Sex differences and the X chromosome

Sarah Medland

Boulder 2019

Thinking about sex differences using the language of heterogeneity

- Are these differences due to differences in the magnitude of the effects (quantitative)?
  - e.g. Is the contribution of genetic effects greater/smaller in males than in females?

- Are the differences due to differences in the source/nature of the effects (qualitative)?
  - e.g. Are there different genetic effect influencing the trait in males and females?

The language of heterogeneity

- Sex differences = Sex limitation

1948

ON SEX LIMITATION IN HUMAN GENETICS*

By H. HARRIS, M.B., B.Chir.(Camb.)

IT is well known that in many instances of hereditary disease the condition is observed to occur more frequently in one

cases, the sons never inherit the peculiarity directly from their fathers, but the daughters, and the daughters alone, transmit the latent tendency, so that the sons of the daughters

1840

L'HÉRÉDITÉ DANS LES MALADIES,

PAR P. A. PIORRY,

Docteur en médecine, Chevalier de la Légion-d'Honneur, Médecin de l'Hôpital de la Pitié, Agrégé à la Faculté de Médecine de Paris, Professeur de Clinique et de Pathologie interne, Membre de l'Académie Royale de Médecine, des Sociétés médicales de Tours, de Boulogne, de Goëttingue, de l'Académie Royale de Médecine de Madrid, etc.

1861

THE

BRITISH AND FOREIGN

MEDICO-CHIRURGICAL

REVIEW

OR

QUARTERLY JOURNAL

OF

PRACTICAL MEDICINE AND SURGERY.

VOL. XXVII.

JANUARY — APRIL, 1861.

ART. III.

On Sexual Limitation in Hereditary Disease. By WILLIAM SEDGWICK.

(*Concluded from our last.*)

FROM hereditary diseases of the organ of vision, the transition is easy to those affecting the organ of hearing, for there are some defects which these organs seem, as it were, to share in common. This connexion has been already referred to by some writers, amongst whom Mr. White Cooper* states that imperfection of the two senses (of sight and hearing) not infrequently co-exist, especially in the curious class of cases we have just been considering, where the inability to distinguish colours is often associated with a corresponding inability to distinguish musical sounds. Dr. Earle relates, in his case of colour-blindness, that "the whole family, of which the chart has been exhibited, is probably no less generally characterized by a defective musical ear than an imperfect appreciation of colours. Several of the individuals comprised in it are utterly incapable of distinguishing one tune from another."†

* Cyclopædia of Anatomy and Physiology, art. "Vision," p. 1453.
† American Journal of Medical Science, vol. xxxv. p. 847. 1845.

The language of heterogeneity

Quantitative

- differences in the magnitude of the effects

Models

- Scalar

Qualitative

- differences in the source/nature of the effects

Models

- Non-scalar

The language of heterogeneity

- Scalar limitation (Quantitative)
  - % of variance due to A,C,E are the same between groups
  - The total variance is not ie:
    - $var_{Female} = k*var_{Male}$
    - $A_{Female} = k*A_{Male}$
    - $E_{Female} = k*E_{Male}$

k here is the *scalar*

The language of heterogeneity

- Non-Scalar limitation
  - $var_{Female} \neq var_{Male}$
  - $A_{Female} \neq A_{Male}$
  - $E_{Female} \neq E_{Male}$

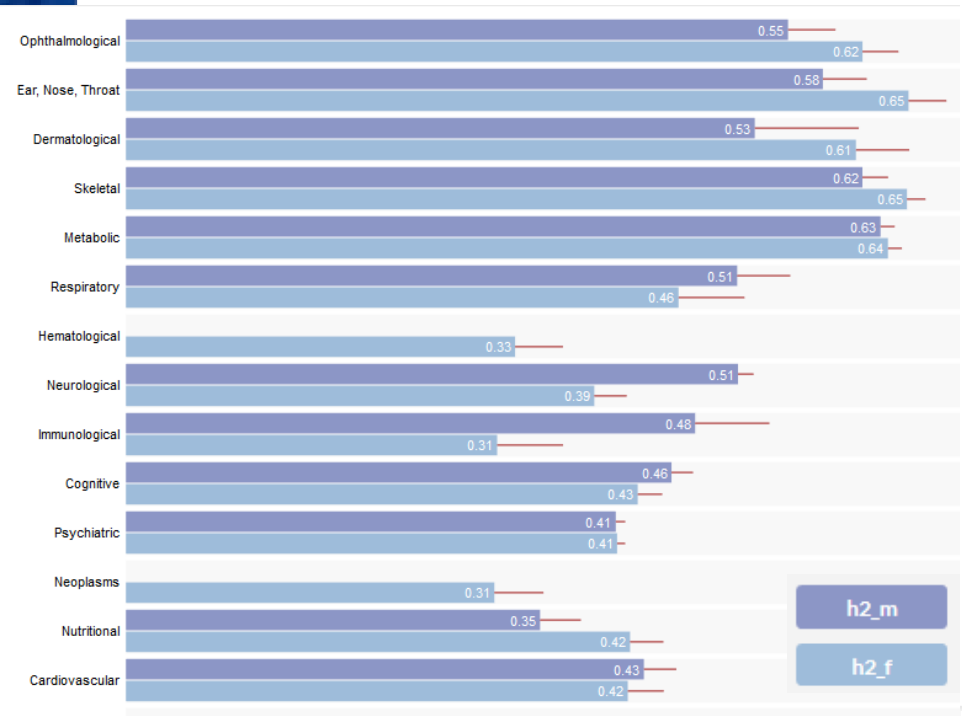
nature genetics

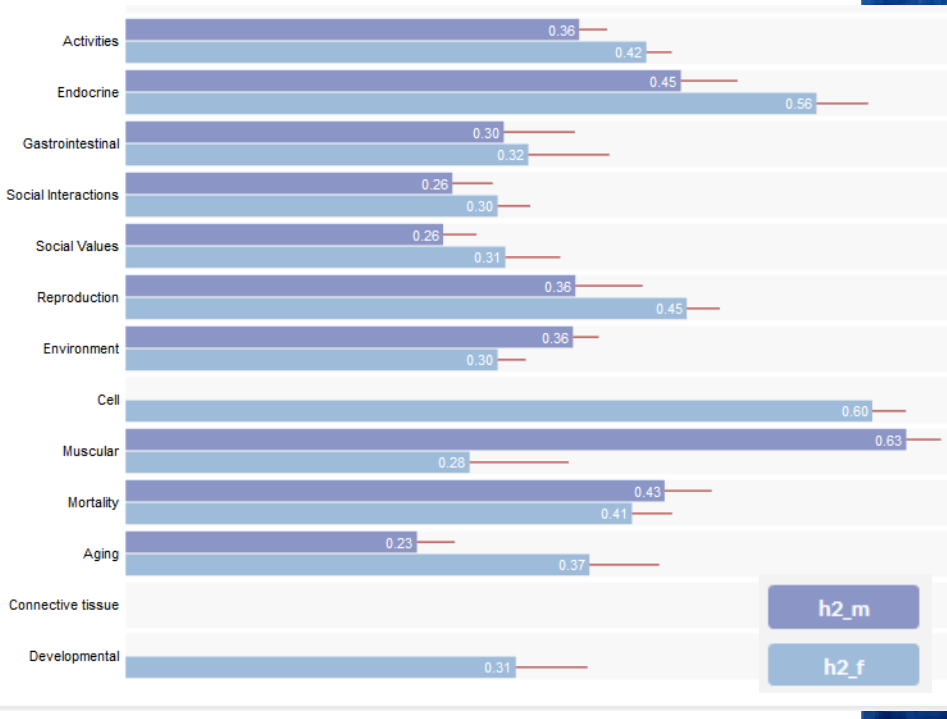
Analysis | Published: 18 May 2015

Meta-analysis of the heritability of human traits based on fifty years of twin studies

Tinca J C Polderman, Beben Benyamin, Christiaan A de Leeuw, Patrick F Sullivan, Arjen van Bochoven, Peter M Visscher & Danielle Posthuma

Nature Genetics **47**, 702–709 (2015) | Download Citation





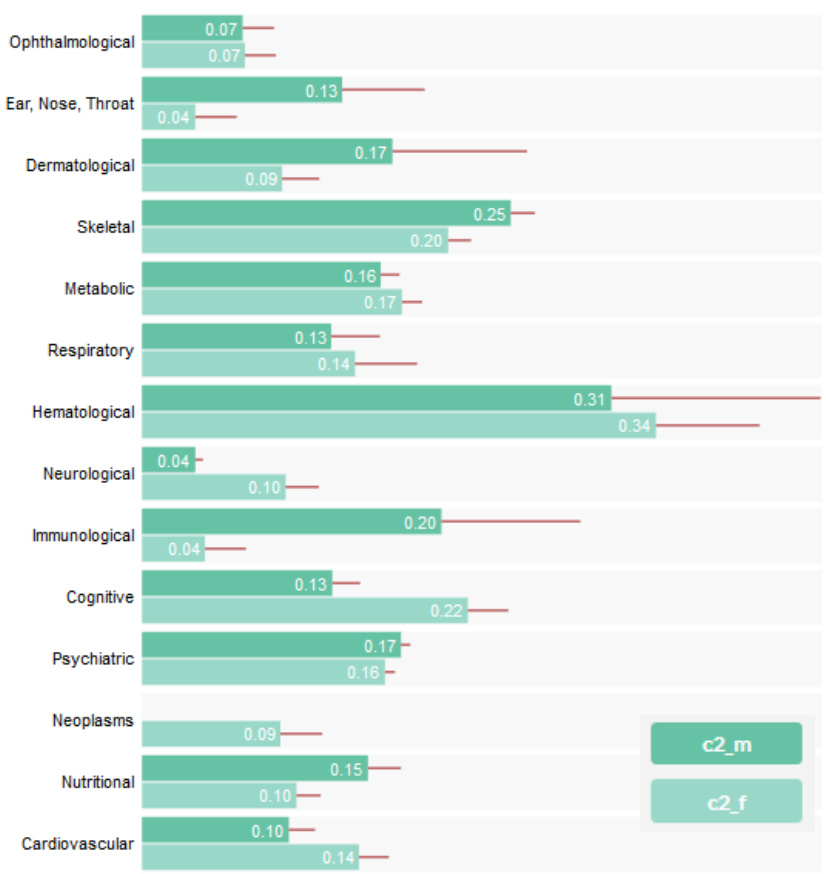
nature genetics

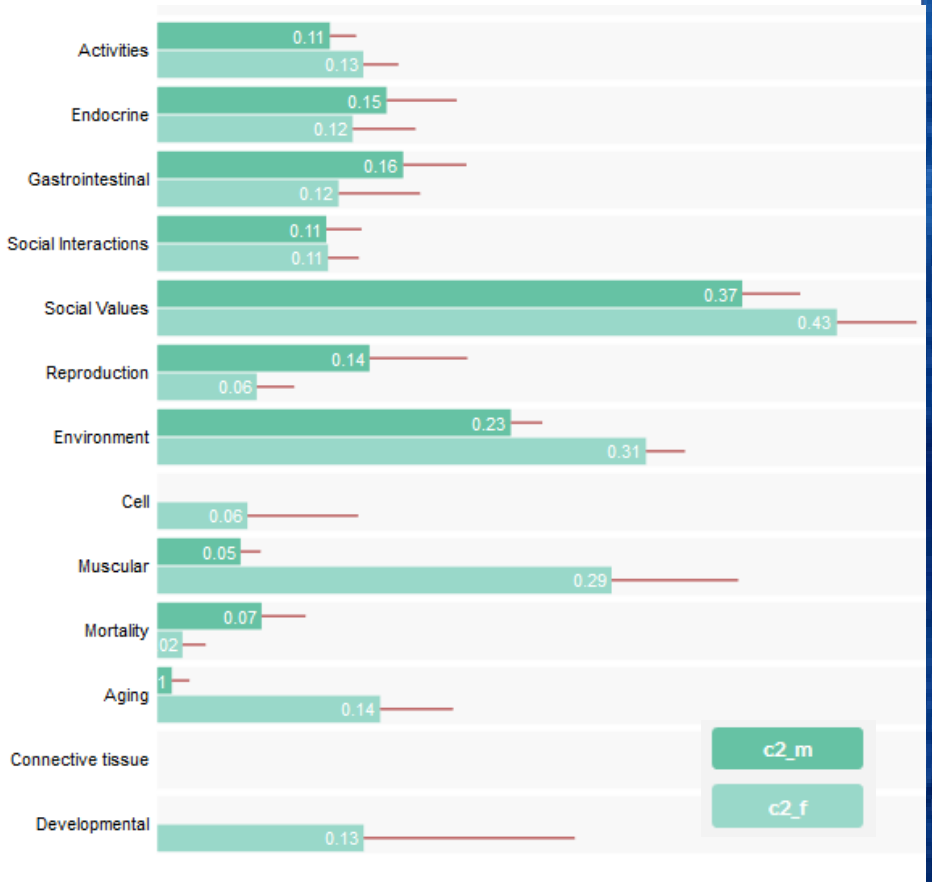
Analysis | Published: 18 May 2015

Meta-analysis of the heritability of human traits based on fifty years of twin studies

Tinca J C Polderman, Beben Benyamin, Christiaan A de Leeuw, Patrick F Sullivan, Arjen van Bochoven, Peter M Visscher & Danielle Posthuma

Nature Genetics **47**, 702–709 (2015) | Download Citation





How can we test for this in a GWAS?

- Check for sex differences the phenotypic distributions or frequencies
- Check for interactions between SNP effects and sex
  - Make sure you include a main effect of sex when you do this
  - Plink 1.9 – include sex in the cov file and request main effects and interactions

```
plink --bfile mydata --linear interaction
--covar my.cov --sex
```

Other ways of doing this?

Genetic Epidemiology 34:846–853 (2010)

Meta-Analysis of Sex-Specific Genome-Wide Association Studies

Reedik Magi, Cecilia M. Lindgren, and Andrew P. Morris*



- MALE-SPECIFIC. Analyze males only in each GWAS. Combine allelic effect estimates in a fixed-effects meta-analysis, weighted by the inverse variance, and test for association with the trait using X^2_{Mj}. $X^2_{Mj} = B^2_{Mj}/V_{Mj}$ 1df
- FEMALE-SPECIFIC. Analyze females only in each GWAS. Combine allelic effect estimates in a fixed-effects meta-analysis, weighted by the inverse variance, and test for association with the trait using X^2_{Fj}. $X^2_{Fj} = B^2_{Fj}/V_{Fj}$, 1df
- SEX-DIFFERENTIATED. Analyze males and females separately in each GWAS. Obtain male- and female-specific allelic effect estimates in a fixed-effects meta-analysis, and test for association with the trait, allowing for sex-differentiation using X^2_{Dj}. $X^2_{Dj} = X^2_{Mj} + X^2_{Fj}$, 2df
- HETEROGENEITY. Analyze males and females separately in each GWAS. Obtain male- and female-specific allelic effect estimates in a fixed-effects meta-analysis, and test for heterogeneity between the sexes using X^2_{Hj}. $X^2_{Hj} = X^2_{Dj} - X^2_{Cj}$, 1df
- SEX-COMBINED. Analyze males and females combined in each GWAS of , ambivalent to sex. Combine allelic effect estimates in a fixed-effects meta-analysis, weighted by the inverse variance, and test for association with the trait. $X^2_{Cj} = B^2_{Cj}/V_{Cj}$ 1df

Implementation in the GWASMA software package

PLOS ONE

PUBLISH

OPEN ACCESS PEER-REVIEWED

RESEARCH ARTICLE

Approaches to detect genetic effects that differ between two strata in genome-wide meta-analyses: Recommendations based on a systematic evaluation

Thomas W. Winkler, Anne E. Justice, L. Adrienne Cupples, Florian Kronenberg, Zoltán Kutalik, Iris M. Heid,
the GIANT consortium

Published: July 27, 2017 • https://doi.org/10.1371/journal.pone.0181038

Several consortia conduct *stratified GWAMAS*, where study analysts are asked to perform the analyses separately by stratum—for example separately for men and women or for persons with and without diabetes [4,5,7,10]. For study analysts, this is relatively straight forward to implement with existing genome-wide analysis software. For meta-analysts, stratified GWAS allow a stratified meta-analyses, opening up multiple options: (i) to test for stratum-specific effects *(stratified association test)*, (ii) to combine stratified results together and to test for stratum-combined effects (*overall association test* [11]), (iii) to test for difference between stratum-specific effects (*difference test* [12]), or (iv) to test for joint effects accounting for potential GxS by using the sum of squared stratum-specific test statistics (*alternative joint test* [13]). The *alternative joint test* was shown to be equivalent to the *joint test* combining the main and the interaction effect for a dichotomous factor S [8]. Numerous variants with GxS have already been identified via stratified

$$Z_{Diff} = (\hat{\beta}_1 - \hat{\beta}_2)/\sqrt{se_1^2 + se_2^2}.$$

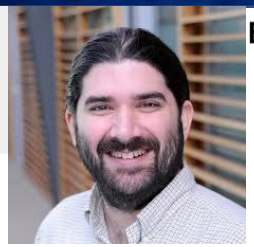
$$Z_{Diff} = (\hat{\beta}_1 - \hat{\beta}_2)/\sqrt{se_1^2 + se_2^2 - 2 \cdot Cov(\hat{\beta}_1, \hat{\beta}_2)}$$ and the $Cov(\hat{\beta}_1, \hat{\beta}_2)$

PLOS | GENETICS

BROWSE

OPEN ACCESS PEER-REVIEWED

RESEARCH ARTICLE

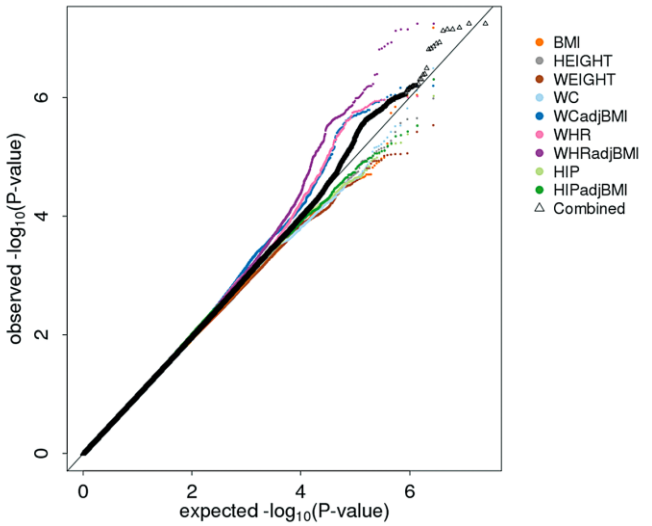


Sex-stratified Genome-wide Association Studies Including 270,000 Individuals Show Sexual Dimorphism in Genetic Loci for Anthropometric Traits

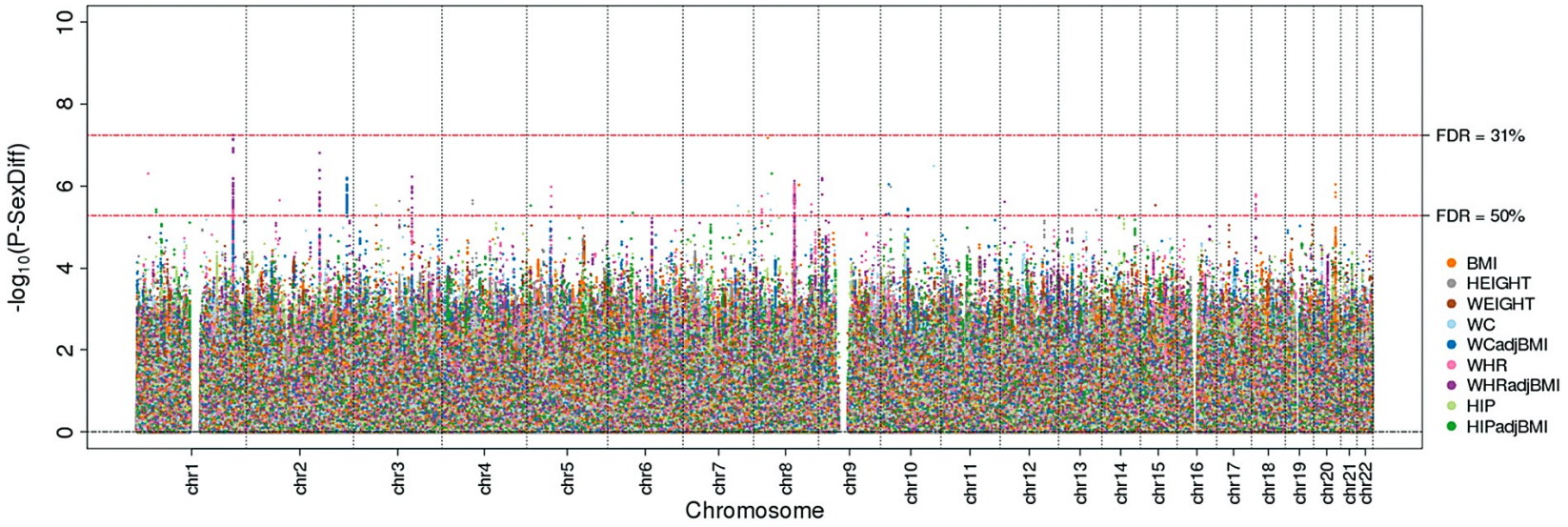
Joshua C. Randall, Thomas W. Winkler, Zoltán Kutalik, Sonja I. Berndt, Anne U. Jackson, Keri L. Monda, Tuomas O. Kilpeläinen, Tõnu Esko, Reedik Mägi, Shengxu Li, Tsegaselassie Workalemahu, Mary F. Feitosa, Damien C. Croteau-Chonka, [...], Iris M. Heid [view all]

Published: June 6, 2013 • https://doi.org/10.1371/journal.pgen.1003500

B



A



PLOS | GENETICS

BROWSE

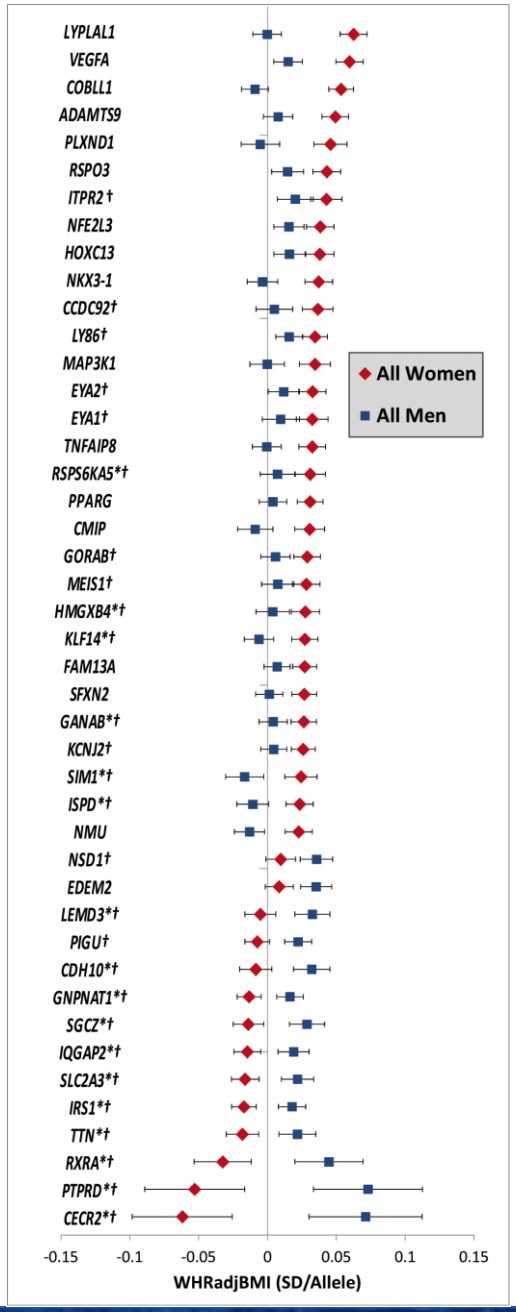
OPEN ACCESS PEER-REVIEWED

RESEARCH ARTICLE

The Influence of Age and Sex on Genetic Associations with Adult Body Size and Shape: A Large-Scale Genome-Wide Interaction Study

Thomas W. Winkler, Anne E. Justice, Mariaelisa Graff, Llilda Barata, Mary F. Feitosa, Su Chu, Jacek Czajkowski, Tõnu Esko, Tove Fall, Tuomas O. Kilpeläinen, Yingchang Lu, Reedik Mägi, Evelin Mihailov, [...], Ruth J. F. Loos [view all]

Published: October 1, 2015 • https://doi.org/10.1371/journal.pgen.1005378



What about X?

NATURE MEDICINE VOLUME 23 | NUMBER 11 | NOVEMBER 2017

EDITORIAL

nature medicine

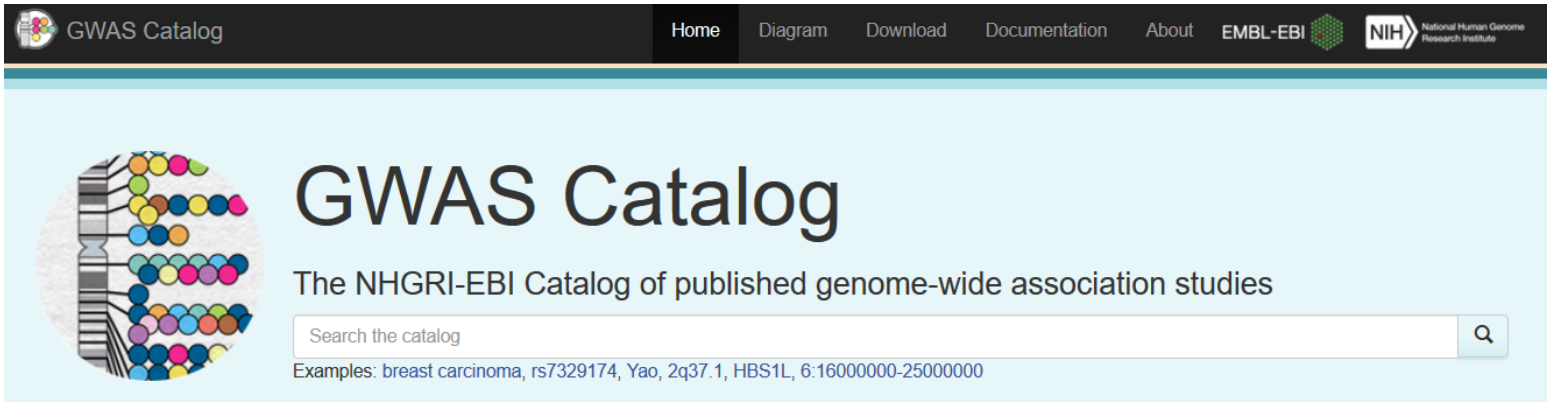
Accounting for sex in the genome

The X chromosome makes up about 5% of the haploid human genome, and carries just around 800 protein-coding genes out of our total of 20,000 such genes. Even so, in some genetics research, the X chromosome has featured prominently: mutations within it contribute to almost 10% of Mendelian disorders.

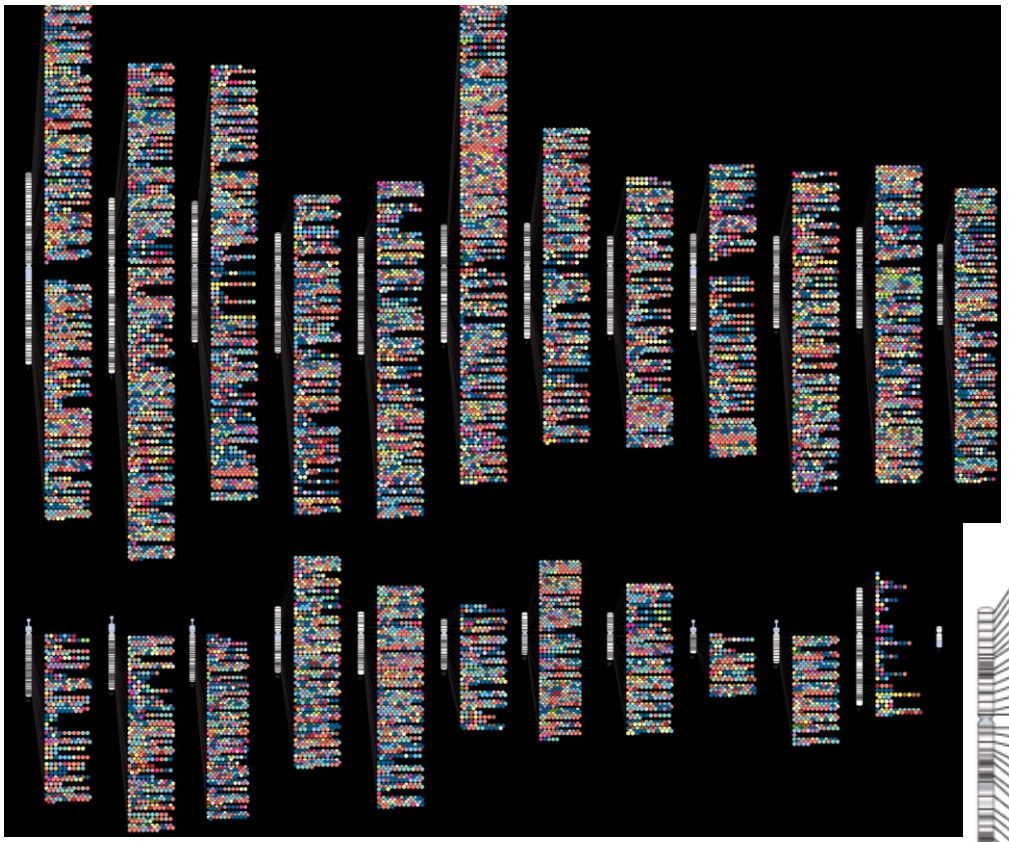
This disparity was highlighted earlier this year by Whitehead Institute Director David C. Page at the Keystone Symposia's meeting on Sex and Gender Factors Affecting Metabolic Homeostasis, Diabetes and Obesity.

Let’s fact check this

- GWAS catalogue
  - https://www.ebi.ac.uk/gwas/



As of 2019-01-31, the GWAS Catalog contains 3764 publications and 107785 unique SNP-trait associations.



537 of these associations (0.5%) are on the X chromosome
2 are on the Y chromosome

Why is this is case?

In the past, genotyping chips contained very few X-chromosome markers, which created a bottleneck on data. This has since improved, but the significance of variants on the X chromosome still remains harder to assess than for variants on autosomal chromosomes. One reason is simply that there are two copies of X in women and one in men, so the signals for variants on this chromosome obtained with standard array genotyping platforms are comparatively lower for men. Another reason is the phenomenon of X inactivation—the process by which one of the two X chromosomes is randomly silenced in women's cells. It is not yet possible for standard sequencing technologies to discern which genetic variants are on the silenced version of the X chromosome. To make matters more complicated, X inactivation can vary within the body.

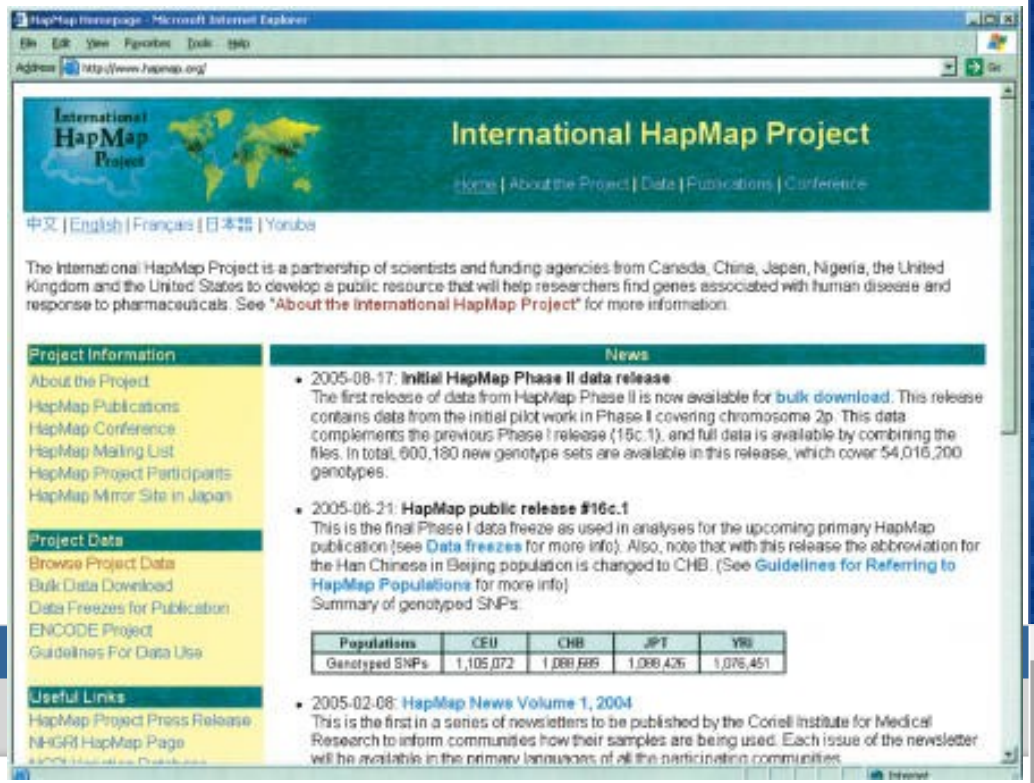
NATURE MEDICINE VOLUME 23 | NUMBER 11 | NOVEMBER 2017

2 points here

- Coverage on chips
  - Still a major issue
  - Many annotation errors
- Modelling of X effects
  - X Inactivation
  - Dosage compensation

A critical 3rd point

- HapMap2 never released X references for imputation



NCBI Resources How To

Variation

NCBI retiring HapMap Resource

June 16, 2016

A recent computer security audit has revealed security flaws in the legacy HapMap site that require NCBI to take it down immediately. We regret the inconvenience, but we are required to do this. That said, NCBI was planning to decommission this site in the near future anyway (although not quite so suddenly), as the 1,000 genomes (1KG) project has established itself as a research standard for population genetics and genomics. NCBI has observed a decline in usage of the HapMap dataset and website with its available resources over the past five years and it has come to the end of its useful life.

Dosage compensation

- Dosage compensation is the process by which organisms equalize the expression of genes between members of different biological sexes
- Different species have different methods
  - Random inactivation of one ♀ X (eg mammals)
    - not all genes along the X chromosome are subject to X-inactivation
    - active expression at some loci is required for homologous recombination with the pseudo-autosomal region (PAR) of the Y chromosome during meiosis
    - 10-25% of human X chromosome genes outside of the PARs show weak expression from the inactive X chromosome.
  - Two-fold increased transcription of a single ♂ X (eg Drosophila)
  - Decreased transcription of both hermaphroditic Xs by half (eg C. elegans)

X inactivation (the canonical version)

ScienceDirect

X-Inactivation

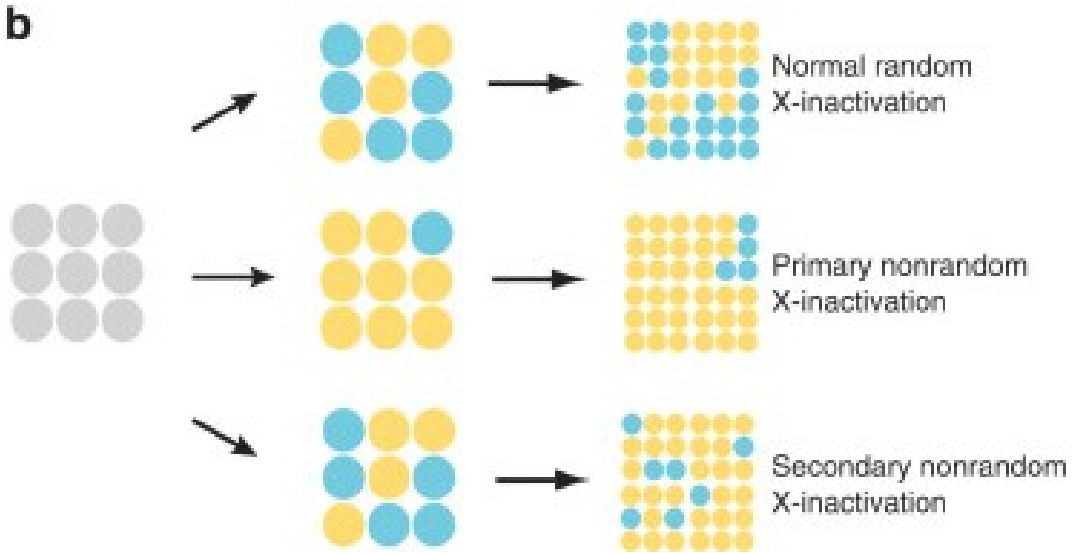
X-inactivation is a method of dosage compensation whereby somatic cells have one X-chromosome randomly repressed, or inactivated, at an early embryonic stage in development.

From: Handbook of Stem Cells, 2004

the inactive chromosome is silenced through methylation

X inactivation (the canonical version)

In humans inactivation happens during the late blastocyst stage, after implantation. It is theoretically random and inherited by the daughter cells



What this means analytically

- The variance of females for expression/effect of X linked variants will be ½ that of males
  - Using a test designed for autosomes will not account for this properly
  - To the extent that X has a substantial effect on a phenotype you are analysing you might expect to see $Var_{MALE} > Var_{FEMALE}$

Implications for QC & Type I error

X-specific considerations that are important to account for in GWAS include, but are not limited to: (1) correlation between X-linked genotype calling error rate and the sex composition of an assay plate, which can lead to plate effects that correlate with sex and, hence, with any sexually dimorphic trait; (2) X-linked variants being more likely to exhibit different effects between males and females [40], suggesting enhanced power of sex-stratified statistical tests; (3) power of the analyses being affected by the smaller allelic sample size (due to males carrying one allele and X-inactivation in females), reduced diversity on X and other unique population genetic patterns [41]–[47], and a lower density of X-linked SNPs on genotyping arrays; (4) quality control (QC) criteria need to account for sex information to prevent filtering the entirety or a large fraction of the chromosome [1], while at the same time accounting for confounding sex-specific effects; (5) sex-specific population structure leading to differential effects of population stratification (which could lead to false positives [48]–[50]) between X and the autosomes; and (6) application of association tests designed for the autosomes potentially leading to statistical inaccuracies.

Accounting for eXentricities: Analysis of the X Chromosome in GWAS Reveals X-Linked Genes Implicated in Autoimmune Diseases

Diana Chang, Feng Gao, Andrea Slavney, Li Ma, Yedael Y. Waldman, Aaron J. Sams, Paul Billing-Ross, Aviv Madar, Richard Spritz, Alon Keinan

Published: December 5, 2014 • https://doi.org/10.1371/journal.pone.0113684 • >> See the preprint

How can we analyse X properly?

DOI: 10.1002/gepi.22132

RESEARCH ARTICLE

WILEY Genetic Epidemiology

OFFICIAL JOURNAL INTERNATIONAL GENETIC EPIDEMIOLOGY SOCIETY www.geneticepi.org

Statistics for X-chromosome associations

Umut Özbek[1,2] | Hui-Min Lin[3] | Yan Lin[3] | Daniel E. Weeks[3,4] | Wei Chen[3,4,5] | John R. Shaffer[4] | Shaun M. Purcell[6,7,8,9,10] | Eleanor Feingold[3,4]

P, G, and S stand for phenotype, genotype, and sex

Regression model G1: $P \sim G\ (0,1)$. (1)

Regression model G2: $P \sim G\ (0,2)$. (2)

Regression model G1S: $P \sim G\,(0,1) + S$. (3)

Regression model G2S: $P \sim G\,(0,2) + S$. (4)

Regression model G1xS:

$$P \sim G\,(0,1) + S + G\,(0,1) * S. \quad (5)$$

Regression model G2xS:

$$P \sim G\,(0,2) + S + G\,(0,2) * S. \quad (6)$$

The statistics proposed by Clayton (2008) improve on these regression models (at least in theory) by using generalized linear model score tests based on genotype–phenotype covariance. They treat males the same as homozygote females (0,2 coding), but also account for variance differences. They do not lose power (in contrast to a stratified analysis) even if the phenotype varies between sexes as long as allele frequency does not (Clayton, 2008).

Since the Clayton 1-df statistic uses the (0,2) male genotype coding, we will refer to this statistic as the "C2" statistic.

. Clayton also proposed a regression generalization of C2, where phenotype is a dependent variable and sex is added as a covariate, which we will refer to as the "C2S" statistic.

Zheng et al. (2007) proposed a very different test statistic for X-chromosome association of a dichotomous trait, which is essentially a weighted average of separate male and female statistics. Their statistic (which we will refer to as the "Z" statistic) is

$$Z_{mfG}^2 = \left(\sqrt{\frac{n_f}{n_m + n_f}} Z_{fG} + \sqrt{\frac{n_m}{n_m + n_f}} Z_m \right)^2 \sim \chi_1^2, \quad (13)$$

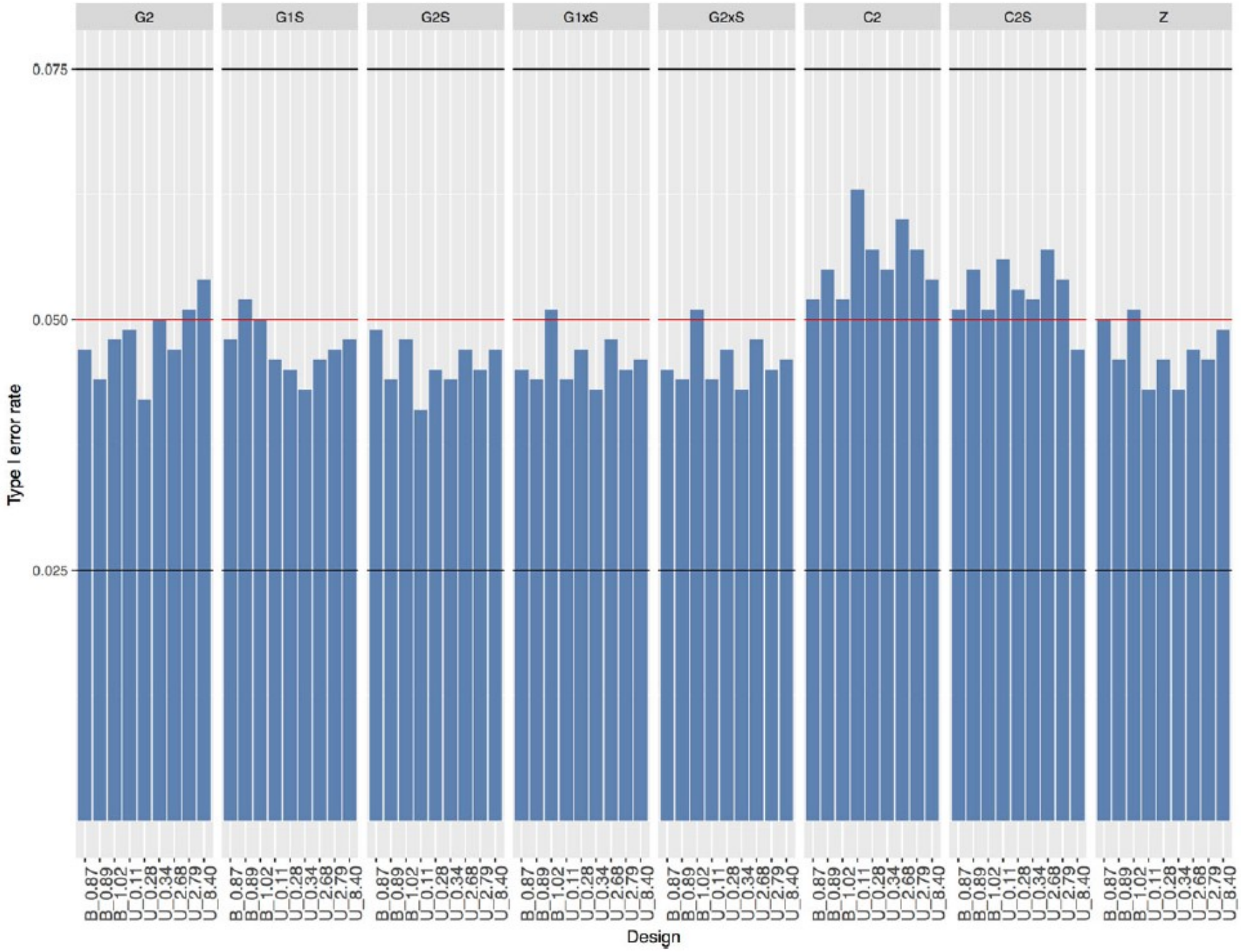


FIGURE 2 Type I error rates of the methods for dichotomous phenotypes and 12,242 real SNPs

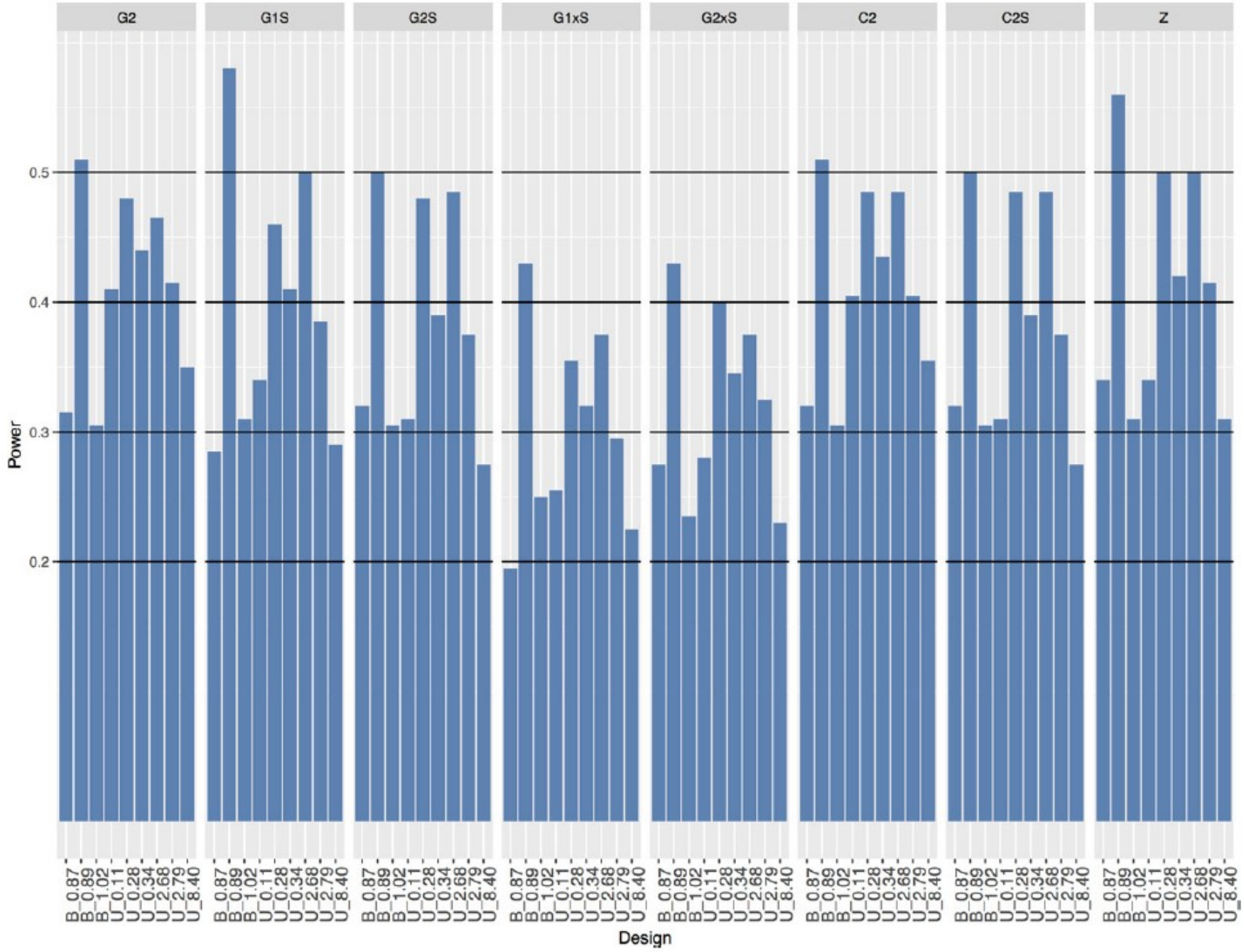


FIGURE 3 Power of the methods for dichotomous phenotypes

Type I error results for quantitative phenotypes are given in Supporting Information Table S4. Results are very similar to those for dichotomous phenotypes. For regression model G1, we observed very high type I error rates when male and female phenotype means were different (essentially equivalent to an unbalanced design). However, regression models with male genotypes coded as (0,2), and/or the models with a sex covariate have well-controlled type I error rates.

In quantitative phenotype power analysis (Supporting Information Table S5), among the methods that have well-controlled type I error rates, regression model G2S and Clayton's C2S statistic again have highest power when males with the B genotype have the same mean as homozygous BB females. However, when the B males' mean is the same as heterozygous AB females, G1S is more powerful.

My reading of this is these are currently the best models

Regression model G1S: $P \sim G(0,1) + S.$

Regression model G2S: $P \sim G(0,2) + S.$

```
G1S = plink --bfile mydata --linear --xchr-model 1
--covar mydata.cov --out Xchromosome

G2S = plink --bfile mydata --linear --xchr-model 2
--covar mydata.cov --out Xchromosome
```

0. Exclude all sex and haploid chromosomes from the analysis. If the 'genotypic', 'hethom', 'dominant', or 'recessive' modifier was used with --linear/--logistic, this mode is forced.

1. (default) Add sex as a covariate on the X chromosome; don't do anything else differently.

2. Add sex as a covariate on the X chromosome, and code male X chromosome genotypes 0/2 instead of 0/1. If any --condition{,-list} variants are on the X-chromosome (and 'dominant'/'recessive' has not been specified with --condition{,-list}), this coding change applies to them as well.

3. Add sex as a covariate on the X chromosome, and also add a dosage-sex interaction term to the model there. If a permutation test was requested without the --tests flag, this mode causes the X chromosome to be omitted from the permutation test.

Other software to think about

Journal of Heredity

Issues | More Content | Subject | Submit | Purchase | About | All Journal of Here

Volume 106, Issue 5

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, Alon Keinan Author Notes

Journal of Heredity, Volume 106, Issue 5, 1 September 2015, Pages 666–671, https://doi.org/10.1093/jhered/esv059

- Optional sex-aware genotype calling for the X chromosome from raw intensity data
- Quality control tailored to the X chromosome
- Optional sex-aware imputation for the X chromosome
- An array of statistical tests (with various options) for association of X-linked markers
- Association tests for genes on the X chromosome
- Gene-gene interaction tests for genes on the X chromosome and the autosomes
- A visualization suite for association results

A last word on the topic

The failure to assess the influence of sex chromosomes in studies of the genome doesn't necessarily boil down to a lack of tools: there is also a challenge of a lack of will. It takes a bit more effort to include sex chromosomes in certain genomic analyses, and so this step is sometimes skipped. Now is a time to reverse this trend of omission. There are no shortcuts to good science.

NATURE MEDICINE VOLUME 23 | NUMBER 11 | NOVEMBER 2017