



Measuring X inactivation skew for X-linked diseases with adaptive nanopore sequencing

Sena A Gocuk, James Lancaster, Shian Su, et al.

Genome Res. published online September 16, 2024
Access the most recent version at doi:[10.1101/gr.279396.124](https://doi.org/10.1101/gr.279396.124)

P<P	Published online September 16, 2024 in advance of the print journal.
Accepted Manuscript	Peer-reviewed and accepted for publication but not copyedited or typeset; accepted manuscript is likely to differ from the final, published version.
Creative Commons License	This article is distributed exclusively by Cold Spring Harbor Laboratory Press for the first six months after the full-issue publication date (see https://genome.cshlp.org/site/misc/terms.xhtml). After six months, it is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at http://creativecommons.org/licenses/by-nc/4.0/ .
Email Alerting Service	Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or click here .



To subscribe to *Genome Research* go to:
<https://genome.cshlp.org/subscriptions>

Published by Cold Spring Harbor Laboratory Press

Measuring X inactivation skew for X-linked diseases with adaptive nanopore sequencing

Sena A Gocuk^{1,2,3*}, James Lancaster^{4*}, Shian Su^{4,5}, Jasleen K Jolly⁶, Thomas L Edwards^{2,3}, Doron G Hickey², Matthew E Ritchie^{4,5}, Marnie E Blewitt^{4,5}, Lauren N Ayton^{1,2,3†}, and Quentin Gouil^{4,5†}✉

¹Department of Optometry and Vision Sciences, The University of Melbourne, Parkville Victoria, Australia

²Centre for Eye Research Australia, East Melbourne, Victoria, Australia

³Ophthalmology, Department of Surgery, University of Melbourne, Melbourne, Victoria, Australia

⁴Epigenetics and Development Division, The Walter and Eliza Hall Institute of Medical Research, Parkville, Victoria, Australia

⁵Department of Medical Biology, The University of Melbourne, Melbourne, Victoria, Australia

⁶Vision and Eye Research Institute, Anglia Ruskin University, Cambridge, United Kingdom

*These authors contributed equally as co-first

†These authors contributed equally as co-last

Abstract

X-linked genetic disorders typically affect females less severely than males due to the presence of a second X Chromosome not carrying the deleterious variant. However, the phenotypic expression in females is highly variable, which may be explained by an allelic skew in X Chromosome inactivation. Accurate measurement of X inactivation skew is crucial to understand and predict disease phenotype in carrier females, with prediction especially relevant for degenerative conditions.

We propose a novel approach using nanopore sequencing to quantify skewed X inactivation accurately. By phasing sequence variants and methylation patterns, this single assay reveals the disease variant, X inactivation skew, its directionality, and is applicable to all patients and X-linked variants. Enrichment of X-Chromosome reads through adaptive sampling enhances cost-efficiency. Our study includes a cohort of 16 X-linked variant carrier females affected by two X-linked inherited retinal diseases: choroideremia and *RPGR*-associated retinitis pigmentosa. As retinal DNA cannot be readily obtained, we instead determine the skew from peripheral samples (blood, saliva and buccal mucosa), and correlate it to phenotypic outcomes. This revealed a strong correlation between X inactivation skew and disease presentation, confirming the value in performing this assay and its potential as a way to prioritise patients for early in-

tervention, such as gene therapy currently in clinical trials for these conditions.

Our method of assessing skewed X inactivation is applicable to all long-read genomic datasets, providing insights into disease risk and severity and aiding in the development of individualised strategies for X-linked variant carrier females.

X-linked diseases | retinal disorders | skewed X inactivation | long-read sequencing | DNA methylation

Correspondence: gouil.q@wehi.edu.au

Introduction

X Chromosome inactivation (XCI) is a dosage compensation mechanism equalising the expression of the majority of X-linked genes between males and females (Van den Veyver, 2001). In 46,XX females, one X Chromosome is chosen stochastically as the inactive X (Xi) in a small number of cells early in embryonic development, and remains silenced throughout life (Lyon, 1961). Females are thus mosaics of cells where either one of the two X Chromosomes is inactive. Skewed X inactivation, whereby allelic Xi status in a pool of cells departs from the predicted 50:50 ratio, may explain the phenotypic variability in female carriers of X-linked conditions, when the gene in question is subject to X inactivation

(Morleo and Franco, 2008). Skew may be organism-wide (global skew observable in all or most tissues if the skew is due to biased X choice early in development), tissue-specific (when a skewed pool of cells give rise to a tissue, a common occurrence in highly clonal tissues like blood), or geographic (spatial heterogeneity within a tissue due to ontological patches of cells with the same Xi, like the coat of calico cats). XCI skew may be protective when the disease-causing variant is on the preferentially silenced X, or deleterious when on the preferentially active X. Skewed X inactivation is common, with estimates that 50% of females display skews of 65:35 or greater in blood (GTEx Consortium et al., 2017; BIOS consortium et al., 2019).

X-linked inherited retinal diseases, such as choroideremia and X-linked retinitis pigmentosa, most strongly affect males, with night blindness and peripheral vision loss during early adolescence which progress to central vision loss at later stages of the condition. Female carriers of X-linked inherited retinal disorders present with a spectrum of disease severities, ranging from near normal retinæ to severe retinal degeneration, the latter known as “male-pattern” retinal degeneration (Edwards et al., 2015; Gocuk et al., 2023). This variation and the observed prevalences is compatible with skewed X inactivation as a major factor of disease severity in females (Fry and MacLaren, 2023). Retinal phenotypes of female carriers can present as a mosaic pattern comprising regional patches of degeneration due to retinal mottling and/or pigmentary changes, which would again be compatible with a geographic XCI skew. Although it would be ideal to investigate the role of X inactivation in inherited retinal disorders directly in the affected retinal epithelium and photoreceptor cells, retinal biopsy from living patients is not justifiable due to the risk of vision loss directly and indirectly relating to the procedure, as well as the costs and resources involved in obtaining samples (Gocuk et al., 2023).

To date, the evidence for the importance of skewed XCI for X-linked inherited retinal disorders remains limited (Fahim and Daiger, 2016; Fry and MacLaren, 2023). As a proxy for the retina, X inactivation skew of female carriers of X-linked inherited retinal disorders has been previously reported in blood samples, based on DNA methylation measurements. This works because the inactive X Chromosome has hypermethylated CpG islands (Tribioli et al., 1992). These studies used conventional methods based on methylation-sensitive restriction enzymes and PCR fragment length analysis (Fahim et al., 2020). However, such methods provide incomplete information due to the lack of informative polymorphisms found in over 20% of individuals (Onozawa et al., 2015), quantitatively-limited analysis of just a few CpG sites and one or two genomic loci with the potential for PCR bias, the requirement for parental DNA to determine the origin of the variant and direction of the skew and additional genetic testing to ascertain the deleterious variant. A sequencing-based version of this test has recently been proposed (Johansson et al., 2023), directly reading the methylation of the *AR* and *RP2* CpG islands by nanopore sequencing of Cas9-targeted native DNA. DNA modifications trigger recognisable variation in the electrical current measured at the nanopore, enabling base-level modification analysis in long reads. Though this targeted test improves quantitation accuracy, it suffers from most of the same limitations outlined above. Furthermore, blood samples have been found to increase in XCI skew with age and may therefore not accurately reflect the XCI status of the retina and other tissues (Bittel et al., 2008; Sharp et al., 2000).

We set out to develop a new method to measure X inactivation skew based on long-read whole-genome sequencing, to simultaneously identify variants and investigate the contribution of skew to X-linked diseases in females. We used the Oxford Nanopore Technologies PromethION 24 platform and its

adaptive sampling capacity to enrich X Chromosome reads. We sequenced peripheral samples (blood, saliva and buccal mucosa) for a cohort of female carriers of X-linked inherited retinal disorders (choroideremia and *RPGR*-associated X-linked retinitis pigmentosa) to determine X inactivation skew in these tissues, infer the retinal skew and correlate with disease severity and retinal phenotype. We also include a female individual who did not have an inherited retinal disorder but for whom we were able to retrieve fresh retinal tissue (following enucleation for an invasive orbital tumour), allowing a direct comparison of skew in the retina and peripheral samples. Recent advances in the development of gene therapy for retinal diseases in males have sparked interest to determine whether female carriers, particularly those at risk of severe disease, should be considered for upcoming therapeutic intervention. Accurate measurement of skewed X inactivation may assist in determining the most suitable candidates.

Results

Optimising X Chromosome enrichment with adaptive

sampling. In order to increase the number of samples that could be run on one PromethION flow cell while still reaching a target coverage of 15–20X per sample, we made use

of the adaptive sampling capacity of the PromethION P24 (beta feature in MinKNOW). The first few hundred bases of each read are live-basecalled and mapped to the genome reference. If the sequence maps to the X Chromosome, sequencing of that read is allowed to continue, whereas if the sequence does not map or maps elsewhere, the read is deemed off-target and its sequencing is interrupted by voltage reversal to make the pore available for another fragment (Fig. 1A). To guide protocol optimisation, we estimated the relationship between read length and target coverage (see Methods). When targeting the entire X Chromosome (154 Mb), maximising read length is beneficial, up to the point where pore

blocking and decreased pore occupancy (longer fragments mean fewer molecules) become the dominant factors in reducing the output.

Our sequencing results corroborated this estimate, as libraries dominated by small fragments enriched poorly for X Chromosome reads (Fig. 1B, C). Two factors contributed to small fragments dominating the sequencing: either the DNA samples were degraded (buccal swabs had consistently worse DNA integrity than blood or saliva) or in intact samples the unsheared, high-molecular weight fragments could not be efficiently sequenced. Therefore, we strived to increase the average fragment lengths in the libraries by a combination of clean-ups (selective precipitation when DNA amounts allowed, bead-based clean-ups in the case of low input DNA) and shearing to 10–20 kb fragments (Fig. 1B). With optimised libraries, we obtained up to 90X coverage of the X Chromosome per PromethION flow cell, equivalent to a 3–4-fold enrichment over non-adaptive sequencing. We obtained at least 30X total coverage of the X Chromosome for each patient, across 2–4 samples per patient, and providing sufficient data for analysis of alleles and DNA methylation (Fig. 1D).

Calculating X inactivation skew by phasing alleles and

epialleles. Next we developed a method using the phasing of long reads into alleles (maternal and paternal, or haplotype 1 and haplotype 2 when the parent-of-origin is unknown) and their clustering based on DNA methylation (Xi being hypermethylated and Xa being hypomethylated) to determine skewed X inactivation (Figure 2A). We generated a dataset where the maternal and paternal alleles are distinct by crossing two different mouse strains, and where X inactivation is genetically skewed through the inheritance of a mutated *Xist* (Royce-Tolland et al., 2010). We made neural stem cells from the F1 animals and sequenced the DNA by nanopore accord-

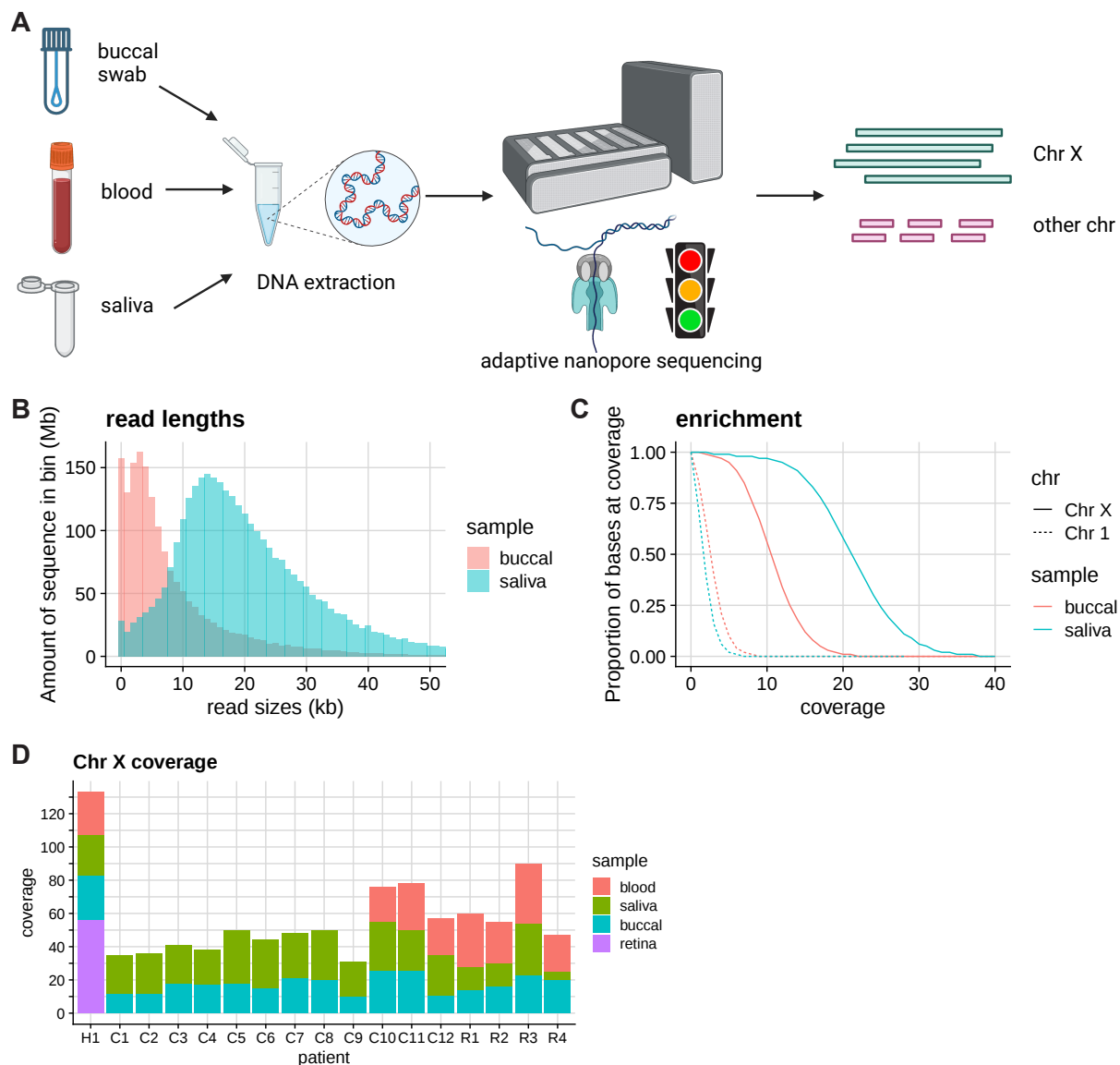


Fig. 1. X inactivation skew sequencing workflow. **A)** Adaptive nanopore sequencing of peripheral samples (blood, saliva, and buccal mucosa) to maximise coverage of the X Chromosome. Sequencing of fragments that do not map to the X Chromosome is terminated early. **B)** Degraded buccal DNA results in shorter read lengths and **C)** reduced X Chromosome enrichment performance. The read lengths for (accepted) reads mapping to the X Chromosome are plotted for patient C9 for illustrative purposes. Rejected reads are typically under 1 kb. **D)** X Chromosome coverage for each patient, split by sample. At least 30X combined X Chromosome coverage was obtained for each patient. H1: healthy individual; C1-C12: female carriers of choroideremia; R1-R4: female carriers of *RPGR*-associated X-linked retinitis pigmentosa.

ing to the ultra-long sequencing protocol.

After small variant calling with DeepVariant (Poplin et al., 2018) and read haplotyping with WhatsHap (Martin et al., 2016), visualisation of the *Xist* locus confirmed the correct SNV-based separation of haplotypes with the presence of the *Xist*^{ΔA} mutation as a ~750-base deletion ~120 nucleotides downstream of the *Xist* transcriptional start site on haplotype 2 (from the 129/Sv strain background). It also illustrated the contrasting DNA methylation patterns between the active and inactive X, correlating with the *Xist* gene expression: highly

195 methylated haplotype 2 with mCpGs in red (silenced, non-
functional *Xist*), and methylation-depleted haplotype 1 with
196 CpGs in blue, expressing a functional *Xist* (Fig. 2B). Of note,
197 the methylation pattern at the *Xist* CpG island is reversed
198 compared to the rest of the X Chromosome CpG islands: the
199 inactive X, which expresses *Xist*, is unmethylated (haplotype
200 1) while the active X (*Xist* silenced) is the methylated allele
201 (haplotype 2). For all other CpG islands on the X Chromo-
202 some, methylation-based clustering of reads assigns Xa to the
203 group of low-methylation reads and Xi to the group of high-

methylation reads (Fig. 2C). This clustering is performed using the density-based clustering algorithm DBSCAN within NanoMethViz (Su et al., 2021). The promoter CpG island of the X-linked *Phf6* gene illustrated how reads clustered in two well-separated groups (Fig. 2C).

Independent of their methylation profile, reads were also assigned to either haplotype 1 or haplotype 2. Regions of low SNV density or regions that are highly repetitive break up the haplotype phasing into haplotype blocks. As a result, haplotype 1 in one block may be the paternal allele while haplotype 1 in the next block may be the maternal allele (phase switching). Haplotype blocks are separated by genomic segments that cannot be haplotyped. In the mouse neural stem cell dataset, out of 469 annotated CpG islands (UCSC mm10 annotation), where we can expect DNA methylation on the inactive X, 454 were haplotyped (Fig. 2D, Supp. Table S1), allowing the assignment of reads to haplotype 1 and haplotype 2.

Of these haplotyped CpG islands, 105 CpG islands clustered into two methylation-level clusters, amounting to 2,300 reads assigned as Xa or Xi (Fig. 2E, Supp. Table S1). The remaining haplotyped CpG islands were not split into Xi and Xa methylation patterns for one of several reasons: similarly high or low methylation across all reads, lack of separation due to continuous variation in methylation, or insufficient read numbers to form a cluster (minimum set as 5 reads, Fig. 2E). We then tallied up the reads that were both haplotypes and clustered as Xa/Xi to calculate the X inactivation skew for each haplotype block. Large haplotype blocks may overlap multiple CpG islands, which allows pooling the counts of haplotype 1/haplotype 2 and Xa/Xi combinations for all CpG islands across the block. The haplotype block overlapping the most CpG islands covered 80 CpG islands (Fig. 2D). The

haplotype-block skew for each block i is expressed as:

$$\text{skew}_i = \frac{n_{\text{HP1_Xa}} + n_{\text{HP2_Xi}}}{n_{\text{HP1_Xa}} + n_{\text{HP2_Xi}} + n_{\text{HP1_Xi}} + n_{\text{HP2_Xa}}}$$

This represents the bias for haplotype 1 to be the active allele.

Since phase switching may occur between blocks, there is a symmetry in outcomes: preferential haplotype 1 silencing in one block is equivalent to preferential haplotype 2 silencing in another block. A folded binomial model allows to represent this symmetry in outcomes. In a series of binomial trials, a folded binomial model arises when the two categories of outcomes (successes and failures) can be distinguished, but we cannot say which were the successes and which were the failures (Mantel, 1970). Therefore we fold the skew: $\text{folded_skew}_i = \min(\text{skew}_i, 1 - \text{skew}_i)$, equivalent to mirroring the skew that is above 0.5 down to below 0.5 with the symmetry axis $x = 0.5$ (Fig. 2F). As expected in the genetically skewed mouse NSC dataset, skews for the haplotype blocks distributed around the extremes at 0 and 1, and the folded distribution peaked around 0 (Fig. 2F).

The global skew p for the sample is derived from the maximum likelihood estimate of the observed read counts in each block under the folded binomial model (Fig. 2G, (Gart, 1970; Mantel, 1970)). The numerical optimisation for p for the mouse NSC data yielded a global skew of 0.031, or a 97:3 ratio allele1_Xa:allele2_Xi, close to the expected 100:0.

Measuring X inactivation skew in human retinal and peripheral samples. Having established the skew analysis pipeline on ground truth data, we next applied it to a patient for whom we could collect retina, blood, saliva and buccal swab samples. This allowed us to investigate the correlation in skew between tissues, as well as the spatial heterogeneity of the retina by collecting discs across four quadrants of the retina (Fig. 3A).

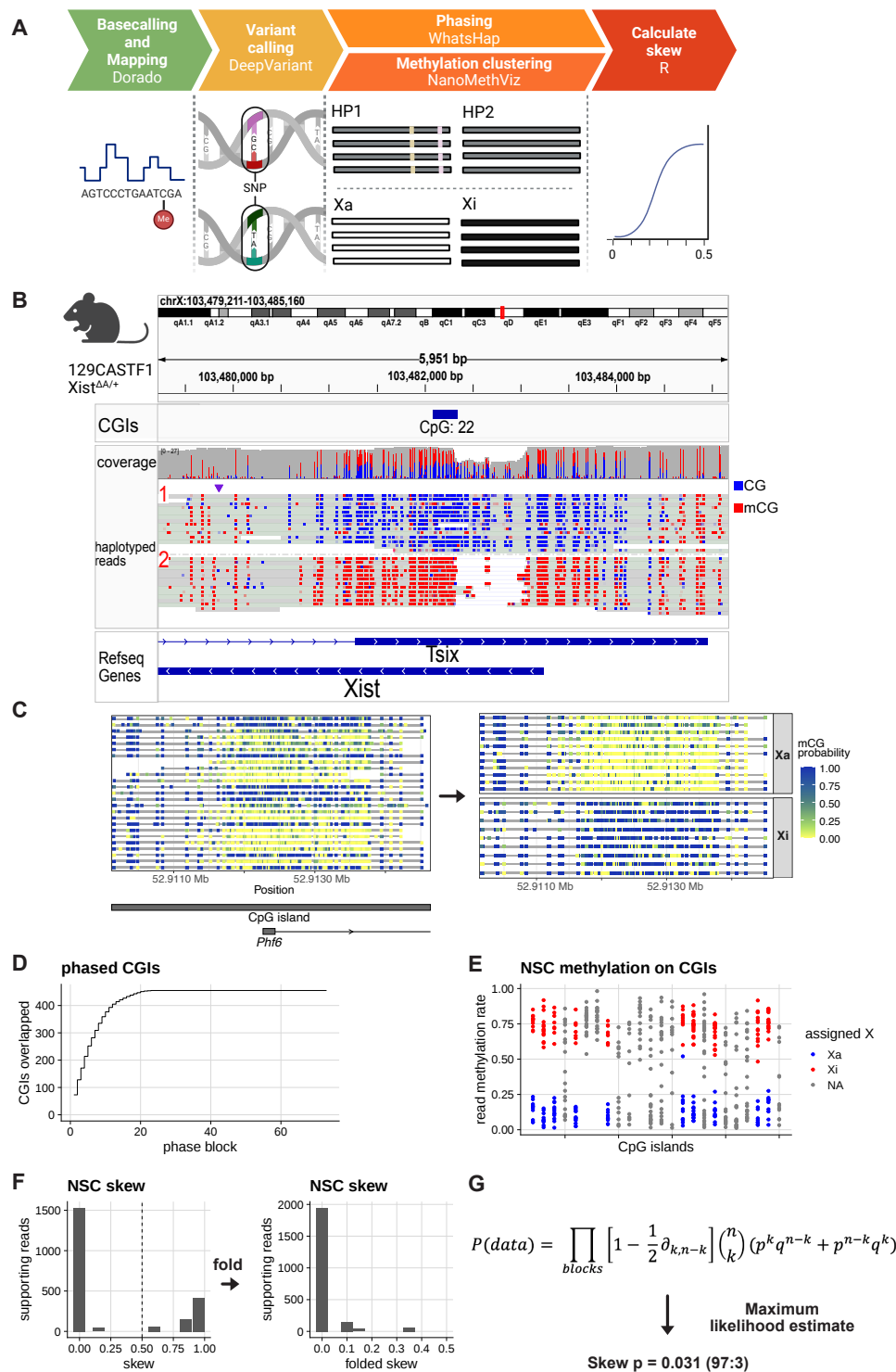


Fig. 2. Bioinformatic workflow to measure X inactivation skew with long-read genomics, demonstrated on ground-truth data. **A)** Five-step bioinformatic pipeline: basecalling and mapping reads including modified bases, SNV calling, phasing alleles and reads based on SNVs, clustering reads based on their methylation profile, tabulating the allele/epiallele combinations to calculate the skew. **B)** Validation on mouse data with genetically engineered skewed X inactivation: a deletion of the *Xist* A-repeat (haplotype 2) triggers its constitutive choice as the Xi, as confirmed by the DNA methylation profile of the *Xist* promoter (red: methylated CpG; blue: unmethylated CpG). **C)** Example of read clustering by methylation profile for the *Phf6* CpG island: reads split into two clusters, one with low methylation (CpG sites with low probability of methylation in yellow, cluster assigned as Xa) and one with high methylation (CpG sites with high probability of being methylated in blue, cluster assigned as Xi). **D)** Cumulative sum of CpG islands overlapped by haplotype blocks. Large haplotype blocks overlap many CpG islands, providing additional power to estimate the X inactivation skew in those haplotype blocks. **E)** CpG island methylation clustering results for one haplotype block. Each point represents a read. CpG islands with grey reads failed to split into Xi/Xa groups. **F)** Non-folded and folded histograms of the haplotype block-wise skews, binned in 10% bins and scaled by the underlying number of reads. Skew is measured as the proportion of reads supporting haplotype 1 as Xa. Because skewed X inactivation is genetically engineered in this experiment, haplotype blocks are expected to yield skews close to either 0 or 1. In the case of perfectly balanced X inactivation, haplotype blocks should return skews of 0.5. Folding is done around the $x = 0.5$ axis, to account for the random assignment of haplotype1/2 labels to the two alleles in each block. Folding would not be necessary if we could assign haplotype 1 to a single parental haplotype consistently. **G)** From the block skews, the sample skew p is calculated by maximising the joint probability of observing the data. $q = 1 - p$; for each haplotype block, k is the number of haplotype 1 reads that are Xa and haplotype 2 reads that are Xi ('successes') and n the total number of haplotyped and clustered reads ('trials').

The 10–20 kb reads optimised for adaptive sampling of the X Chromosome resulted in shorter haplotype blocks than with the ultra-long read mouse data, but overall 579 CpG islands (of 879 CpG islands) were phased into alleles (Fig. 3B, Supp Table S1).

The distribution of block skews revealed a strong X inactivation skew in blood ($p_{\text{blood}} = 0.16$), a slightly smaller skew in saliva ($p_{\text{saliva}} = 0.24$) and a more balanced X inactivation profile in the buccal swab and retinal discs (0.39 to 0.43, Fig. 3C). Each quadrant of the retina had a similarly balanced X inactivation profile, suggesting a well-mixed pattern of cells with the maternal or paternal X inactivated. The lack of correlation between the blood and retinal skews cautions against the ready extrapolation of retinal skew from blood skew, although it is common practice. Here the sampling of multiple tissues enabled us to rule out an organism-wide (global) skew in favour of a tissue-specific skew.

Among the samples collected, the buccal swab is the most likely to correlate with the retina. It is less susceptible to high clonal expansion than the cells present in blood and saliva (GTEx Consortium et al., 2017; Mengel-From et al., 2021). Although in our experience it was more challenging to extract high-quality DNA from it, the buccal swab may be the most informative for evaluating skew in the retina. In this particular patient's case, it was the one that most closely matched the non-skewed retinal X inactivation profile.

Using phase to orient skew and using skew to phase.

Although from block to block the identity of haplotypes can switch, for a single patient the haplotype for a block is the same across tissues. Therefore the correlation in unfolded skew between blocks informs on the direction of the skew: a positive correlation indicates the skew is in the same direction in both samples, whereas a negative correlation would indicate a reversal of skew between samples.

We observed a strong positive correlation between the blood and saliva block skews (Fig. 3D), indicating that the skew direction was the same in both highly skewed samples. The correlation with the more mildly skewed buccal swab sample remained positive, again pointing to the same direction of skew.

We then asked whether we could improve phasing of the X Chromosome by using the skew information: for a highly skewed sample, if in one block haplotype 1 is the predominantly active allele and in the next block haplotype 2 is the predominantly active allele, this would indicate that there was a phase switch between these two blocks. Therefore the unfolded, haplotype-block skews inform which haplotypes H1/H2 belong to the same allele across haplotype blocks, and this allows “scaffolding” of the haplotypes across the chromosome into parent1/parent2 alleles. Even though the haplotype blocks remain disjointed, they can be oriented so that haplotype 1 is consistently the same parental haplotype across haplotype blocks (higher-order phasing). Strong skews gave very high certainty in orienting blocks (Fig. 3E), but even for the buccal sample around 16% of the X Chromosome could be phased across haplotype blocks by using skew (Fig. 3F). Orienting the skew in each block is also helpful to assess whether a particular variant is likely to be on the predominantly active or inactive X.

One outlier haplotype block with a high number of phased and clustered reads could not be confidently oriented (outlier in Fig. 3E, blood panel). Further inspection revealed that this was due to an error in variant phasing in the difficult region harbouring duplicated *RHOXF2* and *RHOXF2B*: ambiguous read mapping over a 48.9-kb segmental duplication (Niu et al., 2011) led to a phase switch within the haplotype block (Supp. Fig. 1). Haplotype 1 was predominantly the Xa upstream of the ambiguous region, whereas it became predominantly the Xi downstream. Thus the skew analysis

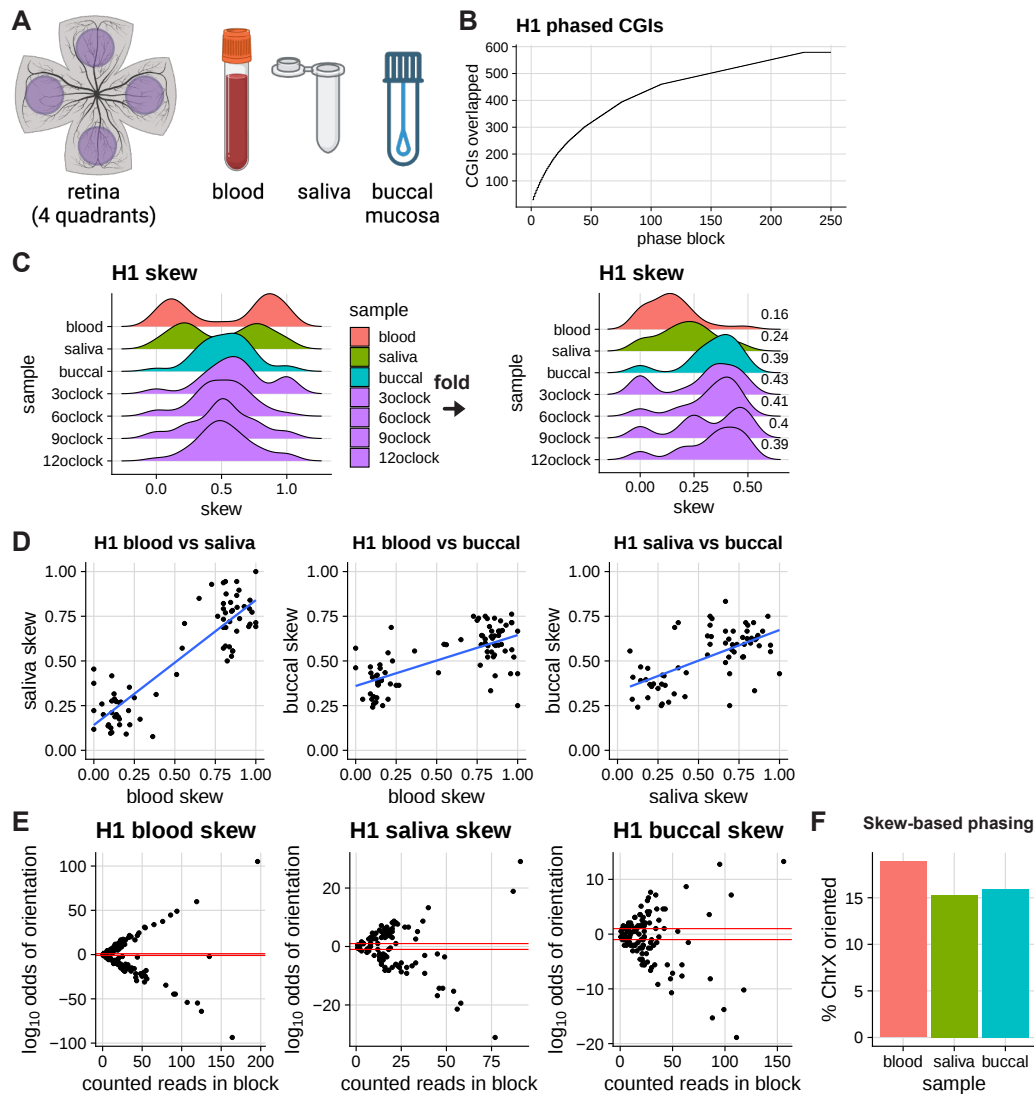


Fig. 3. X inactivation skew across blood, saliva, buccal swab and retina in a healthy patient. **A)** Collected samples: 4 quadrants of the retina from one eye (purple discs), blood, saliva and buccal swab. **B)** Cumulative sum of CpG islands overlapped by haplotype blocks. **C)** Non-folded and folded distributions of haplotype blocks' skews in the seven samples. Blood and saliva show a clear skew, whereas buccal swab and all four retina samples are not highly skewed. **D)** Positive correlation of non-folded haplotype blocks' skews across samples shows that the Xa allelic bias is consistent across tissues (the same X Chromosome in the preferentially active X in all skewed tissues). **E)** Skewed X inactivation allows scaffolding haplotype blocks into alleles. Log-likelihood ratio of haplotype 1 being the preferential Xa over Xi for each block. A red line is plotted at $y=1$ (haplotype 1 ten times more likely than haplotype 2 to be the preferential Xa) and $y=-1$ (haplotype 2 ten times more likely than haplotype 1 to be the preferential Xa). **F)** Up to 18% of the X Chromosome for patient H1 can be arranged in a consistent haplotype (higher-order phasing) based on haplotype block skew.

enabled the identification and correction of a variant phasing error across this complex region.

X inactivation skew in female carriers of inherited retinal disorders. Next we applied our method of measuring X inactivation skew to a cohort of patients with inherited retinal disorders: 12 carrying variants in the *CHM* gene causing

choroideremia and 4 carrying variants in the *RPGR* gene, a major cause of X-linked retinitis pigmentosa (XLRP, Table 1). There were two patient pairs carrying the same variants:

a mother and daughter pair, and two sisters. Seven carriers had grade 4 male-pattern degeneration (four with variants in the *RPGR* gene and three with variants in the *CHM* gene).

All other carriers had variants in the *CHM* gene and had relatively milder retinal phenotypes: one carrier with grade 1 (fine, near normal), four with grade 2 (coarse), and four with grade 3 (geographic) (Edwards et al., 2015).

Long-read sequencing identified disease-causing variants for all patients, including for those where there was no prior genetic diagnostic available. The long reads and chromosome-

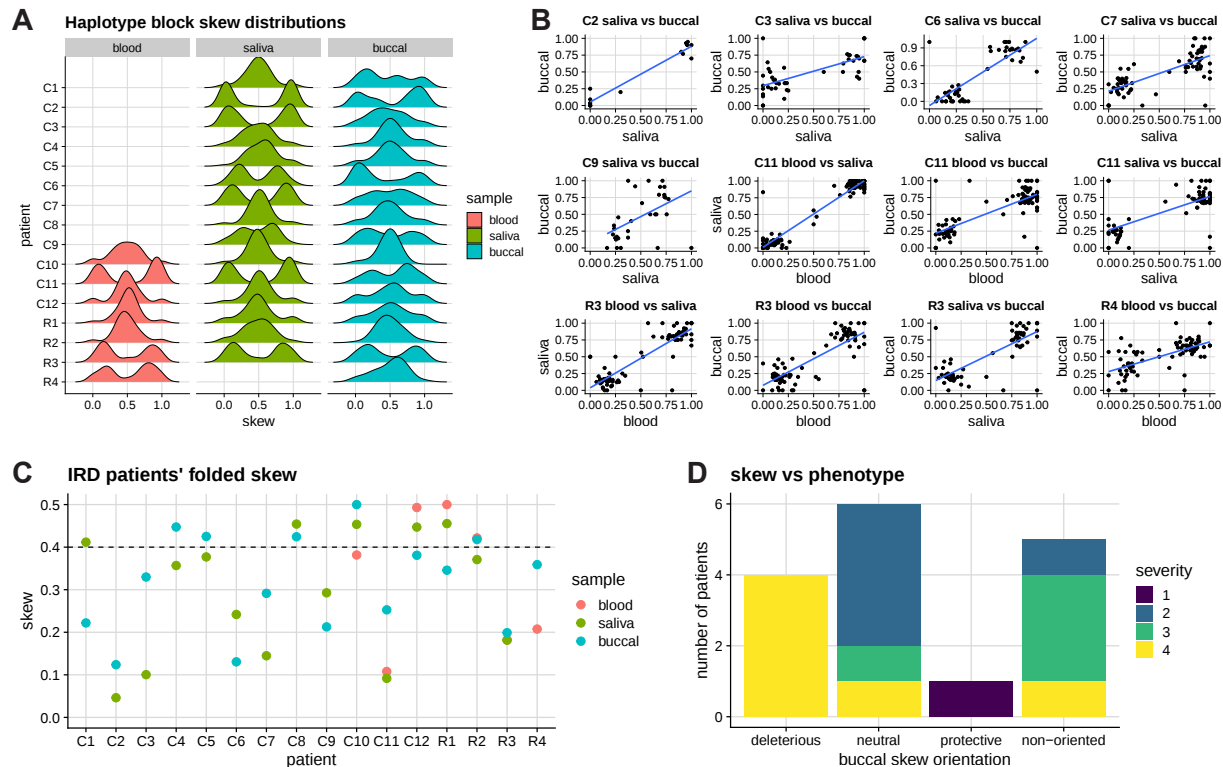


Fig. 4. X inactivation skew in a cohort of patients with X-linked inherited retinal disorders. **A)** Distribution of haplotype blocks' skews in multiple samples from patients carrying mutations in the *CHM* gene (C1-C12) or *RPGR* gene (R1-R4). **B)** Correlations of haplotype blocks' skew between samples indicates consistency in preferential Xa choice across tissues, for each patient where two or more samples show skew ($p < 0.4$). **C)** Summary of tissue skews for each patient. **D)** Correlation of buccal skew with phenotypic severity. The skew is classified as either deleterious (variant allele is preferentially Xa), protective (variant allele is preferentially Xi), neutral (no or little skew, $p \geq 0.4$). For 6 *CHM* variant carriers with skewed buccal swab samples, the skew could not be oriented (non-oriented category). Retinal severity grading presented in this graph is depicted as grades 1–4 to combine retinal phenotypes for choroideremia and *RPGR*-associated X-linked retinitis pigmentosa (Edwards et al., 2015; Nanda et al., 2018): grade 1 is fine and normal phenotypes, grade 2 is coarse and radial pattern phenotypes, grade 3 is geographic and focal pigmentary retinopathy phenotypes, and grade 4 is male pattern phenotypes for choroideremia and *RPGR*-associated X-linked retinitis pigmentosa, respectively.

ID	gene	variant	variant type	impact	grading
C1	<i>CHM</i>	c.529del; p.Glu177Lysfs*20	indel	frameshift	3
C2	<i>CHM</i>	g.85957852del; c.940+3delA	indel	splicing	4
C3	<i>CHM</i>	g.85957852del; c.940+3delA	indel	splicing	2
C4	<i>CHM</i>	c.1214_1215insC; p.Gln405Hisfs*13	indel	frameshift	2
C5	<i>CHM</i>	c.1780delC; p.Leu594Phefs*55	indel	frameshift	2
C6	<i>CHM</i>	g.86045359_86047506del; c.27_49+2125del	SV	deletion	4
C7	<i>CHM</i>	c.757C>T, p.Arg253Ter	SNV	stop	3
C8	<i>CHM</i>	g.85838231_85933652del; c.1167-22313_*26400del	SV	truncation	2
C9	<i>CHM</i>	g.85965588T>C; c.315-1536A>G	SNV	splicing	3
C10	<i>CHM</i>	c.1584_1587delTGTT; p.Val529Hisfs*7	indel	frameshift	2
C11	<i>CHM</i>	c.1584_1587delTGTT; p.Val529Hisfs*7	indel	frameshift	1
C12	<i>CHM</i>	g.84295184_84296004del; c.(1770+1_1771-1)_(*1_?)del	SV	truncation	3
R1	<i>RPGR</i>	c.2362_2366del, p.Glu788Argfs*45	indel	frameshift	4
R2	<i>RPGR</i>	c.619+5G>A	SNV	splicing	4
R3	<i>RPGR</i>	c.2442_2445del, p.Gly817Lysfs*2	indel	frameshift	4
R4	<i>RPGR</i>	c.2362_2366del, p.Glu788Argfs*45	indel	frameshift	4

Table 1. Variants in affected patients. Coordinates are relative to hg38 chrX (NC_000023.11), *CHM* MANE transcript NM_000390.4 and translated protein NP_000381.1, *RPGR* transcript NM_001034853.1 and protein NP_001030025.1. Patient C6's deep intronic variant in *CHM* has been experimentally validated as creating a cryptic splice site (Carss et al., 2017). Patients C1, C4 and C8 had not received a genetic diagnosis prior to the long-read sequencing performed in this study. C2 and C3 are sisters, C10 and C11 daughter and mother, R1 and R4 are mother and daughter.

wide sequencing approach resolved multiple types of events: exonic SNVs, intronic SNVs predicted to affect splicing, and large structural variants removing parts of the gene (Table 1). The availability of longer read lengths for certain samples had a strong effect on increasing the total haplotype block length and the haplotype block N50 (Supp. Table S1). For example for patient C9 saliva (read N50 of 18 kb), 81% of the X Chromosome and 692 CpG islands could be haplotyped. By contrast, samples with more fragmented reads and lower coverage had only 55-58% of the X haplotyped. Higher coverage was associated with a higher proportion of haplotyped CpG islands that could be clustered into Xa and Xi reads, and thus a higher number of informative reads for the skew calculation.

All but one of the 16 patients presented at least one sample with a skew below 0.4 (60:40 or greater, Fig. 4A, B and C). For some patients the extent of the skew varied between samples (Fig 4A and C). For example, C3 demonstrated a highly skewed saliva sample (0.10, or 90:10) whereas their buccal swab sample had a less pronounced skew (0.33). However, when two of a patient's samples were skewed (< 0.4), in all cases they were skewed in the same direction (Fig. 4B).

To correlate skew and phenotype, we used the buccal swab skew as the most likely to reflect the retina's skew, as it is less sensitive to age-related skewing than blood and saliva (GTEx Consortium et al., 2017; Mengel-From et al., 2021). Among the ten severely affected patients (graded as 3 and 4), nine (90%) had a skew of 60:40 or greater, and five (50%) had a skew of 75:25 or greater (Fig. 4C and D). Relative to a baseline expectation of 5% of the population with skews 75:25 or greater (GTEx Consortium et al., 2017), we concluded that there was an over-representation of skew among severely affected patients ($p\text{-value} = 6 \times 10^{-5}$, one-sided binomial test).

We next oriented the skew relative to the allele that carries the variant. By contrast with the skew-based phasing we demon-

strated previously, here we do not focus on scaffolding a consistent parental haplotype across haplotype blocks, but rather we focus on a single haplotype block, the one that carries the diseases-causing variant (e.g. on haplotype 1). Knowing the global skew for the sample, and the ratio of reads supporting haplotype 1 or haplotype 2 as the active X in this block, we can orient the skew relative to the variant. The skew was deemed deleterious when the predominantly active X is the one carrying the variant, and protective when the variant is instead on the predominantly inactive X (Fig. 4D). For all skewed patients with *RPGR* variants, the skew could be oriented, and all were deleterious (3 out of 3, one patient not skewed). Expressing the skew as 'deleterious allele': 'healthy allele', patients R1, R3 and R4 had ratios of 65:35, 80:20 and 64:36, respectively. For *CHM* variant carriers however, skew orientation was less successful as CpG islands are more sparse in this genomic region and the length of the haplotype blocks over *CHM* were often limited by stretches of low complexity sequence. Out of 8 skewed patients (4 were not skewed), two could have their skews oriented: one was strongly deleterious (87:13), consistent with the male pattern/grade 4 phenotype, and one was strongly protective (25:75), consistent with the mildest phenotype observed (fine phenotype/grade 1). Therefore out of 5 oriented skews, all correlated with the phenotype ($p\text{-value} = 0.031$, one-sided binomial test). By contrast, patients with a balanced buccal swab X inactivation ratio (neutral skew) were split across phenotypic severity classification: four coarse phenotype (grade 2), one geographic phenotype (grade 3) and one male pattern phenotype (grade 4). These data suggest that a strong deleterious skew measured in peripheral tissues could be predictive, but a neutral skew does not necessarily mean disease will not be severe.

Investigating the correlation of X inactivation skew

correlation and disease severity in patient pairs. We

further explored the relationship between choroideremia disease severity and X inactivation skew by focusing on a sister and a mother-daughter pair, each pair sharing the same disease-causing variant as well as 50% of their genetic material (5). In addition to explaining differences in overall disease severity, skewed X inactivation may also be able to

explain the geographical patterns of retinal degeneration observed in choroideremia and RPGR-linked retinitis pigmentosa (Fig. 5A and B). In particular, geographic choroideremia and focal pigmentary pattern in retinitis pigmentosa could reflect local patches of cells with elevated deleterious skews, while male patterns may reflect uniformly deleterious skew across the tissue.

In the first pair, carrying a 4-nucleotide deletion in *CHM* causing a frameshift, the non-skewed (50:50) 34-year-old female C10 presented with coarse (grade 2) phenotype whereas her 59-year-old mother C11 with a strong protective 75:25 skew had a milder fine (grade 1) phenotype (Fig. 5C and D). For a degenerative disease and given the common mutation, the mother's higher age would be expected to cause a more severe phenotype than her daughter's; in this case it may have been counteracted by the mother's skewed X inactivation, favouring the expression of the healthy allele, which provided her with protection against choroideremia. While her daughter is affected, the absence of a strong deleterious skew may be beneficial in avoiding a stronger male pattern phenotype.

In the second pair, a one-nucleotide deletion in the *CHM* coding sequence caused both sisters C2 (73 years old) and C3 (64 years old) to develop choroideremia, but to very different extent (Fig. 5E and F). C2 had a very strong 88:12 skew and a grade 4/male pattern retinal degeneration, suggesting that the skew was deleterious. Unfortunately the *CHM* variant

was not contained within a haplotype block with informative CpG islands, so the direction of the skew could not be formally ascertained. However, we could determine that her sister C2 had a skew in the opposite direction (33:67), and therefore protective, which matched with her much milder grade 2/coarse phenotype. These examples thus illustrate the high explanatory power of skewed X inactivation in X-linked diseases.

Discussion

In this study we demonstrate the use of long-read genomic sequencing to measure skewed X Chromosome inactivation. The same data also allows calling of small and large variants: we detected pathogenic intronic and exonic small variants, as well as large structural variants, in a cohort of female patients affected by X-linked retinal disorders. It confirmed previous work outlining the superiority of long reads over short reads to detect variants in the repetitive exon 15 of *RPGR* (Carss et al., 2017; Yahya et al., 2023; Bonetti et al., 2023; Vaché et al., 2024) or structural variants in *CHM* (Fadaie et al., 2021) among other genes (Nakamichi et al., 2023). Where typically multiple assays might be necessary to obtain this information (short-read exome sequencing, short-read genome sequencing, genome CNV microarray, restriction-enzyme/PCR-based tests of skew (Caylor, 2023)), long-read sequencing combined with our new skew analysis method has the potential to replace all of those with one workflow, accelerating the diagnostic journey and reducing its cost. As more and more long-read population genomics projects get underway, factoring in skewed X inactivation (or epigenotype) is likely to improve the genotype-phenotype correlations in females.

Our method takes a chromosome-wide approach, drawing data from hundreds of CpG islands to derive an accurate estimate of the X inactivation skew in the sequenced sample. All

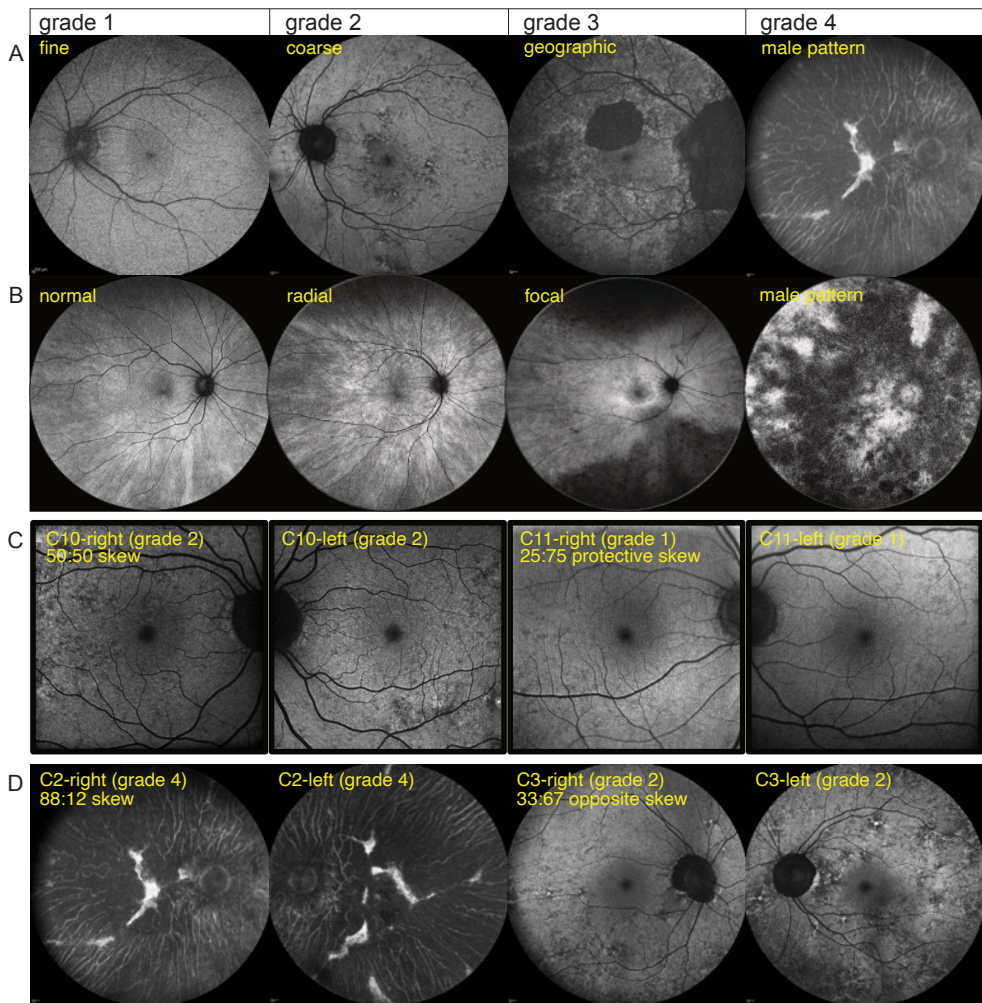


Fig. 5. Clinical phenotypes of *CHM* and *RPGR* variant carriers and correlation with X inactivation skew. **A)** Classification of retinal disease in female carriers of choroideremia based on 55-degree retinal fundus autofluorescence images, illustrative of the classification by [Edwards et al. \(2015\)](#): fine, coarse, geographic and male pattern degeneration. **B)** Classification of retinal disease in female carriers of *RPGR*-associated X-linked retinitis pigmentosa based on 55-degree retinal fundus autofluorescence images, illustrating the classification by [Nanda et al. \(2018\)](#): normal, radial pattern, focal pigmentary retinopathy and male pattern degeneration. **C)** Discordant clinical phenotypes of 34-year-old female C10 and her 59-year-old mother C11 carrying the same pathogenic *CHM* variant. C10 presents with coarse (grade 2) phenotype and C11 presents with fine (grade 1) phenotype. C11 had a pronounced buccal swab X inactivation skew (25:75) in favour of expressing the healthy allele, consistent with a milder disease phenotype. 30-degree fundus autofluorescence images of right and left eyes are depicted for both carriers. **D)** Discordant clinical phenotypes of 73-year-old female C2 and her 64-year-old sister C3, carrying the same *CHM* variant. C2 presents with male pattern degeneration (grade 4, picture also used to illustrate the grade 4 classification in panel A) and C3 presents with a coarse (grade 2) phenotype. C2 and C3 had skews in opposite directions, but the direction relative to the variant could not be determined. 55-degree fundus autofluorescence images of right and left eyes are depicted for both carriers.

samples are informative, contrary to traditional tests based on a single locus that is not sufficiently polymorphic in 20% of cases ([Fahim et al., 2020](#); [Johansson et al., 2023](#)). By using adaptive sampling to enrich for X-Chromosome reads, we obtain a 3–4-fold reduction in sequencing cost per sample while keeping the sample preparation very straightforward (standard ligation-based library preparation protocol). Adaptive sampling is independent of methylation and therefore does not impact skew quantification. However, our analysis is equally applicable to whole-genome, unbiased sequencing as long as X Chromosome coverage is 15 or above. This lower limit on coverage is imposed by the requirement to cluster reads overlapping CpG islands into two groups based on their methylation pattern. With lower coverage, the two groups may not be sufficiently sampled and cluster definition becomes more difficult resulting in fewer informative CpG islands (Supp. Table S1). When possible, sequencing longer reads (20 kb+) has multiple advantages: increases in adaptive sampling efficiency, in the overall proportion of the X Chromosome that is haplotyped, in the haplotype block lengths,

in the number of haplotyped CpG islands, and in the number of CpG island-overlapping reads that are both clustered and phased. Longer reads thus provide more data for estimating the skew, and also facilitate orienting the skew relative to a variant of interest. It is important to note that skew can vary from tissue to tissue within one individual (GTEx Consortium et al., 2017), and certain tissues become more clonal/skewed with age (Juchniewicz et al., 2023; Hatakeyama et al., 2004). Therefore there is a risk in extrapolating to a different, inaccessible, disease-relevant tissue. This risk can be mitigated by sampling multiple tissues. In our study we prioritised the buccal mucosa skew to extrapolate to the retina, while the saliva and blood skews could provide additional evidence of a global, organism-wide skew and increase the confidence in the extrapolation. Orienting the skew relative to the disease-causing variant in the absence of trio data varied in difficulty according to the locus. For a CpG-island rich, repeat-poor locus such as *RPGR*, in all patients (4/4) the haplotype block containing the variant spanned several informative CpG islands and therefore the skew could be oriented with high-confidence without any trio data. By contrast, phasing of the *CHM* locus was often interrupted by complex regions or runs of homozygosity. With the nearest CpG island at the *DACH2* promoter 300 kb away from the 3' end of the *CHM* coding sequence, for 6 out of 12 patients the haplotype block carrying the variant did not overlap a CpG island. Two of the skews could be oriented using relatedness, but for the remaining patients, trio data would be required to definitively orient the skew. This would not require long-read sequencing but could be done simply by PCR genotyping. Compared to transcriptome-based estimates of skewed X inactivation (Werner et al., 2022), sequencing the native genomic DNA allows to orient the skew even for genes that are not expressed in the sampled tissue. Other benefits include a simplified sample collection and preparation, and the identification of all variants rather than exonic variants only. The contribution of X inactivation skew to X-linked inherited retinal disease phenotypes in females remains understudied (Fahim and Daiger, 2016; Fry and MacLaren, 2023). In our cohort of 16 affected females, there was an overrepresentation of skewed X inactivation, and in all severely affected females we found the skew orientation to be deleterious (or in one case predicted to be deleterious with high confidence). By contrast, protective skew in the female relatives of severely affected patients correlated with milder symptoms. Cumulative with previous reports of correlation between skew and *RPGR*-associated retinitis pigmentosa (Fahim et al., 2020), this confirmed the potential for high explanatory power of skewed X inactivation in X-linked inherited retinal disease manifestation in females, and encourages investigating its use in prognostics. As gene therapy solutions for inherited retinal diseases are emerging, including for *RPGR* and *CHM* (Cehajic-Kapetanovic et al., 2020; Martinez-Fernandez de la Camara et al., 2022; Britten-Jones et al., 2022; Xue et al., 2018), early identification of candidate patients who would benefit from this treatment is crucial, in order to rescue expression before irreversible retinal damage. Besides assessing the skew for individual variants, quantifying X inactivation skew can also be used to improve phasing of the X Chromosome. For strongly skewed samples, haplotype blocks overlapping CpG islands can be linked so haplotype 1 is consistently a single parental allele. Analysis of epiallele bias may also be extended to autosomes, for instance in pharmacogenetic panels, if there is evidence of variable expression correlating with differences in DNA methylation. In conclusion, reporting X inactivation skew adds an important additional layer of information for all X-linked diseases or risk variants. As such we expect our method to become

part of the routine analysis pipeline for long-read population genomic datasets.

Methods

Patient sample collection. This multisite study included two cohorts of female carriers of RPGR-associated retinitis pigmentosa and choroideremia. The Australian site received ethics approval from The Royal Victorian Eye and Ear Hospital Human Research Ethics Committee (Ethics ID: 19-1443H). The United Kingdom (UK) site received ethics approval from the Anglia Ruskin University Human Research Ethics Committee (Ethics ID: ETH2223-5360). The wet lab component of this study received ethics approval from the Walter Eliza Hall Institute (Ethics ID: 20/16B). Written informed consent was obtained from each participant prior to commencement of the study, which was undertaken at the respective sites. We used Oragene OG-600 and OCR-100 kits (DNA Genotek) for saliva and buccal mucosa collection, respectively. Patients provided saliva samples by spitting into a tube until the saliva (not foam) passed the fill-line. Buccal swab involved using the cotton tip in the collection kit to rub inside the cheek (no scraping involved). The genetic variants for the participants are outlined in Table 1.

Clinical visits. The clinical component of this study was part of a larger project investigating retinal phenotypes of female carriers of X-linked inherited retinal disorders, which involved participants attending a research visit for a comprehensive ocular examination. Of interest to the current study, participants had fundus autofluorescence imaging performed to classify their retinal severity according to current grading scales for RPGR-associated retinitis pigmentosa (Nanda et al., 2018) and choroideremia (Edwards et al., 2015), and a selection of these images have been utilised in this publica-

tion.

DNA extraction. Blood, saliva and buccal mucosa samples underwent DNA extraction at the Australian Genome Research Facility (AGRF) using a QIAgen DNeasy blood & tissue kit. Retinal DNA was extracted from 3 mm diameter retinal discs with a QIAamp DNA Micro Kit.

Nanopore library preparation. Extracted genomic DNA was quantified with a Qubit and assessed for purity on a NanoDrop and assessed for fragmentation on a TapeStation. Samples that showed excessive fragmentation were size selected in one of two ways: when more than 3 µg of DNA were available, the samples were subjected to semi-selective DNA precipitation (v2, https://community.nanoporetech.com/extraction_methods/size-selection2) based on Jones et al. (2021)); if fewer than 3 µg were available, we used a 0.82X High-Molecular Weight bead clean-up (Promega ProNex). 1 µg of DNA for each sample was then sheared to a 10–20 kb target range using Covaris G-tubes (two 30-second spins at 3,000 g) and taken forward for library preparation following the instructions for genomic DNA sequencing with barcoding kit 14 (SQK-NBD114.24, Oxford Nanopore Technologies). Three-to-four barcoded samples were pooled together for each library to be sequenced on one flow cell.

Nanopore sequencing. 160 ng of library were loaded on R10.4.1 PromethION flow cells (FLO-PRO114M), and run to exhaustion on a PromethION P24 running MinKNOW v23.04.6, using adaptive enrichment of the X Chromosome from the T2T-CHM13v2 genome reference.

Ground truth mouse dataset with genetically skewed X inactivation. All mice were maintained and treated in accordance with Walter and Eliza Hall Institute Animal Ethics Committee approved protocols (approval numbers 2018.004,

2020.048 and 2020.050). *Xist*^{ΔA} mice (Royce-Tolland et al., 2010) were obtained from Dr. Graham Kay, Queensland Institute of Medical Research, and kept on a 129/Sv background. Triplicates of F1 129Cast neural stem cells were derived from E14.5 cortical dissections of three embryos resulting from crosses between *Xist*^{ΔA/+} 129 dams and Cast sires, as described in Jansz et al. (2018). High-molecular weight DNA from 2M neural stem cells was extracted with the Monarch HMW Cell kit (New England Biolab) with agitation at 300 rpm. DNA was resuspended in 750 ul EEB buffer for ultra-long nanopore library preparation. Libraries were prepared with the ultra-long kit (SQK-ULK001) using all the DNA from the extraction. Each sample was run on one R9.4.1 PromethION flow cell (FLO-PRO002) with two intermediate washes and reloads. Adaptive sampling was not performed for this dataset. The triplicates were pooled for the skew analysis.

Adaptive sampling calibration. We start by deriving an estimate of enrichment efficiency as a factor of read lengths to inform library preparation strategy. When enriching for a whole chromosome or in cases where the target length is much larger than read lengths ($L_{\text{target}} \gg L_{\text{read}}$), on-target output is maximised by having the longest read lengths possible. As reads are sampled at random, without adaptive sampling the proportion of the time p that a pore spends acquiring data is simply proportional to the target space ($p \propto L_{\text{target}}/L_{\text{genome}}$) and is independent of read lengths if the wait time between reads t_{wait} is much smaller than the time it takes to sequence a read t_{seq} ($t_{\text{wait}} \ll t_{\text{seq}}$). However with adaptive sampling, on average the pore has to sample $n_{\text{off}} = L_{\text{genome}}/L_{\text{target}}$ off-target reads before finding a read that is on-target. So the proportion of time spent acquiring

on-target data is reduced to

$$p = \frac{t_{\text{seq}}}{t_{\text{seq}} + t_{\text{search}}}$$

where t_{search} is the time spent trying and discarding off-target reads, composed of n_{off} repeats of $t_{\text{reject}} + t_{\text{wait}}$, with t_{reject} the time it takes to sample and reject an off-target read. By denoting the sequencing speed (400 b/s) as s , we have $t_{\text{seq}} = \frac{L_{\text{read}}}{s}$, and thus

$$p = \frac{1}{1 + \frac{sL_{\text{genome}}(t_{\text{reject}} + t_{\text{wait}})}{L_{\text{read}}L_{\text{target}}}}$$

In the case of a human genome (3.1 Gb) and targeting the entire X Chromosome (154 Mb), taking $t_{\text{reject}} = 2.3$ s based on rejected reads of length 900 b and assuming $t_{\text{wait}} = 0.5$ s, with 1 kb reads the proportion of time spent acquiring on-target data is only $p = 4.2\%$ while with 10 kb reads it increases to 31%.

Skew analysis pipeline. Reads from POD5 files were basecalled and mapped to the T2T-CHM13v2 human genome reference using dorado v0.3.2 (<https://github.com/nanoporetech/dorado>) and the dna_r10.4.1_e8.2_400bps_sup@v4.2.0 model (or dna_r10.4.1_e8.2_400bps_sup@v4.0.0 for the first two flow cells that were sequenced with 4 kHz sampling). Dorado uses minimap2 (Li, 2018). Modified basecalling was performed concurrently for 5mCG_5hmCG using the dorado argument `--mod_bases 5mCG_5hmCG` (v2 model). mosdepth v0.3.3 (Pedersen and Quinlan, 2018) and NanoComp v1.21.0 (De Coster and Rademakers, 2023) were used for coverage calculations and quality control. Modbam files from each patient's samples were merged to call SNVs with DeepVariant v1.5.0-gpu (Poplin et al., 2018) on the X Chromosome using the ONT_R104 model, and filtered for high-quality 'PASS' calls using BCFtools

(Li, 2011). The variant call file (VCF) was then phased using WhatsHap v1.7 (Martin et al., 2016) using the merged modbam file. Each sample modbam (saliva, buccal, and blood and retina if available) was then processed individually by NanoMethViz v2.7.8 (Su et al., 2021) to cluster reads by their CpG island methylation pattern, using the HDBSCAN algorithm (Hahsler et al., 2019) and setting the minimum number of reads to form a cluster as 5. The UCSC CHM13v2 (or mm10 for mouse) CpG island annotation was used. CpG islands for which there were exactly two clusters were used to assign Xa and Xi labels to the reads of each cluster (Xa for the low methylation cluster, Xi for the high methylation cluster, reversing these labels in the case of the *XIST/Xist* promoter CpG island). Within each haplotype block, the number of haplotype 1/Xa, haplotype 2/Xi, haplotype 1/Xi and haplotype 2/Xa reads were counted to estimate the block-wise skew, and to estimate the sample skew as the value that maximises the joint probability of observing those counts over all blocks.

Given this sample skew, log-odds of haplotype 1 being the predominantly active vs inactive X for a given haplotype block is given by $\text{logodds} = (2 \times \text{successes} - \text{trials}) \times \log_{10}(p) + (\text{trials} - 2 \times \text{successes}) \times \log_{10}(1 - p)$.

We used R packages *plyranges* v1.20.0 (Lee et al., 2019) for genomic interval manipulation and *cowplot* v1.1.1 for visualisation.

The pipeline is made available as a nextflow pipeline: <https://github.com/QGouil/SkewX> and is archived as Supplementary Material Supplemental code.

Data access

The mouse neural stem cells ultra-long nanopore gDNA sequencing data have been submitted to the EBI ArrayExpress database ([https://www.ebi.ac.uk/biostudies/arrayexpress/studies/E-](https://www.ebi.ac.uk/biostudies/arrayexpress/studies/E-MTAB-13706)

MTAB-13706) under accession number E-MTAB-13706. The patients' genomic nanopore sequencing data generated in this study have been submitted to the European Genome-Phenome Archive (EGA, <https://ega-archive.org/studies/EGAS50000000440>) under accession number EGAS50000000440.

Competing interests statement

QG received travel support from Oxford Nanopore Technologies to attend a conference.

Acknowledgements

We thank Kelsey Breslin for the production of mouse neural stem cells; Sarah MacRaild, Stephen Wilcox and Marek Cmero for their assistance running the PromethION; Jafar Jabbari for sharing his custom bead clean-up protocol; Kathleen Zeglinski, Edward Yang and the WEHI Research Computing team for their help with the Nextflow pipeline. The illustrations were created with BioRender. The authors contributed to the paper as follows: study conception and design: SAG, JL, TLE, DGH, MEB, LNA, QG; data collection: SAG, JL, JKJ; analysis and interpretation of results: SAG, JL, SS, MER, MEB, LNA, QG; draft manuscript preparation: SAG, QG; intellectual contribution and text revisions: SAG, JL, JKJ, TLE, DGH, MER, MEB, LNA, QG. All authors reviewed the results and approved the final version of the manuscript.

Funding

This study was supported by a Choroideremia Research Foundation (USA) grant to SAG, LNA, TLE, JKK, and MEB. QG, LNA, MER and MEB are supported by Australian National Health and Medical Research Council (NHMRC) Investigator Grants (GNT2007996,

- GNT1195713, GNT2017257 and GNT1194345, respectively), a University of Melbourne Driving Research Momentum Fellowship awarded to LNA and an Australian Government Research Training Program Scholarship to SAG. The Walter and Eliza Hall Institute and the Centre for Eye Research Australia receive support from the Victorian State Government through its Operational Infrastructure Support Program. Additional support was provided by the Australian Government through the National Collaborative Research Infrastructure Strategy (NCRIS) program and an Australian National Health and Medical Research Council IRIISS grant (9000719).
- ## References
- BIOS consortium, GoNL consortium, Shvetsova E, Sofronova A, Monajemi R, Gagalova K, Draisma HHM, White SJ, Santen GWE, Chuva de Sousa Lopes SM, et al.. 2019. Skewed X-inactivation is common in the general female population. *European Journal of Human Genetics* **27**: 455–465.
- Bittel DC, Theodoro MF, Kibiryeva N, Fischer W, Talebizadeh Z, and Butler MG. 2008. Comparison of X-chromosome inactivation patterns in multiple tissues from human females. *Journal of Medical Genetics* **45**: 309–313.
- Bonetti G, Cozza W, Bernini A, Kaftalli J, Mareso C, Cristofoli F, Medori MC, Colombo L, Martella S, Staurengi G, et al.. 2023. Towards a long-read sequencing approach for the molecular diagnosis of RPGR ORF15 genetic variants. *International Journal of Molecular Sciences* **24**: 16881.
- Britten-Jones AC, Jin R, Gocuk SA, Cichello E, O'Hare F, Hickey DG, Edwards TL, and Ayton LN. 2022. The safety and efficacy of gene therapy treatment for monogenic retinal and optic nerve diseases: A systematic review. *Genetics in Medicine* **24**: 521–534.
- Martinez-Fernandez de la Camara C, Cehajic-Kapetanovic J, and MacLaren RE. 2022. Emerging gene therapy products for RPGR-associated X-linked retinitis pigmentosa. *Expert Opinion on Emerging Drugs* **27**: 431–443.
- Carss KJ, Arno G, Erwood M, Stephens J, Sanchis-Juan A, Hull S, Megy K, Grozeva D, Dewhurst E, Malka S, et al.. 2017. Comprehensive Rare Variant Analysis via Whole-Genome Sequencing to Determine the Molecular Pathology of Inherited Retinal Disease. *The American Journal of Human Genetics* **100**: 75–90.
- Caylor RC. 2023. Nonrandom X Chromosome Inactivation Detection. *Current Protocols* **3**: e748.
- Cehajic-Kapetanovic J, Xue K, Martinez-Fernandez de la Camara C, Nanda A, Davies A, Wood LJ, Salvetti AP, Fischer MD, Aylward JW, Barnard AR, et al.. 2020. Initial results from a first-in-human gene therapy trial on X-linked retinitis pigmentosa caused by mutations in RPGR. *Nature Medicine* **26**: 354–359.
- De Coster W and Rademakers R. 2023. NanoPack2: population-scale evaluation of long-read sequencing data. *Bioinformatics* **39**: btad311.
- Edwards TL, Groppe M, Jolly JK, Downes SM, and MacLaren RE. 2015. Correlation of Retinal Structure and Function in Choroideremia Carriers. *Ophthalmology* **122**: 1274–1276.
- Fadaie Z, Neveling K, Mantere T, Derks R, Haer-Wigman L, den Ouden A, Kwint M, O'Gorman L, Valkenburg D, Hoyng CB, et al.. 2021. Long-read technologies identify a hidden inverted duplication in a family with choroideremia. *Human Genetics and Genomics Advances* **2**: 100046.
- Fahim AT and Daiger SP. 2016. The Role of X-Chromosome

REFERENCES

- Inactivation in Retinal Development and Disease. In *Retinal Degenerative Diseases* (eds. C Bowes Rickman, MM LaVail, RE Anderson, C Grimm, J Hollyfield, and J Ash), volume 854, pp. 325–331. Springer International Publishing, Cham. Series Title: Advances in Experimental Medicine and Biology.
- Fahim AT, Sullivan LS, Bowne SJ, Jones KD, Wheaton DK, Khan NW, Heckenlively JR, Jayasundera KT, Branham KH, Andrews CA, et al.. 2020. X-Chromosome Inactivation Is a Biomarker of Clinical Severity in Female Carriers of RPGR-Associated X-Linked Retinitis Pigmentosa. *Ophthalmology Retina* **4**: 510–520.
- Fry LE and MacLaren RE. 2023. Comment on Di Giosaffatte et al. A Novel Hypothesis on Choroideremia-Manifesting Female Carriers: Could CHM In-Frame Variants Exert a Dominant Negative Effect? A Case Report. *Genes* 2022, 13, 1268. *Genes* **14**: 2160.
- Gart JJ. 1970. A Locally Most Powerful Test for the Symmetric Folded Binomial Distribution. *Biometrics* **26**: 129.
- Gocuk SA, Jolly JK, Edwards TL, and Ayton LN. 2023. Female carriers of X-linked inherited retinal diseases - Genetics, diagnosis, and potential therapies. *Progress in Retinal and Eye Research* **96**: 101190.
- GTEX Consortium, Tukiainen T, Villani AC, Yen A, Rivas MA, Marshall JL, Satija R, Aguirre M, Gauthier L, Fleharty M, et al.. 2017. Landscape of X chromosome inactivation across human tissues. *Nature* **550**: 244–248.
- Hahsler M, Piekenbrock M, and Doran D. 2019. **dbscan** : Fast Density-Based Clustering with R. *Journal of Statistical Software* **91**.
- Hatakeyama C, Anderson C, Beever C, Peñaherrera M, Brown C, and Robinson W. 2004. The dynamics of X-inactivation skewing as women age. *Clinical Genetics* **66**: 327–332.
- Jansz N, Keniry A, Trussart M, Bildsoe H, Beck T, Tonks ID, Mould AW, Hickey P, Breslin K, Iminoff M, et al.. 2018. Smchd1 regulates long-range chromatin interactions on the inactive X chromosome and at Hox clusters. *Nature Structural & Molecular Biology* **25**: 766–777.
- Johansson J, Lidéus S, Höijer I, Ameer A, Gudmundsson S, Annerén G, Bondeson ML, and Wilbe M. 2023. A novel quantitative targeted analysis of X-chromosome inactivation (XCI) using nanopore sequencing. *Scientific Reports* **13**: 12856.
- Jones A, Torkel C, Stanley D, Nasim J, Borevitz J, and Schwessinger B. 2021. High-molecular weight DNA extraction, clean-up and size selection for long-read sequencing. *PLOS ONE* **16**: e0253830.
- Juchniewicz P, Kloska A, Portalska K, Jakóbkiewicz-Banecka J, Węgrzyn G, Liss J, Głodek P, Tukaj S, and Piotrowska E. 2023. X-chromosome inactivation patterns depend on age and tissue but not conception method in humans. *Chromosome Research* **31**: 4.
- Lee S, Cook D, and Lawrence M. 2019. plyranges: a grammar of genomic data transformation. *Genome Biology* **20**: 4.
- Li H. 2011. A statistical framework for SNP calling, mutation discovery, association mapping and population genetic parameter estimation from sequencing data. *Bioinformatics* **27**: 2987–2993.
- Li H. 2018. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**: 3094–3100.
- Lyon MF. 1961. Gene action in the X-chromosome of the mouse (*Mus musculus* L.). *Nature* **190**: 372–373.

- Mantel N. 1970. 293. Note: An Alternative Test for the Sym-
metric Folded Binomial Distribution. *Biometrics* **26**: 848.
- Martin M, Patterson M, Garg S, O Fischer S, Pisanti N, Klau
GW, Schöenhuth A, and Marschall T. WhatsHap: fast and
accurate read-based phasing. *bioRxiv*. 10.1101/085050.
- Mengel-From J, Lindahl-Jacobsen R, Nygaard M, Soerensen
M, Ørstavik KH, Hertz JM, Andersen-Ranberg K, Tan Q,
and Christensen K. 2021. Skewness of X-chromosome in-
activation increases with age and varies across birth co-
horts in elderly Danish women. *Scientific Reports* **11**:
4326.
- Morleo M and Franco B. 2008. Dosage compensation of
the mammalian X chromosome influences the phenotypic
variability of X-linked dominant male-lethal disorders.
Journal of Medical Genetics **45**: 401–408.
- Nakamichi K, Van Gelder RN, Chao JR, and Mustafi D.
2023. Targeted adaptive long-read sequencing for discov-
ery of complex phased variants in inherited retinal disease
patients. *Scientific Reports* **13**: 8535.
- Nanda A, Salvetti AP, Clouston P, Downes SM, and Ma-
cLaren RE. 2018. Exploring the Variable Phenotypes of
RPGR Carrier Females in Assessing their Potential for
Retinal Gene Therapy. *Genes* **9**: 643.
- Niu AI, Wang Yq, Zhang H, Liao Ch, Wang Jk, Zhang R, Che
J, and Su B. 2011. Rapid evolution and copy number vari-
ation of primate RHOXF2, an X-linked homeobox gene
involved in male reproduction and possibly brain function.
BMC Evolutionary Biology **11**: 298.
- Onozawa M, Goldberg L, and Aplan PD. 2015. Landscape of
Insertion Polymorphisms in the Human Genome. *Genome
Biology and Evolution* **7**: 960–968.
- Pedersen BS and Quinlan AR. 2018. Mosdepth: quick cover-
age calculation for genomes and exomes. *Bioinformatics*
34: 867–868.
- Poplin R, Chang PC, Alexander D, Schwartz S, Colthurst T,
Ku A, Newburger D, Dijamco J, Nguyen N, Afshar PT,
et al.. 2018. A universal SNP and small-indel variant caller
using deep neural networks. *Nature Biotechnology* **36**:
983–987.
- Royce-Tolland ME, Andersen AA, Koyfman HR, Talbot DJ,
Wutz A, Tonks ID, Kay GF, and Panning B. 2010. The
A-repeat links ASF/SF2-dependent Xist RNA processing
with random choice during X inactivation. *Nature Struc-
tural & Molecular Biology* **17**: 948–954.
- Sharp A, Robinson D, and Jacobs P. 2000. Age- and tissue-
specific variation of X chromosome inactivation ratios in
normal women. *Human Genetics* **107**: 343–349.
- Su S, Gouil Q, Blewitt ME, Cook D, Hickey PF, and Ritchie
ME. 2021. NanoMethViz: an R/Bioconductor package for
visualizing long-read methylation data. *PLoS Computa-
tional Biology* **17**: e1009524.
- Tribioli C, Tamanini F, Patrosso C, Milanesi L, Villa A, Per-
golizzi R, Maestrini E, Rivella S, Bione S, Mancini M,
et al.. 1992. Methylation and sequence analysis around
Eagi sites: identification of 28 new CpG islands in XQ24-
XQ28. *Nucleic Acids Research* **20**: 727–733.
- Vaché C, Faugère V, Baux D, Mansard L, Van Goethem C,
Dhaenens CM, Grunewald O, Audo I, Zeitz C, Meunier I,
et al.. 2024. Validation of Nanopore long-read sequencing
to resolve RPGR ORF15 genotypes in individuals with X-
linked retinitis pigmentosa. *European Journal of Human
Genetics* .
- Van den Veyver IB. 2001. Skewed X Inactivation in X-Linked

968 Disorders. *Seminars in Reproductive Medicine* **19**: 183–
 969 192.

970 Werner JM, Ballouz S, Hover J, and Gillis J. 2022. Vari-
 971 ability of cross-tissue X-chromosome inactivation charac-
 972 terizes timing of human embryonic lineage specification
 973 events. *Developmental Cell* **57**: 1995–2008.e5.

974 Xue K, Jolly JK, Barnard AR, Rudenko A, Salvetti AP, Patrí-
 975 cio MI, Edwards TL, Groppe M, Orlans HO, Tolmachova
 976 T, et al.. 2018. Beneficial effects on vision in patients un-
 977 dergoing retinal gene therapy for choroideremia. *Nature*
 978 *Medicine* **24**: 1507–1512.

979 Yahya S, Watson CM, Carr I, McKibbin M, Crinnion LA,
 980 Taylor M, Bonin H, Fletcher T, El-Asrag ME, Ali M, et al..
 981 2023. Long-Read Nanopore Sequencing of RPGR ORF15
 982 is Enhanced Following DNase I Treatment of MinION
 983 Flow Cells. *Molecular Diagnosis & Therapy* **27**: 525–535.