POBYWAJLO S, QUACH HL, RAMSAY PP, REUNANEN M, REYNOLDS R, RIOUX JD, RODEGHER M, ROESNER S, RUBIO JP, RUCKERT IM, SALVETTI M, SALVI E, SANTANIELLO A, SCHAEFER CA, SCHREIBER S, SCHULZE C, SCOTT RJ, SELLEBJERG F, SELMAJ KW, SEXTON D, SHEN L, SIMMS-ACUNA B, SKIDMORE S, SLEIMAN PM, SMESTAD C, SORENSEN PS, SONDERGAARD HB, STANKOVICH J, STRANGE RC, SULONEN AM, SUNDQVIST E, SYVANEN AC, TADDEO F, TAYLOR B, BLACKWELL JM, TIENARI P, BRAMON E, TOURBAH A, BROWN MA, TRONCZYNSKA E, CASAS JP, TUBRIDY N, CORVIN A, VICKERY J, JANKOWSKI J, VILLOSLADA P, MARKUS HS, WANG K, MATHEW CG, WASON J, PALMER CN, WICHMANN HE, PLOMIN R, WILLOUGHBY E, RAUTANEN A, WINKELMANN J, WITTIG M, TREMBATH RC, YAOUANQ J, VISWANATHAN AC, ZHANG H, WOOD NW, ZUVICH R, DELOUKAS P, LANGFORD C, DUNCANSON A, OKSENBERG JR, PERICAK-VANCE MA, HAINES JL, OLSSON T, HILLERT J, IVINSON AJ, DE JAGER PL, PELTONEN L, STEWART GJ, HAFLER DA, HAUSER SL, MCVEAN G, DONNELLY P, COMPSTON A. 2011 Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis. Nature. 476(7359):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(1):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(11):1479-1484.

JIN Y, BIRLEA SA, FAIN PR, FERRARA TM, BEN S, RICCARDI SL, COLE JB, GOWAN K, HOLLAND PJ, BENNETT DC, LUITEN RM, WOLKERSTORFER A, VAN DER VEEN JP, HARTMANN A, EICHNER S, SCHULER G, VAN GEEL N, LAMBERT

bioRxiv preprint doi: https://doi.org/10.1101/009795; this version posted April 1, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

J, KEMP EH, GAWKRODGER DJ, WEETMAN AP, TAIEB A, JOUARY T, EZZEDINE K, WALLACE MR, MCCORMACK WT, PICARDO M, LEONE G, OVERBECK A, SILVERBERG NB, SPRITZ RA. 2012 Genome-wide association analyses identify 13 new susceptibility loci for generalized vitiligo. Nat Genet. 44(6):676-680.
JIN Y, BIRLEA SA, FAIN PR, GOWAN K, RICCARDI SL, HOLLAND PJ, MAILLOUX CM, SUFIT AJ, HUTTON SM, AMADI-MYERS A, BENNETT DC, WALLACE MR, MCCORMACK WT, KEMP EH, GAWKRODGER DJ, WEETMAN AP, PICARDO M, LEONE G, TAIEB A, JOUARY T, EZZEDINE K, VAN GEEL N, LAMBERT J, OVERBECK A, SPRITZ RA. 2010 Variant of TYR and autoimmunity susceptibility loci in generalized vitiligo. N Engl J Med. 362(18):1686-1697.
JU T, CUMMINGS RD. 2005 Protein glycosylation: chaperone mutation in Tn syndrome. Nature. 437(7063):1252.
KORN JM, KURUVILLA FG, MCCARROLL SA, WYSOKER A, NEMESH J, CAWLEY S, HUBBELL E, VEITCH J, COLLINS PJ, DARVISHI K, LEE C, NIZZARI MM, GABRIEL SB, PURCELL S, DALY MJ, ALTSHULER D. 2008 Integrated genotype calling and association analysis of SNPs, common copy number polymorphisms and rare CNVs. Nat Genet. 40(10):1253-1260.
LAAKSOVIRTA H, PEURALINNA T, SCHYMICK JC, SCHOLZ SW, LAI SL, MYLLYKANGAS L, SULKAVA R, JANSSON L, HERNANDEZ DG, GIBBS JR, NALLS MA, HECKERMAN D, TIENARI PJ, TRAYNOR BJ. 2010 Chromosome 9p21 in amyotrophic lateral sclerosis in Finland: a genome-wide association study. Lancet Neurol. 9(10):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(1):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(2):383-390.
LI YR, BRADFIELD JP, MENTCH F, MAGGADOTTIR SM, HOU C, ZHAO SD, ABRAMS DJ, CHANG D, GAO F, GUO Y, ET AL. The Genetic Architecture of Pediatric Autoimmune Disease. Submitted.
LIU JZ, MCRAE AF, NYHOLT DR, MEDLAND SE, WRAY NR, BROWN KM, INVESTIGATORS A, HAYWARD NK, MONTGOMERY GW, VISSCHER PM, MARTIN NG, MACGREGOR S. 2010 A versatile gene-based test for genome-wide association studies. Am J Hum Genet. 87(1):139-145.
LOLEY C, ZIEGLER A, KONIG IR. 2011 Association tests for X-chromosomal markers--a comparison of different test statistics. Hum Hered. 71(1):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(6):1077-1084.
MA L, CLARK AG, KEINAN A. 2013 Gene-based testing of interactions in association studies of quantitative traits. PLoS Genet. 9(2):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, POPOVA N, PRETEL S, ZIYABARI L, LEE M, SHAO Y, WANG ZY, SIROTKIN K, WARD M, KHOLODOV M, ZBICZ K, BECK J, KIMELMAN M, SHEVELEV S, PREUSS D, YASCHENKO E, GRAEFF A, OSTELL J, SHERRY ST. 2007 The NCBI dbGaP database of genotypes and phenotypes. Nat Genet. 39(10):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(9):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(1):88-98.
MUSCAT JE, RICHIE JP, JR., THOMPSON S, WYNDER EL. 1996 Gender differences in smoking and risk for oral cancer. Cancer Res. 56(22):5192-5197.
NAIR RP, DUFFIN KC, HELMS C, DING J, STUART PE, GOLDGAR D, GUDJONSSON JE, LI Y, TEJASVI T, FENG BJ, RUETHER A, SCHREIBER S, WEICHENTHAL M, GLADMAN D, RAHMAN P, SCHRODI SJ, PRAHALAD S, GUTHERY SL, FISCHER J, LIAO W, KWOK PY, MENTER A, LATHROP GM, WISE CA, BEGOVICH AB, VOORHEES JJ, ELDER JT, KRUEGER GG, BOWCOCK AM, ABECASIS GR, COLLABORATIVE ASSOCIATION STUDY OF P. 2009 Genome-wide scan reveals association of psoriasis with IL-23 and NF-kappaB pathways. Nat Genet. 41(2):199-204.
OBER C, LOISEL DA, GILAD Y. 2008 Sex-specific genetic architecture of human disease. Nat Rev Genet. 9(12):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(8):904-909.

bioRxiv preprint doi: https://doi.org/10.1101/009795; this version posted April 1, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

PURCELL S, NEALE B, TODD-BROWN K, THOMAS L, FERREIRA MA, BENDER D, MALLER J, SKLAR P, DE BAKKER PI, DALY MJ, SHAM PC. 2007 PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 81(3):559-575.
QI L, CORNELIS MC, KRAFT P, STANYA KJ, LINDA KAO WH, PANKOW JS, DUPUIS J, FLOREZ JC, FOX CS, PARE G, SUN Q, GIRMAN CJ, LAURIE CC, MIREL DB, MANOLIO TA, CHASMAN DI, BOERWINKLE E, RIDKER PM, HUNTER DJ, MEIGS JB, LEE CH, HU FB, VAN DAM RM, META-ANALYSIS OF G, INSULIN-RELATED TRAITS C, DIABETES GENETICS R, META-ANALYSIS C. 2010 Genetic variants at 2q24 are associated with susceptibility to type 2 diabetes. Hum Mol Genet. 19(13):2706-2715.
SLAVNEY A, ARBIZA L, CLARK AG, KEINAN A. Human X-inactivation escapers show signatures of dosage sensitivity. Submitted.
STOUFFER SA. 1949 *The American soldier*. Princeton University Press, Princeton,.
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(7145):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(5):438-450.
TRYKA KA, HAO L, STURCKE A, JIN Y, WANG ZY, ZIYABARI L, LEE M, POPOVA N, SHAROPOVA N, KIMURA M, FEOLO M. 2014 NCBI's Database of Genotypes and Phenotypes: dbGaP. Nucleic Acids Res. 42(Database issue):D975-979.
TUKIAINEN T, PIRINEN M, SARIN AP, LADENVALL C, KETTUNEN J, LEHTIMAKI T, LOKKI ML, PEROLA M, SINISALO J, VLACHOPOULOU E, ERIKSSON JG, GROOP L, JULA A, JARVELIN MR, RAITAKARI OT, SALOMAA V, RIPATTI S. 2014 Chromosome X-wide association study identifies Loci for fasting insulin and height and evidence for incomplete dosage compensation. PLoS Genet. 10(2):e1004127.
UK IBD GENETICS CONSORTIUM, BARRETT JC, LEE JC, LEES CW, PRESCOTT NJ, ANDERSON CA, PHILLIPS A, WESLEY E, PARNELL K, ZHANG H, DRUMMOND H, NIMMO ER, MASSEY D, BLASZCZYK K, ELLIOTT T, COTTERILL L, DALLAL H, LOBO AJ, MOWAT C, SANDERSON JD, JEWELL DP, NEWMAN WG, EDWARDS C, AHMAD T, MANSFIELD JC, SATSANGI J, PARKES M, MATHEW CG, WELLCOME TRUST CASE CONTROL C, DONNELLY P, PELTONEN L, BLACKWELL JM, BRAMON E, BROWN MA, CASAS JP, CORVIN A, CRADDOCK N, DELOUKAS P, DUNCANSON A, JANKOWSKI J, MARKUS HS, MATHEW CG, MCCARTHY MI, PALMER CN, PLOMIN R, RAUTANEN A, SAWCER SJ, SAMANI N, TREMBATH RC, VISWANATHAN AC, WOOD N, SPENCER CC, BARRETT JC, BELLENGUEZ C, DAVISON D, FREEMAN C, STRANGE A, DONNELLY P, LANGFORD C, HUNT SE, EDKINS S, GWILLIAM R, BLACKBURN H, BUMPSTEAD SJ, DRONOV S, GILLMAN M, GRAY E, HAMMOND N, JAYAKUMAR A, MCCANN OT, LIDDLE J, PEREZ ML, POTTER SC, RAVINDRARAJAH R, RICKETTS M, WALLER M, WESTON P, WIDAA S, WHITTAKER P, DELOUKAS P, PELTONEN L, MATHEW CG, BLACKWELL JM, BROWN MA, CORVIN A, MCCARTHY MI, SPENCER CC, ATTWOOD AP, STEPHENS J, SAMBROOK J, OUWEHAND WH, MCARDLE WL, RING SM, STRACHAN DP. 2009 Genome-wide association study of ulcerative colitis identifies three new susceptibility loci, including the HNF4A region. Nat Genet. 41(12):1330-1334.
VOSKUHL R. 2011 Sex differences in autoimmune diseases. Biol Sex Differ. 2(1):1.
WELTER D, MACARTHUR J, MORALES J, BURDETT T, HALL P, JUNKINS H, KLEMM A, FLICEK P, MANOLIO T, HINDORFF L, PARKINSON H. 2014 The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. Nucleic Acids Res. 42(Database issue):D1001-1006.
WILLER CJ, LI Y, ABECASIS GR. 2010 METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics. 26(17):2190-2191.
WILLIAMS OD. 1989 The Atherosclerosis Risk in Communities (Aric) Study - Design and Objectives. Am J Epidemiol. 129(4):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(5):643-647.
ZAYKIN DV, ZHIVOTOVSKY LA, WESTFALL PH, WEIR BS. 2002 Truncated product method for combining P-values. Genet Epidemiol. 22(2):170-185.
ZHENG G, JOO J, ZHANG C, GELLER NL. 2007 Testing association for markers on the X chromosome. Genet Epidemiol. 31(8):834-843.