Title:

Identifying, understanding, and correcting technical biases on the sex chromosomes in next-generation sequencing data

Authors and Affiliations:

Timothy H. Webster, Madeline Couse, Bruno M. Grande, Eric Karlins, Tanya Phung, Phillip A. Richmond, Whitney Whitford, Melissa A. Wilson Sayres

Supplementary Item	**Page**
Table S1. Samples included in this study.	2
Table S2. Coordinates of major X chromosome features in hg19.	3
Figure S1. Relative sequencing depth on the X and Y chromosomes in the 1000 Genomes Project high-coverage samples.	4
Figure S2. Relative mapping quality (MAPQ) on the X and Y chromosomes in the 1000 Genomes Project high-coverage samples.	5
Figure S3. Relative number of reads mapped to the X and Y chromosomes across different sequencing strategies.	6
Figure S4. Relative number of reads mapped to the X and Y chromosomes across in the 1000 Genomes Project high-coverage samples.	7
Archive S1. Zipped directory containing Snakemake pipeline and all files necessary to reproduce analyses.	Separate file

Table S1. Samples included in this study.

ID	Sex[a]	Sequencing[b]	Dataset[c]	Citation[d]
HG00512	M	E, LC, DC	Dataset 1	1
HG00513	F	E, LC, DC	Dataset 1	1
HG00419	F	DC	Dataset 2	2
NA20845	M	DC	Dataset 2	2
NA19625	F	DC	Dataset 2	2
NA19017	F	DC	Dataset 2	2
HG03052	F	DC	Dataset 2	2
HG01595	F	DC	Dataset 2	2
NA18525	F	DC	Dataset 2	2
NA20502	F	DC	Dataset 2	2
HG02568	F	DC	Dataset 2	2
NA18939	F	DC	Dataset 2	2
HG03642	F	DC	Dataset 2	2
HG00759	F	DC	Dataset 2	2
HG01112	M	DC	Dataset 2	2
HG01583	M	DC	Dataset 2	2
HG01051	M	DC	Dataset 2	2
HG00268	F	DC	Dataset 2	2
HG03742	M	DC	Dataset 2	2
NA19648	F	DC	Dataset 2	2
HG00096	M	DC	Dataset 2	2
HG02922	F	DC	Dataset 2	2
HG01565	M	DC	Dataset 2	2
HG01879	M	DC	Dataset 2	2
HG01500	M	DC	Dataset 2	2
HG03006	M	DC	Dataset 2	2

[a]Reported sex.

[b]Sequencing strategy: exome (E), low-coverage whole-genome (LC), and deep-coverage whole-genome (DC).

[c]Dataset membership for this study.

[d]1. The 1000 Genomes Project Consortium, 2015. 2. Sudmant *et al.*, 2015. References in main text.

Table S2. Coordinates of major X chromosome features in hg19.[a]

Chromosome	Start	End	Feature[b]
chrX	60000	2649520	PAR1
chrX	2649520	58632012	XAR
chrX	61632012	88395830	XCR1
chrX	88395830	92683067	XTR
chrX	92683067	154931044	XCR2
chrX	154931044	155260560	PAR2

[a]Coordinates in zero-index, half open format (i.e., bed format).
[b]PAR1: pseudoautosomal region 1; XAR: X-added region; XCR: X-conserved region; XTR: X-transposed region.

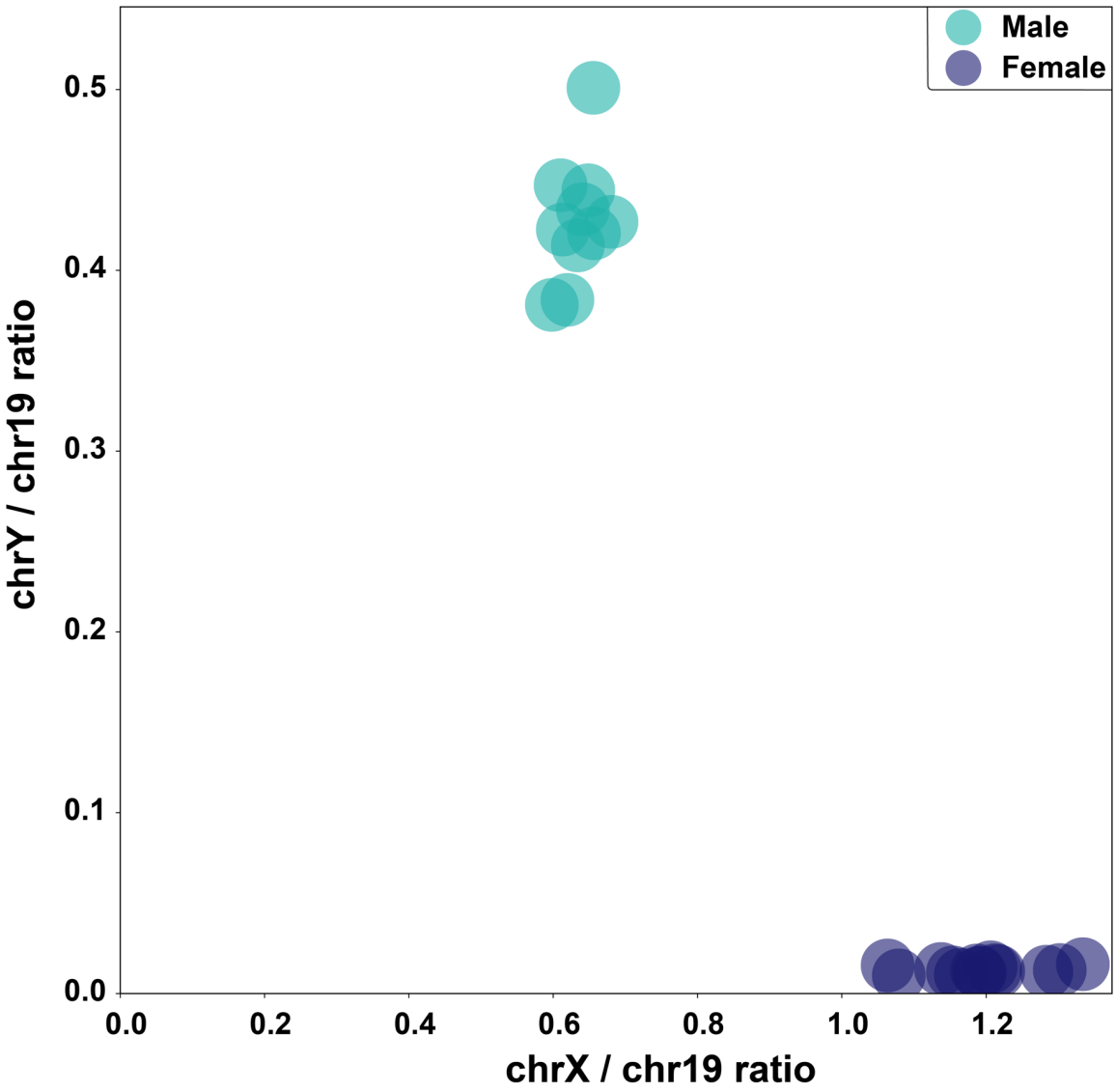

Figure S1. Relative sequencing depth on the X and Y chromosomes in the 1000 Genomes Project high-coverage samples. Males are plotted in green, while females are plotted in blue. Mean depth on chromosome 19 was used to normalize the sex chromosomes.

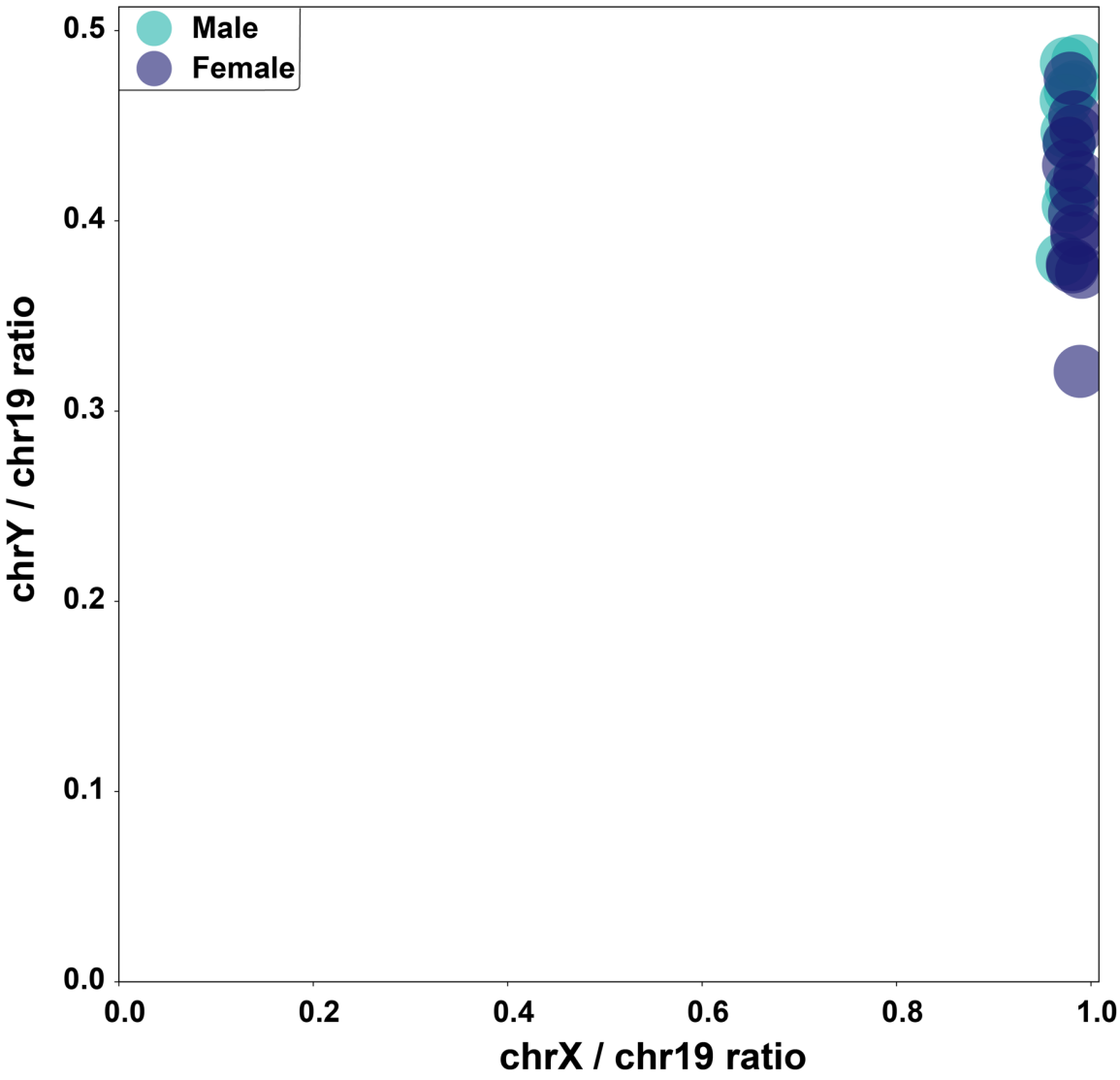

Figure S2. Relative mapping quality (MAPQ) on the X and Y chromosomes in the 1000 Genomes Project high-coverage samples. Males are plotted in green, while females are plotted in blue. Mean MAPQ on chromosome 19 was used to normalize the sex chromosomes.

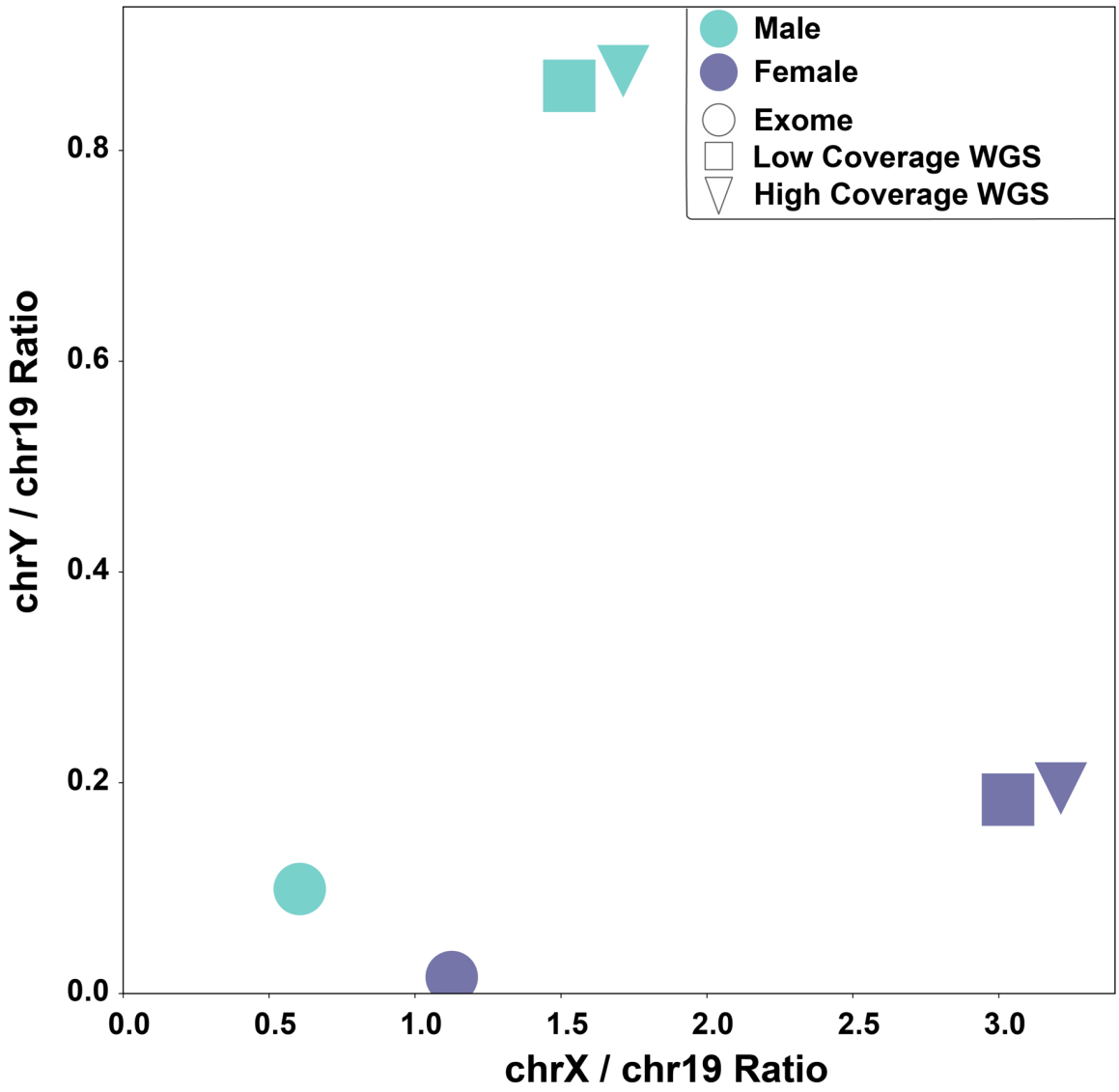

Figure S3. Relative number of reads mapped to the X and Y chromosomes across different sequencing strategies. Values come from exome (circles), low-coverage whole-genome sequencing (squares), and high-coverage whole-genome sequencing (triangles) for a single male (green) and female (blue) individual. The number of reads mapped to chromosome 19 was used to normalize the sex chromosomes.

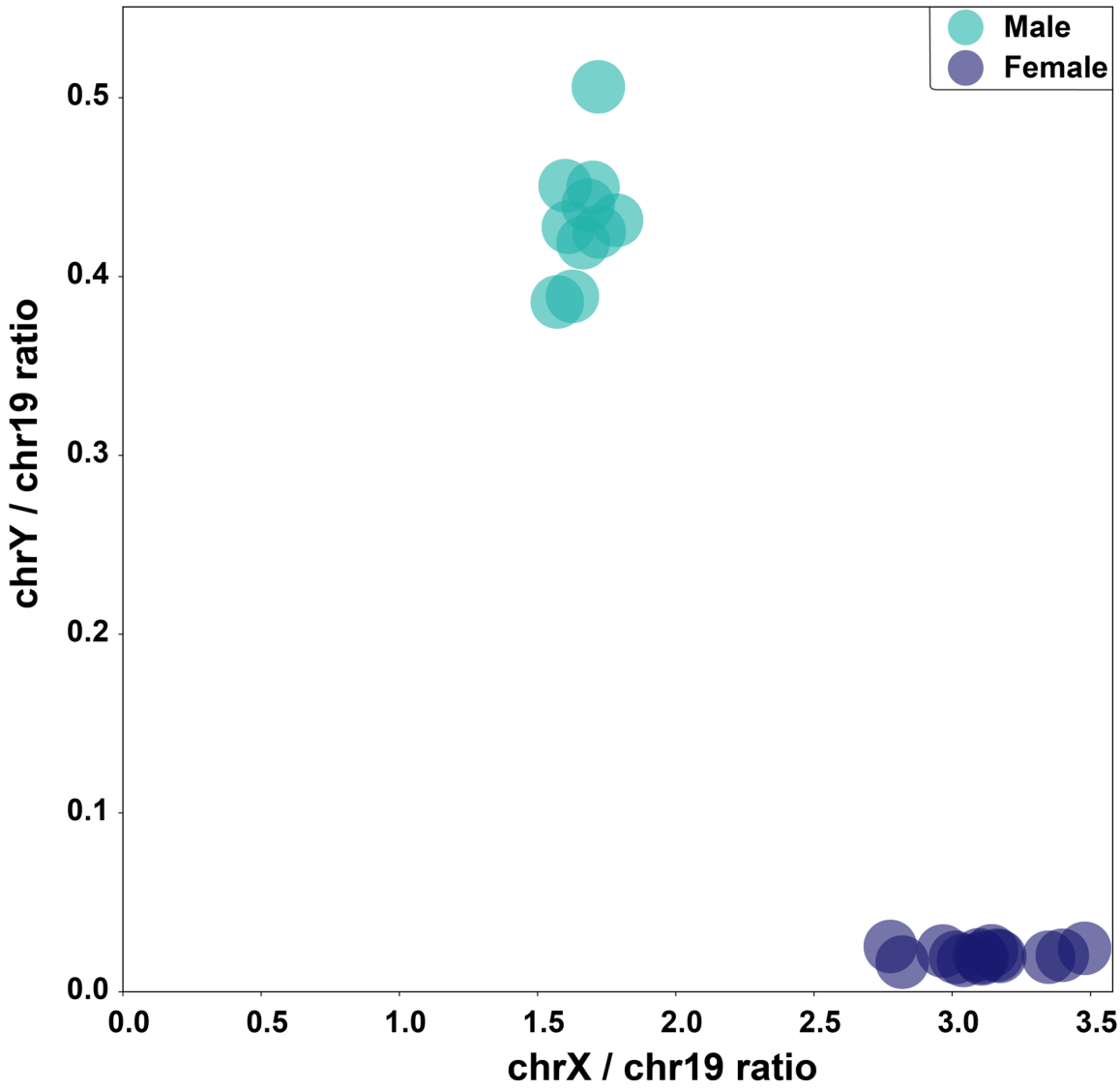

Figure S4. Relative number of reads mapped to the X and Y chromosomes across in the 1000 Genomes Project high-coverage samples. Males are plotted in green, while females are plotted in blue. The number of reads mapped to chromosome 19 was used to normalize the sex chromosomes.