

Accounting for sex in the genome

Genetic association studies of the human genome often omit the X chromosome because of the unique analytical challenges it presents. A concerted effort to undo this exclusion could offer medically relevant insights into basic biology that might otherwise be missed.

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. There is also a broad appreciation that certain illnesses, such as depression and most autoimmune diseases, occur more often in females than in males, suggesting an influence of the X chromosome (either directly or indirectly). Likewise, other diseases, such as autism, are more commonly diagnosed in males, underscoring that sex chromosomes might exert a significant influence on health. Despite these insights, the X chromosome is often less scrutinized in the era of population genetics analyses because of the unique statistical challenges it presents.

A literature search published several years ago found that just 242 among 743 genome-wide association studies (GWAS) in the review included the X chromosome in their analyses. It's not surprising then that, of the almost 300 traits explored using GWAS, only 15 of the 2,800 significant variants, or 0.5%, were reported on the X chromosome¹. 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. Riffing on the observed shrinking of the Y chromosome over time, Page remarked that, whereas it may take ten million years for the Y chromosome to disappear, it has taken only ten years of GWAS for the X chromosome to do so.

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.

The statistical methods available to manage these complexities require additional expertise and effort to be incorporated into studies, but new tools are becoming available. For example, in 2015, one team of researchers proposed a tool known as XWAS, a software suite for analysis of the X chromosome in genetic studies, including those examining genetic associations². The team behind this tool used it to reanalyze data from 16 GWAS of different autoimmune and related diseases; they discovered associations between several X-linked genes and illnesses, including an association between inflammatory bowel disease (IBD) and the *COSMC* gene encoding the core 1 β 3GalT-specific molecular chaperone³. Building from this finding, as well as the knowledge that early onset of IBD is seen

more often in men than in women, another group showed that deleting *Cosmc* in mice contributes to gut inflammation in males but not in females. The study also reported gut microbiome changes in the rodents resulting from loss of *Cosmc*⁴. A follow-up published earlier this year characterized functional domains of the COSMC chaperone protein⁵. It's too early to say whether these initial insights will pave the way forward for new IBD therapeutics, but they offer an interesting starting point for exploration. This example underscores that there may be a lot to be gained in terms of discovering drug targets by including sex chromosomes in analyses.

The apparent exclusion of the X chromosome extends to association analyses conducted with more recent DNA sequencing tools as well. As of May 2016, of the 41 published genetic association studies for complex traits using whole-genome or whole-exome sequencing data, 25 completely omitted the X chromosome from their analyses. The majority of the remaining 16 did not apply any specialized computational and statistical methods to accurately scrutinize the X chromosome for genetic variants associated with disease (A. Keinan, personal communication).

It is not just the X chromosome that has been neglected in genetic analyses. A review paper last year noted that the Y chromosome is “too often ignored by researchers but could potentially be the key to understanding the [coronary artery disease] prevalence differences between men and women” (ref. 6). And it's not only genetic association studies that need to take the influence of sex chromosomes into consideration. The agenda for a workshop facilitated by the US National Institutes of Health late last month on sex as a biological variable included a session on sex differences in gene expression. Studies in the last couple of years have, in fact, begun to explore the influence of the X chromosome and sex on gene expression^{7,8}. Earlier this summer, a study appearing in this journal offered a characterization of male and female transcriptional profiles associated with major depressive disorder across six brain regions⁹. These findings form an important starting point for understanding such associations, but there are relatively few publications as of yet and plenty of room for far more study in this area.

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.

1. Wise, A.L., Gyi, L. & Manolio, T.A. *Am. J. Hum. Genet.* **92**, 643–647 (2013).
2. Gao, F. *et al. J. Hered.* **106**, 666–671 (2015).
3. Chang, D. *et al. PLoS One* **9**, e113684 (2014).
4. Kudelka, M.R. *et al. Proc. Natl. Acad. Sci. USA* **113**, 14787–14792 (2016).
5. Hanes, M.S., Moremen, K.W. & Cummings, R.D. *PLoS One* **12**, e0180242 (2017).
6. Molina, E. *et al. Heart Lung Circ.* **25**, 791–801 (2016).
7. Kukurba, K.R. *et al. Genome Res.* **26**, 768–777 (2016).
8. Gershoni, M. & Pietrovski, S. *BMC Biol.* **15**, 7 (2017).
9. Labonté, B. *et al. Nat. Med.* **23**, 1102–1111 (2017).