

## EDUCATION

## Eight quick tips for including chromosome X in genome-wide association studies

Justin Bellavance<sup>1,2</sup>, Linda Wang<sup>1,2</sup>, Sarah A. Gagliano Taliun<sup>2,3,4\*</sup>

**1** Faculty of Medicine, Université de Montréal, Montréal, Québec, Canada, **2** Research Centre, Montréal Heart Institute, Montréal, Québec, Canada, **3** Department of Medicine, Faculty of Medicine, Université de Montréal, Montréal, Québec, Canada, **4** Department of Neurosciences, Faculty of Medicine, Université de Montréal, Montréal, Québec, Canada

☞ These authors contributed equally to this work.

\* [sarah.gagliano-taliun@umontreal.ca](mailto:sarah.gagliano-taliun@umontreal.ca)



## OPEN ACCESS

**Citation:** Bellavance J, Wang L, Gagliano Taliun SA (2024) Eight quick tips for including chromosome X in genome-wide association studies. *PLoS Comput Biol* 20(6): e1012160. <https://doi.org/10.1371/journal.pcbi.1012160>

**Editor:** Patricia M. Palagi, SIB Swiss Institute of Bioinformatics, SWITZERLAND

**Published:** June 6, 2024

**Copyright:** © 2024 Bellavance et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** SAGT was funded by a Junior 1 award from the Fonds de Recherche du Québec - Santé (FRQS; <https://frq.gouv.qc.ca>) and acknowledges funding through the Alzheimer Society Research Program by the Alzheimer Society of Canada and co-funding by the Canadian Institutes of Health Research – Institute of Aging (Funding Reference Number 189919), as well as support from two Canadian Institutes of Health Research (CIHR) Project Grants (PJT 183817 and PJT 191687) and a CIHR Catalyst Grant (AD6 192920). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

## Introduction

All individuals carry a minimum of 1 copy of chromosome X. Despite being a relatively long chromosome with more than 150 million base pairs [1], similar in length to chromosome 8, association testing of genetic variants on chromosome X is still not routinely conducted. Genome-wide association studies (GWAS) have been used to identify a vast range of genomic loci of interest for a variety of complex human diseases and traits by quantifying genetic variants that are statistically associated with a given disease/trait [2,3]. However, a lack of testing for variants on the X chromosome limits our ability to identify vital loci and subsequently understand potential mechanisms linked to this chromosome. There was a call for the inclusion of chromosome X into genome-wide association analyses presented in 2013. At that time, a scan of published GWAS from 2010 and 2011 showed that only 33% of the studies had tested variants on the X chromosome in their analyses [4]. Despite this call for inclusion, the lack of representation of this chromosome has not improved according to a 2023 study. Of the 136 publications that submitted at least 1 summary statistics file to the NHGRI-EBI GWAS Catalog in 2021, only 25% reported chromosome X results [5].

Indeed, there are several characteristics of this chromosome that make it unique compared to the autosomes, which can pose analytical challenges in association testing. Such challenges include how to account for X inactivation in individuals with an XX karyotype, how to model the hemizygous state of genotypes in individuals with an XY karyotype, or how to best code genotypes at the 2 pseudo-autosomal regions, short stretches at either end of the X with high homology with the Y chromosome, known as PAR1 and PAR2. The non-pseudo-autosomal region (nonPAR) denotes the middle sequence of the X chromosome.

Furthermore, there are many well-used software that take GWAS summary statistics as input and ignore chromosome X information [6,7]. This practice can make it difficult and unintuitive for researchers to run association testing on the X chromosome.

Inclusion of chromosome X routinely in GWAS and downstream analyses will serve to enhance our understanding of the genetic contributors to complex diseases and traits. Here, we propose 8 tips to help move towards the inclusion of X in GWAS to provide a suggested set of concrete actions that can be taken to overcome the challenges or obstacles preventing routine analysis of this chromosome.

**Competing interests:** The authors have declared that no competing interests exist.

## Tip 1: Retain chromosome X variants in your quality control pipeline

Chromosome X variants tend to have lower quality whether genotyped, sequenced, or imputed compared to autosomal variants [8,9]. Perform careful quality control checks genome-wide, including variants on the X chromosome instead of discarding genotyping or sequencing calls for variants on the X chromosome. For sequencing data, depth on the X can be used to get a sense of each study participant's sex chromosome karyotype. For genotyping array data, F statistics of X chromosome heterozygosity, as implemented in PLINK, for example, can be estimated to accomplish this task. Chromosome X-specific variant filters as reviewed by Keur and colleagues [10] should be incorporated, including testing for missingness by sex, call rate by sex, and the proportion of heterozygotes in individuals with XY karyotypes should be assessed. Additionally, Hardy–Weinberg Equilibrium models accounting for different ploidy on chromosome X (for example, diploid for females and haploid for males in the nonPAR region) with different assumptions with regard to X inactivation (see [Tip 6](#)) have been proposed [11].

## Tip 2: Impute chromosome X variants

For genotyping array data, maximize the number of genetic variants available for association testing by imputing genotypes of variants that are not present on the array, including variants on the X chromosome. Various free online resources can be used to carry out the computationally intensive task of genotype imputation using available whole-genome sequencing imputation panels ([Table 1](#)), including the Trans-Omics for Precision Medicine (TOPMed) Imputation Server [12], Sanger Imputation Service [13], and Michigan Imputation Server [14]. The TOPMed and Michigan Imputation Servers will perform some basic chromosome X-specific quality control on the uploaded genotypes, including verifying that all variants in the nonPAR region are either haploid or diploid. For these 2 servers, chromosome X is split into 3 parts (PAR1, nonPAR, and PAR2) for statistical phasing and genotype imputation, and then subsequently merged. In contrast, the Sanger Imputation Service does not perform quality control steps on the genotypes. A thorough assessment of the various imputation panels and tools available to handle chromosome X imputation is pertinent but is beyond the scope of this short educational article.

## Tip 3: Use genetic association software that supports X chromosome testing

It is necessary to understand how conventional genetic association study software handles the X chromosome because different tools handle association testing on the X chromosome differently ([Table 2](#)). These differences can cause some missteps for researchers trying to include the X chromosome in their analyses for the first time. While there is no one recommended

**Table 1. Descriptive summary of some available genotype imputation panels.**

Panel	N samples in panel	N variants	Types of variation available in the panel
TOPMed (Version r2)	97,256	308,107,085 genetic variants (of which 286,068,980 are single nucleotide variants) across the 22 autosomes and the X chromosome	single nucleotide variants and short indels
Haplotype Reference Consortium (Version r1.1 2016)	32,470	39,635,008	single nucleotide variants
1000 Genomes (Phase 3, version 5)	2,504	49,143,605	single nucleotide variants and short indels

<https://doi.org/10.1371/journal.pcbi.1012160.t001>

**Table 2. Non-comprehensive list of software for association testing, including chromosome X variants.** Flags for single variant association testing are described.

Software	Base Considerations	Specific Flags	Description of Flags
PLINK v1.90 [17]	Treats X chromosome the same as autosomes unless a model is specified. Non-PAR variants assigned to chr23 and PAR variants to chr25.	(1)--xchr-model (2)--merge-x	(1) Choose between 4 options: (i) exclude X chromosome, (ii) treat X chromosome the same as autosomes (and add sex as covariate), (iii) code male genotypes as 0/2 instead of 0/1, or (iv) add dosage–sex interaction to the model alongside sex as a covariate. (2) Merge PAR variants to be treated the same as the non-PAR variants.
PLINK v2.00 [18]	Treats X chromosome the same as autosomes unless a model is specified. Non-PAR variants assigned to chr23 and PAR variants to chr25.	(1)--xchr-model (2)--merge-par	(1) Choose between 3 options: (i) keep constant dosage, (ii) have different dosages of X chromosome depending on sex, or (iii) exclude the X chromosome. (2) Merge PAR variants to be treated the same as the non-PAR variants.
SAIGE v1.0.0 [19]	Treats X chromosome the same as autosomes.	(1)--X_ParRegion	Dosages of non-PAR regions of male chromosome X will be multiplied by 2.
REGENIE v3.3 [20]	Treats X chromosome the same as autosomes. Sex chromosomes will be collapsed into a single chromosome by default.	(1)--par-regions	Determine bounds for PAR1/PAR2 regions by using reference genome build code. Default is GRCh38.
BOLT-LMM v2.4.1 [21]	For the X chromosome, males should be coded as diploid, such that male genotypes are coded as 0/2 and female genotypes are coded as 0/1/2 (corresponding to a model of random X inactivation); PAR and non-PAR variants should be merged beforehand.	N/A	N/A
GCTA v1.94.2 [22]	NonPAR variants coded as 0/2 for males and 0/1/2 for females. Like PLINK, non-PAR variants assigned to chr23 and PAR variants to chr25.	--dc 1 or 0	Specify dosage compensation model for X chromosome variants where 1 = full (default) and 0 = no dosage compensation.
XWAS v3.0 [23]	XWAS adds X-specific functionalities on top of PLINK functions. For analysis of variants on the X-chromosome, XWAS removes all variants other than the ones on chromosome X, including variants of chromosome X's PAR. Not to mention, X chromosome variants are filtered if they have significant differences in minor allele frequencies and/or missingness between male and female controls, with thresholds set in QC parameters.	(1)--xwas (2)--clayton-x (3)--strat-sex (4)--xchr-model (5)--run-all	(1) Proceed with XWAS unique commands, such as the ones below. (2) Test for association on the X chromosome using the method published by Clayton, 2008 [24]. (3) This flag will carry out sex-stratified analysis, and then combine the 2 results with options to modify male genotype output as either 0/1 or 0/2. (4) See (1) from PLINK v1.9. (5) Run Clayton's test (see flag 2), sex-stratified test (see flag 3), and tests for difference in effect sizes or phenotypic variance in females vs. males

<https://doi.org/10.1371/journal.pcbi.1012160.t002>

tool or approach for X chromosome association analysis, one study observed that in samples with a skewed male:female ratio, coding the male chromosome X genotypes as 0/2 (rather than 0/1) alleviated type 1 error [15]. Therefore, when male:female imbalance is present in an analysis dataset, coding of male genotypes on the X should be considered. We recommend implementing accepted genome-wide significant thresholds that account for the multiple testing inherent to genome-wide scans (for example,  $5 \times 10^{-8}$ , which adjusts for 1 million independent tests) [16], rather than deriving X-specific significance thresholds.

#### Tip 4: Perform stratified association analyses for X chromosome variants

Stratified analyses carrying out association tests in individuals with XX karyotypes separately from those with XY karyotypes can be conducted to identify associations on the X (as well as for autosomal variants) that differ in magnitude, direction, or significance depending on sex chromosome karyotype. Tests of heterogeneity to identify genetic variants that exhibit significant association differences in the stratified association tests should also be carried out to quantify possible association differences. As sample sizes increase, individuals with sex-chromosome

**Table 3. Summary of selected types of genetic variation with known associations for complex human diseases or traits on chromosome X.**

Type of variation	Brief description	Chr X example in the literature
Single-nucleotide polymorphism/variant (SNP/ SNV)	A difference in a single nucleotide at a specific position in the DNA sequence	<i>GATA1</i> intronic SNP (X:48787974:G:A; GRCh38) associated with eosinophil count [25]
Insertion or deletion (indel)	Insertions or deletions of genetic sequence of up to 50 base pairs	<i>LINC01496</i> intronic deletion (X:51502424:GA/G; GRCh38) associated with prostate carcinoma [26]
Copy number variants (CNVs)	A type of structural variant with genomic segments of extra (duplications) or loss (deletions) of genetic sequence between 1,000 base pairs and 5 megabases [27]	Duplication in Xp22.33 region associated with urea [28]
Structural variants (SVs)	DNA rearrangements (e.g., inversions, insertions, deletions, translocations) involving at least 50 base pairs [29]	gnomAD SV provides a catalogue of SVs in the human genome, including on chromosome X [30]

<https://doi.org/10.1371/journal.pcbi.1012160.t003>

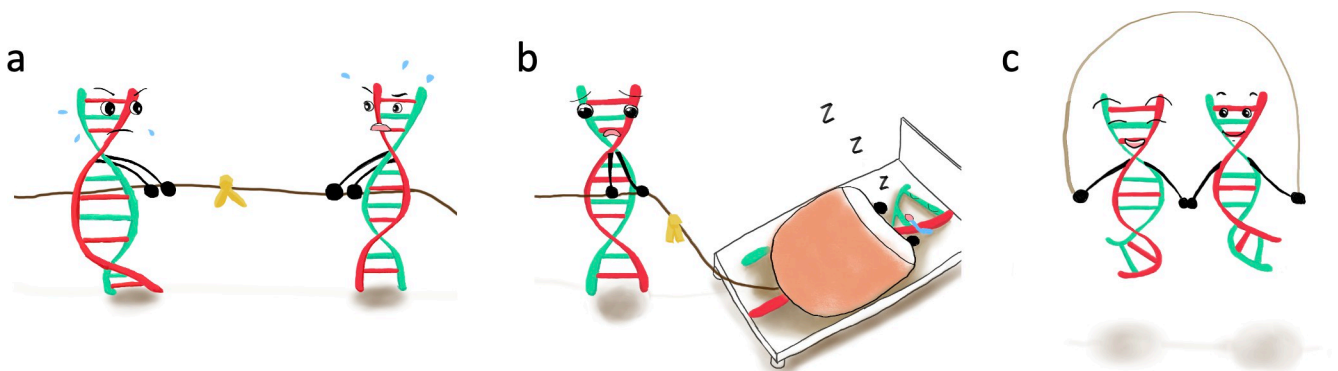
aneuploidy should be included in analyses, possibly stratified according to sex chromosome karyotype. However, stratifying study individuals into subgroups will reduce the total sample size for the association test. In datasets where statistical power is a concern, there are alternative methods to test for effect of sex, such as testing for genetic variant-by-sex interactions.

### Tip 5: At the minimum, assess biallelic single-nucleotide polymorphisms and variants as well as small insertions and deletions (indels) on chromosome X

Limited work has been done on investigating genetic variant–trait associations for more complex types of genetic variation on the X chromosome. Nevertheless, there is evidence of many types of genetic variation on the X chromosome contributing to complex disease risk, thus warranting additional investigations across traits and across types of variation (Table 3). When possible, expand to other variation types such as copy number and structural variants and include variants with more than 2 alleles (i.e., multi-allelic variants).

### Tip 6: Consider implementing various statistical models to account for X inactivation

In XX individuals, there can be compensatory mechanisms to reduce the dosage of gene expression in which one copy is subject to inactivation, resulting in only one copy of the gene being expressed (Fig 1). This biological factor must be considered while performing genetic



**Fig 1. Representation of proposed X inactivation models in XX individuals.** In random X inactivation, 50% of cells have 1 allele inactive, whereas the other 50% have the other allele inactive (a). In skewed (also known as nonrandom) inactivation, more than 50% of cells (say more than 75% or more extreme) have the same allele inactive (b). Finally, escape of X inactivation describes the scenario where both alleles are active in all cells (c).

<https://doi.org/10.1371/journal.pcbi.1012160.g001>

association analyses on chromosome X. Statistical methods have been proposed to account for different inactivation models: random, skewed, or escape from inactivation [23,24,31,32]. PLINK [17,18] and XWAS [23] can model escape from inactivation or random inactivation, and there are models for skewed inactivation [31,33]. Given the lack of widely used standards and guidelines for association testing on the X chromosome, it can be useful to report association results from more than one applied method to further research in this direction using the most recent map of X inactivation across human tissues [34]. Applying and comparing association models accounting for different models of X inactivation for genetic variants on the X chromosome in XX individuals will be an important step forward to foster chromosome X association testing.

### **Tip 7: Make your code and association summary statistics, including chromosome X results, publicly available**

Data availability increases transparency and reproducibility and facilitates future research within the broader research community. Share your results with the broader community in a peer-reviewed publication describing your work. Sharing of summary statistics is commonplace and now required by many journals for publication of primary research articles. These summary statistics are used by the broader research community for a variety of tasks including replication of association signals, meta-analysis of cohorts in which the same or similar trait is measured, carrying out two-sample mendelian randomization, identification of variants to include in the development of a polygenic score and their corresponding weights, and many more downstream analyses. Not to mention, including chromosome X in summary statistics facilitates the inclusion of variants on this chromosome in these downstream analyses, furthering knowledge on the contribution of chromosome X variants to complex traits and diseases.

Code availability goes hand in hand with data availability in the call for open science. Making code available can provide other researchers with a starting point to facilitate the inclusion of X variants in their analyses. To this end, it is important to provide annotated code where the implemented steps are well documented. Guidelines described elsewhere should be followed to ensure code reproducibility that supports open science [35,36].

### **Tip 8: Embrace the biological complexity and value of X chromosome analyses**

By following these tips, we can all do our part in increasing the inclusion of chromosome X in association analyses and embracing our biological complexity through association analyses.

## **Conclusions**

Testing variants on the X chromosome is important for increasing understanding of this chromosome's role in complex traits and diseases. Not to mention, making these results available to the broader community will facilitate downstream analyses including meta-analyses, the creation of polygenic risk scores and causal inference using mendelian randomization that incorporate chromosome X to advance the field. That being said, the inclusion of X chromosome genetic variants in association analyses is not a "one size fits all" approach. There is no single software or X inactivation model that will always give optimal results for every trait and in every study design. We encourage researchers to conduct further investigations to choose the best tool, resource, or method for their specific task using the information provided here as a starting point. All in all, addressing the challenges associated with chromosome X analyses will be crucial to foster future opportunities for scientific discovery related to topics that are

not yet well understood, including, but not limited to, understanding haploid versus diploid states, sex chromosome aneuploidy states, and the biology of nonrandom X inactivation.

## References

- Ross MT, Grafham DV, Coffey AJ, Scherer S, McLay K, Muzny D, et al. The DNA sequence of the human X chromosome. *Nature*. 2005; 434(7031):325–337. <https://doi.org/10.1038/nature03440> PMID: 15772651; PubMed Central PMCID: PMC2665286.
- Watanabe K, Stringer S, Frei O, Umicovic Mirkov M, de Leeuw C, Polderman TJC, et al. A global overview of pleiotropy and genetic architecture in complex traits. *Nat Genet*. 2019; 51(9):1339–48. Epub 20190819. <https://doi.org/10.1038/s41588-019-0481-0> PMID: 31427789.
- MacArthur J, Bowler E, Cerezo M, Gil L, Hall P, Hastings E, et al. The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). *Nucleic Acids Res*. 2017; 45(D1):D896–D901. Epub 20161129. <https://doi.org/10.1093/nar/gkw1133> PMID: 27899670; PubMed Central PMCID: PMC5210590.
- Wise AL, Gyi L, Manolio TA. eXclusion: toward integrating the X chromosome in genome-wide association analyses. *Am J Hum Genet*. 2013; 92(5):643–647. <https://doi.org/10.1016/j.ajhg.2013.03.017> PMID: 23643377; PubMed Central PMCID: PMC3644627.
- Sun L, Wang Z, Lu T, Manolio TA, Paterson AD. eXclusionary: 10 years later, where are the sex chromosomes in GWASs? *Am J Hum Genet*. 2023; 110(6):903–912. <https://doi.org/10.1016/j.ajhg.2023.04.009> PMID: 37267899; PubMed Central PMCID: PMC10257007.
- Bulik-Sullivan BK, Loh PR, Finucane HK, Ripke S, Yang J, Schizophrenia Working Group of the Psychiatric Genomics C, et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet*. 2015; 47(3):291–5. Epub 20150202. <https://doi.org/10.1038/ng.3211> PMID: 25642630; PubMed Central PMCID: PMC4495769.
- Brown BC, Asian Genetic Epidemiology Network Type 2 Diabetes C, Ye CJ, Price AL, Zaitlen N. Trans-ethnic Genetic-Correlation Estimates from Summary Statistics. *Am J Hum Genet*. 2016; 99(1):76–88. Epub 20160616. <https://doi.org/10.1016/j.ajhg.2016.05.001> PMID: 27321947; PubMed Central PMCID: PMC5005434.
- Chen DZ, Roshandel D, Wang Z, Sun L, Paterson AD. Comprehensive whole-genome analyses of the UK Biobank reveal significant sex differences in both genotype missingness and allele frequency on the X chromosome. *Hum Mol Genet*. 2024; 33(6):543–551. <https://doi.org/10.1093/hmg/ddad201> PMID: 38073250.
- Schurz H, Muller SJ, van Helden PD, Tromp G, Hoal EG, Kinnear CJ, et al. Evaluating the Accuracy of Imputation Methods in a Five-Way Admixed Population. *Front Genet*. 2019; 10:34. Epub 20190205. <https://doi.org/10.3389/fgene.2019.00034> PMID: 30804980; PubMed Central PMCID: PMC6370942.
- Keur N, Ricano-Ponce I, Kumar V, Matzaraki V. A systematic review of analytical methods used in genetic association analysis of the X-chromosome. *Brief Bioinform*. 2022; 23(5). <https://doi.org/10.1093/bib/bbac287> PMID: 35901513; PubMed Central PMCID: PMC9764208.
- Graffelman J, Weir BS. Testing for Hardy-Weinberg equilibrium at biallelic genetic markers on the X chromosome. *Heredity (Edinb)*. 2016; 116(6):558–68. Epub 20160413. <https://doi.org/10.1038/hdy.2016.20> PMID: 27071844; PubMed Central PMCID: PMC4868269.
- Taliun D, Harris DN, Kessler MD, Carlson J, Szpiech ZA, Torres R, et al. Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. *Nature*. 2021; 590(7845):290–9. Epub 20210210. <https://doi.org/10.1038/s41586-021-03205-y> PMID: 33568819; PubMed Central PMCID: PMC7875770.
- McCarthy S, Das S, Kretschmar W, Delaneau O, Wood AR, Teumer A, et al. A reference panel of 64,976 haplotypes for genotype imputation. *Nat Genet*. 2016; 48(10):1279–83. Epub 20160822. <https://doi.org/10.1038/ng.3643> PMID: 27548312; PubMed Central PMCID: PMC5388176.
- Das S, Forer L, Schonherr S, Sidore C, Locke AE, Kwong A, et al. Next-generation genotype imputation service and methods. *Nat Genet*. 2016; 48(10):1284–7. Epub 20160829. <https://doi.org/10.1038/ng.3656> PMID: 27571263; PubMed Central PMCID: PMC5157836.
- Ozbek U, Lin HM, Lin Y, Weeks DE, Chen W, Shaffer JR, et al. Statistics for X-chromosome associations. *Genet Epidemiol*. 2018; 42(6):539–50. Epub 20180613. <https://doi.org/10.1002/gepi.22132> PMID: 29900581; PubMed Central PMCID: PMC6394852.
- Uffelmann E, Huang QQ, Munung NS, de Vries J, Okada Y, Martin AR, et al. Genome-wide association studies. *Nat Rev Methods Primers*. 2021; 1(1):59. <https://doi.org/10.1038/s43586-021-00056-9>
- Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 2007; 81(3):559–75. Epub 20070725. <https://doi.org/10.1086/519795> PMID: 17701901; PubMed Central PMCID: PMC1950838.

18. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience*. 2015; 4:7. Epub 20150225. <https://doi.org/10.1186/s13742-015-0047-8> PMID: 25722852; PubMed Central PMCID: PMC4342193.
19. Zhou W, Nielsen JB, Fritsche LG, Dey R, Gabrielsen ME, Wolford BN, et al. Efficiently controlling for case-control imbalance and sample relatedness in large-scale genetic association studies. *Nat Genet*. 2018; 50(9):1335–41. Epub 20180813. <https://doi.org/10.1038/s41588-018-0184-y> PMID: 30104761; PubMed Central PMCID: PMC6119127.
20. Mbatchou J, Barnard L, Backman J, Marcketta A, Kosmicki JA, Ziyatdinov A, et al. Computationally efficient whole-genome regression for quantitative and binary traits. *Nat Genet*. 2021; 53(7):1097–103. Epub 20210520. <https://doi.org/10.1038/s41588-021-00870-7> PMID: 34017140.
21. Loh PR, Kichaev G, Gazal S, Schoech AP, Price AL. Mixed-model association for biobank-scale datasets. *Nat Genet*. 2018; 50(7):906–908. <https://doi.org/10.1038/s41588-018-0144-6> PMID: 29892013; PubMed Central PMCID: PMC6309610.
22. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet*. 2011; 88(1):76–82. Epub 20101217. <https://doi.org/10.1016/j.ajhg.2010.11.011> PMID: 21167468; PubMed Central PMCID: PMC3014363.
23. Gao F, Chang D, Biddanda A, Ma L, Guo Y, Zhou Z, et al. XWAS: A Software Toolset for Genetic Data Analysis and Association Studies of the X Chromosome. *J Hered*. 2015; 106(5):666–71. Epub 20150812. <https://doi.org/10.1093/jhered/esv059> PMID: 26268243; PubMed Central PMCID: PMC4567842.
24. Clayton D. Testing for association on the X chromosome. *Biostatistics*. 2008; 9(4):593–600. Epub 20080425. <https://doi.org/10.1093/biostatistics/kxn007> PMID: 18441336; PubMed Central PMCID: PMC2536723.
25. Vuckovic D, Bao EL, Akbari P, Lareau CA, Mousas A, Jiang T, et al. The Polygenic and Monogenic Basis of Blood Traits and Diseases. *Cell*. 2020; 182(5):1214–31 e11. <https://doi.org/10.1016/j.cell.2020.08.008> PMID: 32888494; PubMed Central PMCID: PMC7482360.
26. Conti DV, Darst BF, Moss LC, Saunders EJ, Sheng X, Chou A, et al. Trans-ancestry genome-wide association meta-analysis of prostate cancer identifies new susceptibility loci and informs genetic risk prediction. *Nat Genet*. 2021; 53(1):65–75. Epub 20210104. <https://doi.org/10.1038/s41588-020-00748-0> PMID: 33398198; PubMed Central PMCID: PMC8148035.
27. Eichler EE. Copy Number Variation and Human Disease. *Nat Educ*. 2008; 1(3):1.
28. Auwerx C, Lepamets M, Sadler MC, Patxot M, Stojanov M, Baud D, et al. The individual and global impact of copy-number variants on complex human traits. *Am J Hum Genet*. 2022; 109(4):647–68. Epub 20220302. <https://doi.org/10.1016/j.ajhg.2022.02.010> PMID: 35240056; PubMed Central PMCID: PMC9069145.
29. Sudmant PH, Rausch T, Gardner EJ, Handsaker RE, Abyzov A, Huddleston J, et al. An integrated map of structural variation in 2,504 human genomes. *Nature*. 2015; 526(7571):75–81. <https://doi.org/10.1038/nature15394> PMID: 26432246; PubMed Central PMCID: PMC4617611.
30. Collins RL, Brand H, Karczewski KJ, Zhao X, Alföldi J, Francioli LC, et al. A structural variation reference for medical and population genetics. *Nature*. 2020; 581(7809):444–51. Epub 20200527. <https://doi.org/10.1038/s41586-020-2287-8> PMID: 32461652; PubMed Central PMCID: PMC7334194.
31. Wang J, Yu R, Shete S. X-chromosome genetic association test accounting for X-inactivation, skewed X-inactivation, and escape from X-inactivation. *Genet Epidemiol*. 2014; 38(6):483–93. Epub 20140708. <https://doi.org/10.1002/gepi.21814> PMID: 25043884; PubMed Central PMCID: PMC4127090.
32. Chen B, Craiu RV, Sun L. Bayesian model averaging for the X-chromosome inactivation dilemma in genetic association study. *Biostatistics*. 2020; 21(2):319–335. <https://doi.org/10.1093/biostatistics/kxy049> PMID: 30247537.
33. Su Y, Hu J, Yin P, Jiang H, Chen S, Dai M, et al. XCMAX4: A Robust X Chromosomal Genetic Association Test Accounting for Covariates. *Genes (Basel)*. 2022; 13(5). Epub 20220509. <https://doi.org/10.3390/genes13050847> PMID: 35627231; PubMed Central PMCID: PMC9141238.
34. Tukiainen T, Villani AC, Yen A, Rivas MA, Marshall JL, Satiya R, et al. Landscape of X chromosome inactivation across human tissues. *Nature*. 2017; 550(7675):244–248. <https://doi.org/10.1038/nature24265> PMID: 29022598; PubMed Central PMCID: PMC5685192.
35. Sandve GK, Nekrutenko A, Taylor J, Hovig E. Ten simple rules for reproducible computational research. *PLoS Comput Biol*. 2013; 9(10):e1003285. Epub 20131024. <https://doi.org/10.1371/journal.pcbi.1003285> PMID: 24204232; PubMed Central PMCID: PMC3812051.
36. Tonzani S, Fiorani S. The STAR Methods way towards reproducibility and open science. *iScience*. 2021; 24(4):102137. Epub 20210401. <https://doi.org/10.1016/j.isci.2021.102137> PMID: 33997663; PubMed Central PMCID: PMC8100894.