An approach for quantifying allele-specific expression estimates on the X chromosome

Kimberly Olney[1,2+], Tanya Phung[1,2+], and Melissa A. Wilson[1,2]

1. School of Life Sciences, Arizona State University, Tempe, AZ,
2. Center for Evolutionary Medicine, Arizona State University, Tempe
+ these authors contributed equally

kcolney@asu.edu @olneykimberly github.com/olneykimberly

Background: X and Y share homologous regions

- Genetic males (46, XY) and females (46, XX) share highly similar genomes, only differing in the sex chromosomes (X and Y). To equalize dosage on the X chromosome, one of the two X chromosomes in genetic females is transcriptionally silenced; this process is called X chromosome inactivation (XCI)[1] (**Fig 1**).
- Between ~15-30% of genes on the inactive X chromosome escape silencing, resulting in biallelic expression of these genes on the X chromosome in genetic females[1].
- Allele-specific expression analysis requires heterozygous variants from DNAseq and counts from RNAseq. However, quantifying allele-specific expression on the X chromosome can be challenging due to the shared sequence homology between the X and Y chromosome that results in the mismapping of reads[2,3]. It is unknown how mismapping due to homologous regions between the X and Y chromosomes affects allele-specific expression analyses.
- **This work aims to quantify allele-specific expression estimates on the X chromosome by employing a sex chromosome complement alignment approach.**

Fig 1. X chromosome inactivation. Gray cells are expressing the paternal X copy (X^p). White cells are expressing the maternal X copy (X^m). Some tissues are patchy with sections of cells showing biased expression for either the maternal or paternal X copy, while other tissues are mosaic[4].

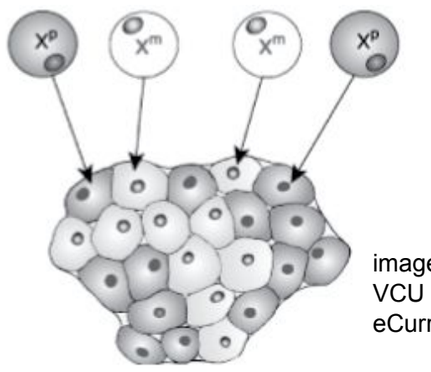


image credit: VCU eCurriculum

Methods: Sex chromosome complement alignment

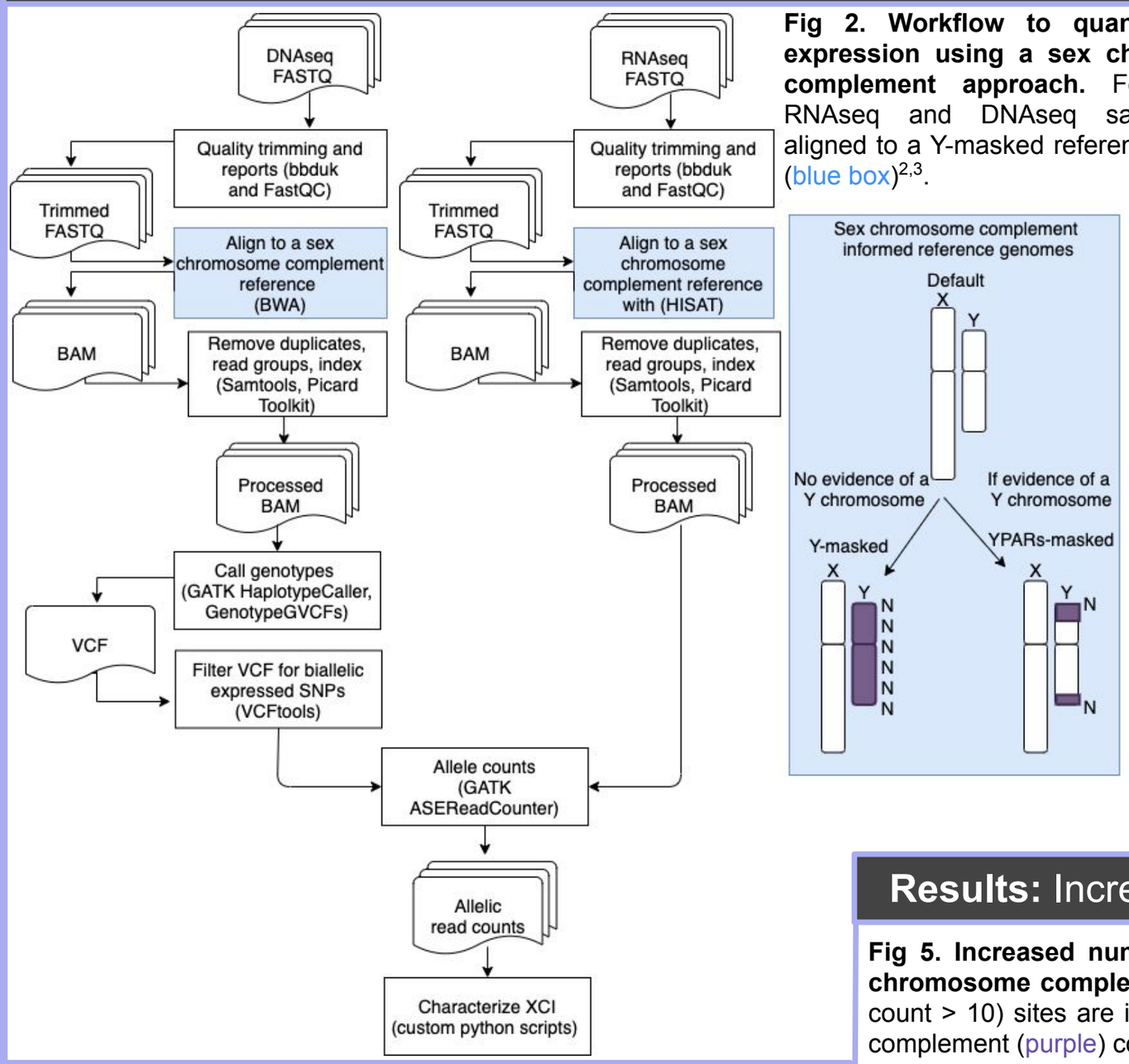
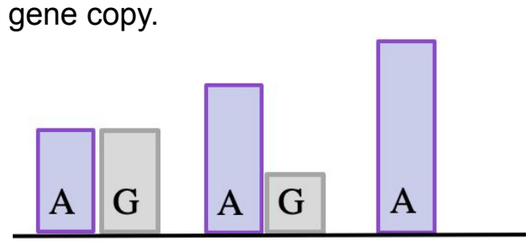


Fig 2. Workflow to quantify allelic expression using a sex chromosome complement approach. Female XX RNAseq and DNAseq samples are aligned to a Y-masked reference genome (blue box)[2,3].

Fig 3. Allele-specific expression estimates are used to determine if a gene is silenced or escapes XCI. If both alleles are expressed equally at heterozygous sites on the X chromosome, this gene escapes XCI. There may also be an allelic imbalance in which the gene escapes XCI but shows bias expression for one of the gene copies. A gene is silenced if it is only expressing one gene copy.

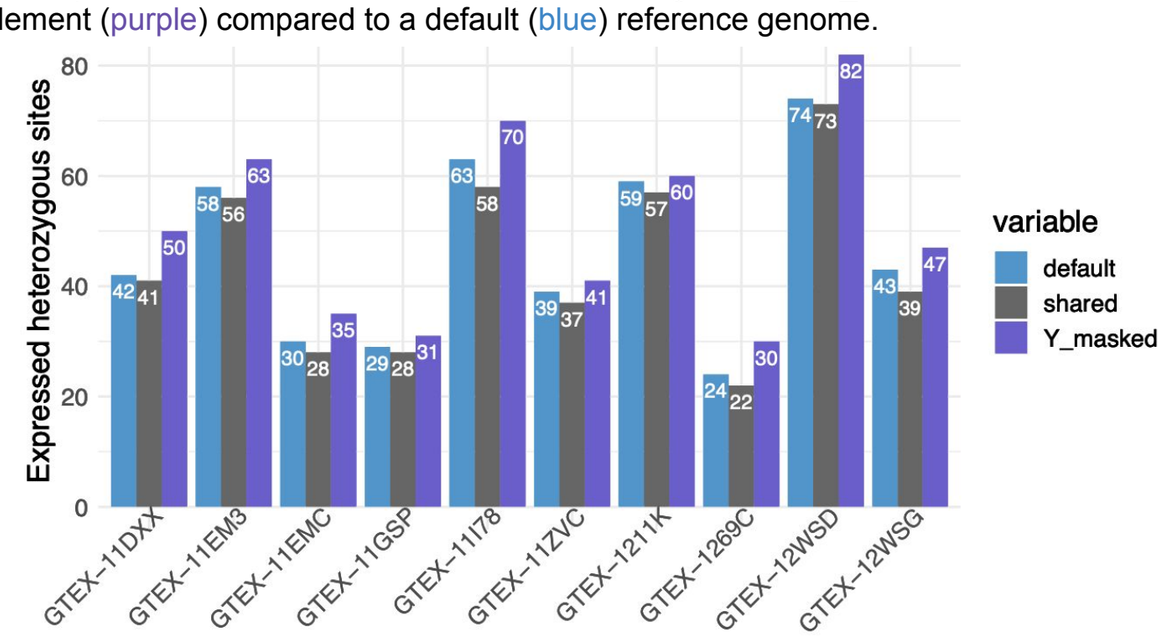


Equal expression (escaped) | Allelic imbalance (escaped) | Allele-specific expression (silenced)

- Here, we process ten female (46, XX) heart RNAseq and DNAseq samples from GTEx[5].
- RNAseq and DNAseq samples were aligned to both a default reference genome and a Y-masked reference genome (**Fig 2**).
- Compare the number of expressed (total allele count > 10) heterozygous sites between default and Y-masked approach (**Fig 5**).

Results: Increased number of expressed heterozygous sites

Fig 5. Increased number of heterozygous sites when samples are aligned to a sex chromosome complement reference genome. More heterozygous expressed (total allele count > 10) sites are identified when samples are aligned to a Y_masked sex chromosome complement (purple) compared to a default (blue) reference genome.



- Uniquely identified sites when using the sex chromosome complement approach are enriched in genes located within the pseudoautosomal regions (PAR1 and PAR2).

References

(1) Tukiainen et al. 2017. *Nature* 550, 244–248 https://doi.org/10.1038/nature24265.
(2) Olney et al. 2020. Biology of Sex Differences 11 (1): 42. https://doi.org/10.1186/s13293-020-00312-9.
(3) Webster et al. 2019. GigaScience 8 (7). https://doi.org/10.1093/gigascience/giz074.
(4) Phung et al. 2019. BioRxiv https://doi.org/10.1101/785105.
(5) The Genotype-Tissue Expression (GTEx) Project. Approved for project #8834 for General Research Use to MAW.

Acknowledgements

This work was supported by the National Institute of General Medical Sciences (NIGMS) of the National Institutes of Health (NIH) grant R35GM124827 to MAW and by the National Institute of Childhood Development (NICHD) of the NIH award FP00019155 to KO. KO was additionally supported as an ARCS Spetzler Scholar. We acknowledge Research Computing at Arizona State University for providing high-performance computing and storage resources that have contributed to the research results reported here.