scientific reports

OPEN

A novel quantitative targeted analysis of X-chromosome inactivation (XCI) using nanopore sequencing

Josefin Johansson[1], Sarah Lidéus[1], Ida Höijer[1], Adam Ameur[1], Sanna Gudmundsson[2,3], Göran Annerén[1], Marie-Louise Bondeson[1] & Maria Wilbe[1]✉

X-chromosome inactivation (XCI) analyses often assist in diagnostics of X-linked traits, however accurate assessment remains challenging with current methods. We developed a novel strategy using amplification-free Cas9 enrichment and Oxford nanopore technologies sequencing called XCI-ONT, to investigate and rigorously quantify XCI in human androgen receptor gene (*AR*) and human X-linked retinitis pigmentosa 2 gene (*RP2*). XCI-ONT measures methylation over 116 CpGs in *AR* and 58 CpGs in *RP2*, and separate parental X-chromosomes without PCR bias. We show the usefulness of the XCI-ONT strategy over the PCR-based golden standard XCI technique that only investigates one or two CpGs per gene. The results highlight the limitations of using the golden standard technique when the XCI pattern is partially skewed and the advantages of XCI-ONT to rigorously quantify XCI. This study provides a universal XCI-method on DNA, which is highly valuable in clinical and research framework of X-linked traits.

X-chromosome inactivation (XCI) is a compensation mechanism for the difference in number of X-chromosomes between males (46XY), females (46XX) and individuals with X-aneuploidies[1]. The human molecular XCI mechanism is not completely understood[2,3] but studies have shown that XCI is controlled by cis-acting and trans-acting factors in the X-inactivation center (Xic) including gene expression of the X-inactive specific transcript (*XIST*) and X-active specific transcript (*XACT*) genes[3]. This eventually leads to monoallelic expression, and accumulation of H3K27me3[3] and methylation of CpG sites near promoters of genes on the inactive X-chromosome (Xi) (Fig. 1A)[4–6]. Methylation has shown to be important for the maintenance of Xi's inactive state[7–10]. XCI in humans is initiated in the early embryonic implantation stage[3], and the choice of which X-chromosome to be inactivated is in general described as a random process, where both X-chromosomes are equally represented[11,12]. However, preferential inactivation of one X-chromosome, so-called skewed XCI, has been observed to modify X-linked disease manifestation in carrier females, indicating that the genotype is important in the choice of active X-chromosome (Xa), possibly by cell selection due to a mutant cell disadvantage[13–16]. Firstly, preferential silencing of the pathogenic allele has been associated with a selective female survival or a less severe effect of X-linked traits. Extreme skewing has been shown to be a good indicator for the presence of pathogenic X-linked variants in carrier females[15,17,18] and XCI analyses in family relatives therefore often assist in the interpretation of X-linked variants. Secondly, expression of the pathogenic allele may lead to manifestation of phenotypes in carrier females of X-linked traits[19–28] and can explain phenotypic differences observed in affected carrier females and individuals with X-aneuploidies[21,24,29–36]. Of note, XCI analyses are recommended to be performed in relevant tissues, at different ages and in non-smokers[37–44]. In general, the definition of skewing has been > 80:20 but due to limitations in the golden standard technique used for XCI analysis, quantification of the skewedness has not been recommended for other than 100:0 silencing, highlighting the need of a quantitative method.

The golden standard XCI analysis uses Methylation Sensitive Restriction Enzymes (MSREs), PCR and Fragment Length Analysis (FLA) targeting two regions on the X-chromosome; the first exon of the human androgen receptor (*AR*) gene[45] and the promoter region of the human X-linked retinitis pigmentosa 2 (*RP2*) gene[46]. Methylation of CpG sites in these regions are associated with XCI in humans[46–48] and by using different MSREs with

[1]Department of Immunology, Genetics and Pathology, Science for Life Laboratory, Uppsala University, Husargatan 3, Box 815, SE-751 08 Uppsala, Sweden. [2]Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA, USA. [3]Division of Genetics and Genomics, Boston Children's Hospital, Harvard Medical School, Boston, MA, USA. ✉email: maria.wilbe@igp.uu.se
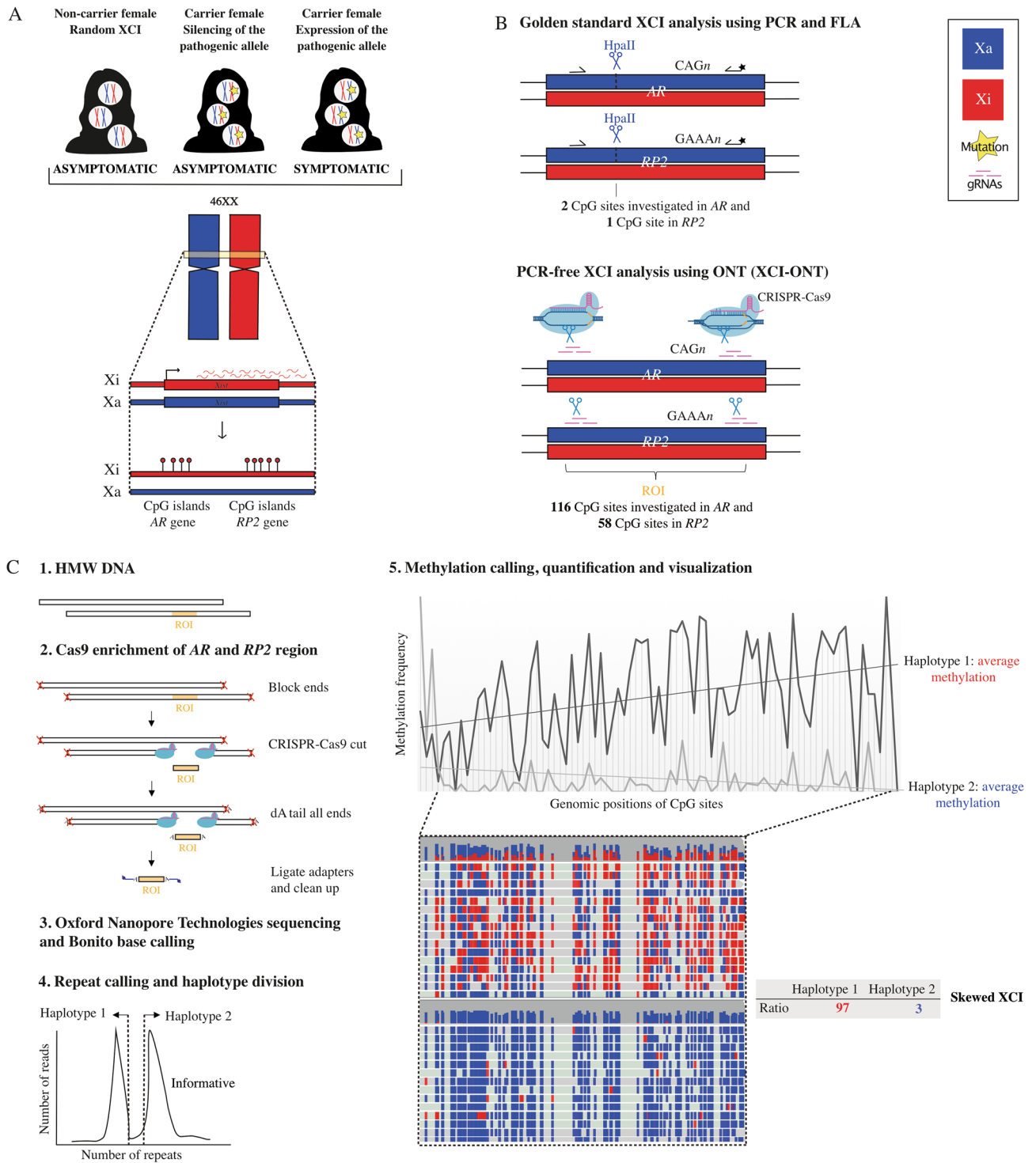
Figure 1. (**A**) Schematic of the X-chromosome inactivation mechanism and its effect in carrier females of X-linked traits. Blue: active X-chromosome (Xa), red: inactive X-chromosome (Xi), yellow star: pathogenic X-linked variant. Illustrations created by using adobe illustrator 2022 (available at https://adobe.com/products/illustrator). (**B**) The golden standard XCI methods uses methylation sensitive restriction enzymes (HpaII) that cuts the Xa but leaves the Xi intact. HpaII targets two and one CpG site in the Androgen receptor (*AR*) and Retinis pigmentosa 2 (*RP2*) respectively, and is followed by PCR and fragment length analysis (FLA) spanning polymorphic repetitive regions (CAG$_n$ or GAAA$_n$) that separates the parental alleles. Illustrations created by using Adobe Illustrator 2022 (available at https://adobe.com/products/illustrator). The XCI-ONT approach cuts the DNA independent of methylation status using CRISPR-Cas9 enrichment with three gRNAs (pink) flanking a ~ 3 kb region of interest (ROI) spanning 116 CpG sites in *AR* and 58 CpG sites in *RP2*. (**C**) XCI-ONT includes: (**1**) Dephosphorylation of 5′ ends to reduce ligation of sequencing adapters to off-target strands. (**2**) CRISPR-Cas9 system bind and cuts the ROI, and the DNA is dA-tailed for adapter ligation to Cas9 cut sides, which are both 3′ dA-tailed and 5′ phosphorylated. (**3**) The library is sequenced using Oxford Nanopore Technologies and Bonito base calling. (**4**) Calling repeats by aligning the reads to reference genomes containing all possible repeats in the regions, and the reads are divided into haplotypes by plotting the number of reads with different repeats. Lastly, methylation calling and quantification are performed using Nanopolish, and the data is visualized using Integrative Genomics Viewer (IGV). The average methylation of the ROI is calculated and the XCI ratio between the two X-chromosomes is determined.

various target sequences, one or a few CpG sites in these genes are often investigated using the golden standard analysis. Methylation of CpGs in *AR* and *RP2* are of specific interest in XCI analyses due to their proximity to highly conserved polymorphic repetitive elements (CAG$_n$ in *AR* and GAAA$_n$ in *RP2*, heterozygosity rate 0.97 combined) that can be used for separating the parental alleles using FLA (Fig. 1B). However, the golden standard analysis is not robust since it relies on MSREs for digestion, is PCR-based and semi-quantitative. The results are often difficult to interpret due to PCR stutter peaks, secondary structures and/or polymorphisms in the fragment affecting the fragment size and thereby biasing the separation of alleles. In addition, PCR and FLA are not suitable for quantifying imbalanced expression of the two X-chromosomes often leaving the smaller fragment in favor[49].

Quantification is necessary when the level of skewedness is < 100:0, and in the grey zone of the skewing definition (> 80:20), often used in the interpretation of X-linked variants. In addition, quantification can be very useful when XCI is modifying disease e.g. in the investigation of carrier females presenting symptoms due to expression of the pathogenic allele (incomplete silencing). Approaches for quantifying the levels of XCI on the maternal and paternal allele have been proposed before but this requires either bisulfite conversion and PCR[50] or paired RNA and DNA sequencing data of the *XIST* gene using informative transcribed heterozygous single nucleotide variants (SNVs) to separate the parental alleles[41,51]. A SNV is not as informative as repeated elements and can therefore not be used to distinguish between individuals. This highlights the need for a universal quantitative XCI analysis on DNA to be used for research applications and clinical diagnostics related to X-linked traits.

Oxford nanopore technologies (ONT) sequencing can be used to detect CpG methylation revealed by changes in the raw electrical signal during the sequencing process. Long reads are powerful to investigate methylation over repetitive regions and GC-rich sequences such as CpG islands. In addition, high coverage without amplification is a very useful quantitative measurement of methylation. However, high read depth of the whole genome is expensive and therefore an enrichment approach could be advantageous. Recently, an amplification-free targeted library preparation prior to long-read sequencing was released using the CRISPR-Cas9 system to effectively sequence complicated targeted regions with high-coverage and no PCR bias[52–56]. Analysis without PCR steps are important to avoid allelic dropout and allow quantitative measurement of methylation. Here, we demonstrate a novel strategy to quantify XCI using CRISPR-Cas9 enrichment of the *AR* and *RP2* regions along with ONT sequencing for repeat and methylation detection. The novel strategy is referred to as XCI-ONT in this paper and is compared to the golden standard XCI method (Fig. 1B,C).

Results

XCI investigation using golden standard method of *AR* and *RP2*.
In this study, XCI was first investigated using the golden standard PCR-based strategy of the *AR* and *RP2* genes. The investigated family was previously described with X-linked intellectual disability (OMIM #300966) caused by a *TAF1* variant (c.3568C > T;p. (Arg1190Cys), NM_004606.4) where carrier females had ~ 100:0 skewed XCI when investigating the *AR* gene[57]. The XCI result was confirmed in this study by investigating the *RP2* locus of two asymptomatic carrier females (IV:8, III:10) and one asymptomatic non-carrier female (III:7) (Supplementary Fig. 1). The two asymptomatic carrier females presented a ~ 100:0 skewed XCI status consistent with the maternal X-chromosome being silent (carrying the pathogenic variant). The asymptomatic non-carrier female had expression of both alleles i.e. a random XCI status, however the method presents variable result between experiments and contrasting results in the different genes (63:37 in *AR* vs. 47:53 in *RP2*). In addition, three females (female I, II and III) were investigated using the golden standard method. All individuals showed different ratios of XCI for both *AR* (female I 62:38, female II 47:53 and female III 81:19) and *RP2* (female I 81:19, female II 60:40 and female III 87:13) (Table 1, Supplementary Fig. 2).

XCI investigation using XCI-ONT of *AR* and *RP2*.
To address the limitations of the golden standard technique and quantify the ratio of active maternal or paternal X-chromosomes, we developed XCI-ONT and applied it on the same samples (Fig. 1B,C). XCI-ONT uses the Cas9-enrichment protocol[53] with three gRNAs on both sides of a ~ 3 kb region spanning the same repeats and CpG sites as the golden standard method of the *AR* and *RP2* genes (Supplementary Table 1). The region enrichments were followed by DNA sequencing and simultaneous methylation detection using ONT sequencing. This enables direct detection of repeats without PCR-bias and methylation detection of 116 CpGs in *AR* (chrX: 67543761–67546170, hg38) and 58 CpGs in *RP2* (chrX:

Sample	OMIM	Female disease status	XCI status AR (FLA*/ONT)	XCI status RP2 (FLA*/ONT)
IV:8	#300966	Carrier	100:0/95:5	100:0/91:9
III:10	#300966	Carrier	100:0/97:3	100:0/91:9
III:7	#300966	Non-carrier	63:37/31:69	47:53/34:66
Female I.I	NA	NA	62:38/28:72	81:19/26:74
Female I.II	NA	NA	NA/27:73	NA/27:73
Female II	NA	NA	47:53/54:46	60:40/67:33
Female III	NA	NA	81:19/81:19	87:13/92:8

Table 1. XCI status of all samples in the *AR* and *RP2* gene. *FLA* fragment length analysis/golden standard analysis, *ONT* Oxford nanopore technologies. *FLA is a semi-quantitative method and measuring peak heights is not recommended for other than extreme skewedness (> 90:10).

46836539–46837273, hg38), in comparison to the golden standard technique investigating only one or two CpGs per gene[45,46]. The XCI-ONT investigation generated between 43 and 155 reads on target of which 32–105 were used to calculate the methylation frequencies to separate the haplotypes (Supplementary Table 2). The repeats were detected by alignment to reference genome containing a range of naturally occurring repeats lengths in each gene respectively. The number of repeats in the novel XCI-ONT assay in *AR* and *RP2* was consistent with the base pair differences detected using the golden standard assay (Supplementary Figs. 1, 2 and 3). Methylation calling and calculation of the methylation frequency were performed using Nanopolish[58]. The methylation status was visualized, and the XCI status was calculated using the average methylation frequency in said regions followed by calculating the ratio of the average methylation between the alleles in each gene. The two asymptomatic carrier females (IV:8 and III:10) presented a skewed XCI status using XCI-ONT with a methylation ration of 95:5 (IV:8) and 97:3 (III:10) in *AR*, and 91:9 (IV:8) and 91:9 (III:10) in *RP2*. The asymptomatic non-carrier female (III:7) presented a random XCI status with a methylation ratio of 31:69 in *AR*, and 34:66 in *RP2* (Table 1, Fig. 2 and Supplementary Table 3). Females I and II displayed a random XCI status, and female III presented a random XCI status for *AR* but a skewed XCI status for *RP2*. Female I was analyzed two times (female I.I and female I.II) with a new DNA sample to investigate the robustness of the method. The ratios were as following; 28:72 (female I.I), 27:73 (female I.II), 54:46 (female II), and 81:19 (female III) in *AR* and 26:74, (female I.I) 27:73 (female I.II), 67:33 (female II), and 92:8 (female III) in *RP2*. The results from female I was highly consistent between the two separate runs (28:72 vs. 27:73), demonstration the robustness and specificity for XCI-ONT (Table 1, Supplementary Fig. 4 and Supplementary Table 3). The variability of methylation calls between the two haplotypes across
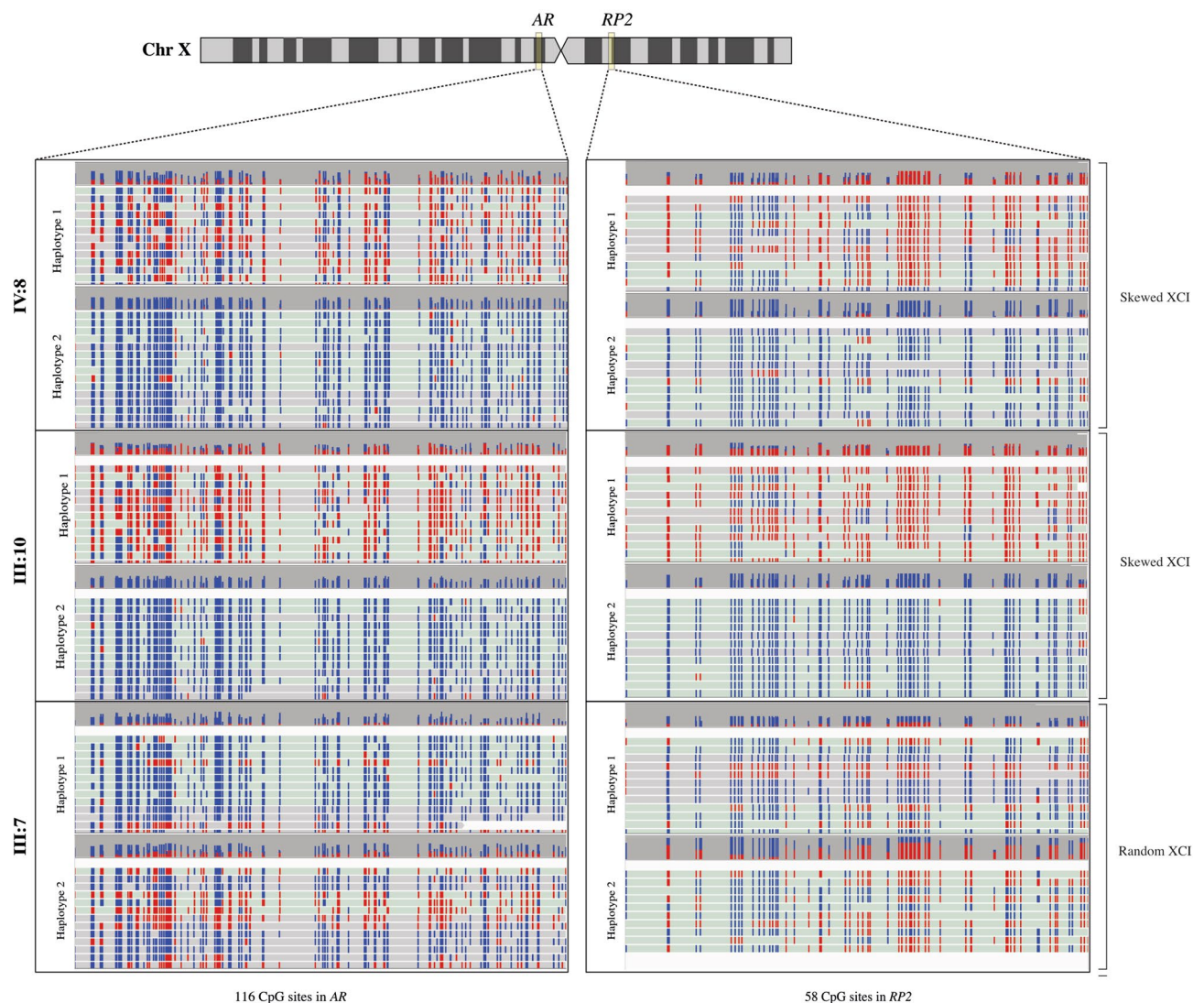

Figure 2. Visualization of XCI-ONT result using Integrative Genomics Viewer (IGV) presenting methylated CpG sites (red; Xi) and unmetylated CpG sites (blue; Xa) across the reads in the Androgen receptor (*AR*) and Retinis pigmentosa 2 (*RP2*) genes. Asymptomatic carrier females (IV:8, III:10) present a skewed XCI pattern and asymptomatic non-carrier females (III:7) of the same family have a random XCI status. Top bar in each haplotype visualization indicates the read coverage (height) and the percentage of methylated and unmethylated calls at each position.

4

all reads and analyzed CpG sites were investigated and visualized. A distinct pattern for the majority of CpG sites between methylated or unmethylated haplotypes can be seen for IV:8 and III:10 for both *AR* and *RP2* (Supplementary Fig. 5A–D). For females I and II the different haplotypes cannot be distinguished using methylation status (Supplementary Fig. 5E–L). Female III exhibits no distinction between the haplotypes for *AR* but shows a clear division between the haplotypes for *RP2* (Supplementary Fig. 5M,N).

Discussion

We demonstrate a combination of CRISPR-Cas9 enrichment and ONT sequencing (XCI-ONT) as a novel successful strategy to quantify XCI in females, which can be used to investigate X-linked traits. Here, XCI was investigated in a family with X-linked intellectual disability[57], using the golden standard PCR-based strategy of the *AR* and *RP2* genes, and then compared to the XCI-ONT result of the same genes. The golden standard XCI investigation of carrier females (IV:8 and III:10) revealed a ~ 100:0 skewed XCI status in the *AR* and *RP2*, resulting in silencing the pathogenic allele as a protective mechanism to disease. The golden standard XCI investigation of an asymptomatic non-carrier female of the same family (III:7) as well as three additional individuals (female I-III) expressed both alleles (random XCI), however the golden standard method presented variable result between experiments and contrasting results in the different genes (63:37 in *AR*, 47:53 in *RP2* for III:7 and 62:38 in *AR*, 81:19 in *RP2* for female I). The results highlight the limitations of using the golden standard method when the XCI pattern is only partially skewed[45,46].

To address this, and quantify the ratio of active maternal or paternal X-chromosomes, we developed XCI-ONT, and applied it on the same samples as in the golden standard analysis. The XCI-ONT method detects 116 CpGs in the *AR* gene and 58 CpGs in *RP2* gene, compared to only one or two CpGs per gene using the golden standard method[45,46]. The two asymptomatic carrier females presented a skewed XCI status with 95:5 (IV:8) or 97:3 (III:10) in the *AR* gene, and 91:9 in the *RP2* gene (IV:8 and III:10), which is consistent with the golden standard results, affirming the established XCI-ONT protocol presented in this paper. The asymptomatic non-carrier female (III:7) and female I presented a random XCI status using XCI-ONT with a methylation ratio of 31:69 and 28:72 in *AR* and 34:66 and 26:74 in *RP2* respectively, implying a distinctly different and robust detection using XCI-ONT compared to the golden standard investigation that presented contrasting results between genes and experiments (63:37 and 62:38 in *AR*, 47:53 and 81:19 in *RP2*). These results display a robust XCI detection of different ratios in both genes and confirm that the quantitative XCI-ONT investigation is preferred when XCI pattern is partially skewed. Despite the small sample size of this study, the results show a more accurate XCI investigation with great quantitative potential using the novel XCI-ONT strategy compared with the semi-quantitative golden standard method. We also show that reanalyzing an individual using a new DNA sample yields the same result demonstrating the robustness of the method.

To distinguish the alleles, both the golden standard and XCI-ONT strategy uses the advantage of highly conserved polymorphic repetitive elements in *AR* and *RP2*. Repeats are useful for separating chromosomes since parental alleles often differ in number of repeat units. The golden standard investigation uses FLA for an indirect detection of the repeats and separates the alleles by fragment size. Consequently, the haplotype division can be biased by indels and fluorophores or secondary structures affecting the detected fragment length. ONT has the power of direct detection of repeats, underlining the power of using XCI-ONT for haplotype division. This conclusion was supported because the difference in repeat lengths between the two haplotypes was consistent when comparing the two methods, except in female I where the second haplotype show 22 repeats by FLA and 39 repeats using XCI-ONT. This is most likely due to the fact that FLA is a PCR based method which cannot handle the large number of repeats.

To investigate XCI skewedness, both the golden standard and XCI-ONT strategy uses the methylation level of the different haplotypes in *AR* and *RP2*. However, the golden standard method has limitations since it relies on MSREs for digestion and only investigate methylation of one or a few CpG sites in the region. In comparison, the XCI-ONT strategy investigates methylation over 116 CpG sites in *AR* and 58 CpG sites in *RP2* using ONT sequencing. To consider the methylation calling error frequency, only reads with steady methylation calls have been included in this analysis (see "Methods" section). Thus, some CpG sites may end up with no methylation status (Supplementary Fig. 5), and therefore the XCI-ONT status is reported as a ratio of the average methylation between the chromosomes. Future studies that include more samples can reveal which sites that are the most reliable to use in analysis and therefore fine-tune this method and improve accuracy even further. In theory, confirmed by the results in this study, XCI-ONT results in a more accurate XCI measurement in comparison to golden standard methods, especially with partially skewed females, indicating the power of using ONT and more than one CpG site in XCI analyses. This is confirmed by the variability of methylation seen across the CpG sites in the two haplotypes of all investigated individuals (Supplementary Fig. 5).

In this study, both the *AR* and *RP2* gene have been used to confirm the XCI status. The methylation ratio of the alleles may differ slightly in the genes due to poor haplotype division caused by ≤ 2 repeats in difference in *RP2*, more CpGs in the investigated *AR* region or methylation calling errors. XCI-ONT holds the potential of increasing the number of gRNAs targeting additional X-linked genes, improving the accuracy further. In addition, to date, a limitation in the methylation calling only allows methylation detection within 10 base pair windows and therefore XCI-ONT cannot be compared to the golden standard method at a CpG site level. However, although this is not a possibility with Nanopolish yet, this research is constantly evolving and is expected to be possible soon.

In parallel to our study, ONT sequencing was demonstrated as a promising tool to investigate XCI across the X-chromosome[6], but ONT sequencing as a targeted approach to separate the haplotypes and quantify XCI has to our knowledge not been reported before. Moreover, the performance of ONT sequencing compared to golden standard XCI methods have not been demonstrated before.

To summarize, we demonstrate for the first time, XCI-ONT as a novel successful strategy to quantify XCI in carrier females of X-linked traits. This is useful in clinical work and research when the patient is in the grey zone of the skewing definition (> 80:20), often assisting in the interpretation of X-linked variants. In addition, quantification can be highly useful in the investigation of carrier females manifesting symptoms due to expression of the pathogenic allele. Quantification of XCI can illuminate the mechanism of disease, and provide useful information for improving prenatal risk assessment[36], diagnostics and treatment development of X-linked traits. Lastly, we would like to illuminate the possibility that different disorders and pathogenic variants requires different levels of expression to develop symptoms and stress the use of a quantitative XCI investigation to answer these questions. A related application for XCI-ONT is monitoring pharmacological Xi reactivation, which has been suggested as treatment for X-linked disorders due to skewed XCI[59–62].

Methods

Patient samples and ethical considerations.
Ethical approval was received from the local ethics committee of the Swedish Ethical Review Authority for human research in Uppsala, Sweden (Dnr 2012/321). The samples were recruited at the Department of Clinical genetics, Uppsala University hospital and informed consent was received from all participants. The study was conducted according to the guidelines of the Declaration of Helsinki. HMW genomic DNA were extracted from 200 µl blood of six females using Nanobind CBB Big DNA kit (Circulomics) according to standard procedures or using the salting-out method (standard protocol available upon request) followed by clean-up with Nanobind CBB Big DNA kit (Circulomics). The family investigated in this study was previously described with X-linked intellectual disability (OMIM #300966) caused by a c.3568C > T;p.(Arg1190Cys) variant in the *TAF1* gene (NM_004606.4)[57]. DNA from two carrier females (III:10, IV:8) and one non-carrier female (III:7) as well as three unrelated females (female I-III) was included in the current analysis.

XCI investigation using golden standard methods of the *AR* and *RP2* genes.
Two hundred ng of genomic DNA were cut using methylation sensitive HpaII FastDigest in a total volume of 20 µl following manufacturer's instructions (Thermo Fisher Scientific). PCR spanning the *AR* gene was performed using Long-range PCR kit (Qiagen) with input of 50 ng DNA or 2 µl digested DNA with primers and cycling conditions described before[45]. The *RP2* gene was amplified in a 20 µl reaction containing 1X PCR reaction buffer, 250 nM dNTPs, 1uM of each primer, 1U Taq polymerase, 50 ng DNA or 2 µl digested DNA with primers and thermal cycler conditions described before[46]. Genotyping was performed using FLA on the 3130xl ABI Genetic Analyzer with ROX500 Size Standard (Thermo Fisher Scientific). The Amplified Fragment Length Polymorphism (AFLP) was determined using Geneious Prime version 2022.01 with microsatellite plugin (https://www.geneious.com). The peak heights were measured using ImageJ software[63] followed by calculation of the height ratio. DNA from a female sample with known XCI pattern (100:0) were used as a control throughout the analysis pipeline.

XCI investigation using XCI-ONT of the *AR* and *RP2* genes.
Three CRISPR RNAs (crRNAs) on both sides of each region of interest in the *AR* and *RP2* genes were designed using CHOPCHOP[64] and selected according to the previously described instructions (Cas-mediated PCR-free enrichment protocol version: ENR_9084_v109_revD_04Dec2018). All custom designed Alt-R crRNAs (Integrated DNA Technologies, Supplementary Table 2) were pooled as an equimolar mix of crRNAs (100 µM) and assembled with trans-activating crRNAs (tracrRNAs) (Integrated DNA Technologies, cat 1073190) using Duplex buffer (Integrated DNA Technologies, cat 11010301) according to manufacturer's instructions. Cas9 enrichment was performed according to manufacturer's instructions using on 4–5 µg DNA and Long fragment buffer (Cas-mediated PCR-free enrichment protocol SQK-CS9109). Samples were run on a MinION R9.4.1 flow cell and operated using the MinKNOW software (version 20.10.3).

Methylation analysis of Oxford nanopore data.
Base calling was performed using Bonito basecaller (version 0.3.8) and reads were aligned to the human reference genome (GRCh38/hg38) using minimap2 (version 2.18-r1015). Only reads targeting the regions of interest in the *AR* (chrX:67543761–67546170, hg38) and *RP2* gene (chrX:46836539–46837273, hg38) were considered for further analysis. To enable repeat-calling, an in-house database was created with reference genomes in containing 5–40 CAG repeats (*AR* gene) and 5–30 GAAA repeats (*RP2* gene). The range of repeat numbers used in the alignment has previously been described in the unaffected general population (ranging from 9 to 38 in *AR* and 10–20 in *RP2*)[46,65]. Repeat calling was performed by alignment to the in-house database using minimap2 (version 2.18-r1015) and number of repeats in the reads were identified using Integrative Genomics Viewer (version 2.10.0, base quality > 20). The data were visualized by plotting the number of reads (y-axis) containing any repeat in the range of 5–40 repeats in *AR* or 5–30 repeats in *RP2* (x-axis) (Supplementary Fig. 3). The reads were divided into haplotypes based on the repeat count and comparison with the golden standard results. Methylation calling and calculation of the methylation frequency was performed using Nanopolish software package (version 0.12.0). Only sites with a log-likelihood ratio > 2.5 (methylated) or < − 2.5 (unmethylated) were included in the methylation analysis (Supplementary Fig. 5). Bam-files were converted using the "converting bam for igv" package[66] and the methylation status was visualized using Integrative Genomics Viewer bisulfite mode CG (version 2.10.0). The XCI status was established by using the average methylation frequency in the chrX:67543761–67546170 region (*AR* gene, 116 CpG sites, hg38) and chrX:46836539–46837273 region (*RP2* gene, 58 CpG sites, hg38) followed by calculating the ratio of the average methylation between the alleles in each gene.

Data availability

References

1. Lyon, M. F. Gene action in the X-chromosome of the mouse (*Mus musculus* L.). *Nature* **190**, 372–373. https://doi.org/10.1038/190372a0 (1961).
2. Loda, A., Collombet, S. & Heard, E. Gene regulation in time and space during X-chromosome inactivation. *Nat. Rev. Mol. Cell Biol.* https://doi.org/10.1038/s41580-021-00438-7 (2022).
3. Patrat, C., Ouimette, J. F. & Rougeulle, C. X chromosome inactivation in human development. *Development* https://doi.org/10.1242/dev.183095 (2020).
4. Sharp, A. J. *et al.* DNA methylation profiles of human active and inactive X chromosomes. *Genome Res.* **21**, 1592–1600. https://doi.org/10.1101/gr.112680.110 (2011).
5. Yasukochi, Y. *et al.* X chromosome-wide analyses of genomic DNA methylation states and gene expression in male and female neutrophils. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 3704–3709. https://doi.org/10.1073/pnas.0914812107 (2010).
6. Gershman, A. *et al.* Epigenetic patterns in a complete human genome. *Science* **376**, eabj5089. https://doi.org/10.1126/science.abj5089 (2022).
7. Venolia, L. *et al.* Transformation with DNA from 5-azacytidine-reactivated X chromosomes. *Proc. Natl. Acad. Sci. U. S. A.* **79**, 2352–2354. https://doi.org/10.1073/pnas.79.7.2352 (1982).
8. Mohandas, T., Sparkes, R. S. & Shapiro, L. J. Reactivation of an inactive human X chromosome: Evidence for X inactivation by DNA methylation. *Science* **211**, 393–396. https://doi.org/10.1126/science.6164095 (1981).
9. Lock, L. F., Takagi, N. & Martin, G. R. Methylation of the Hprt gene on the inactive X occurs after chromosome inactivation. *Cell* **48**, 39–46. https://doi.org/10.1016/0092-8674(87)90353-9 (1987).
10. Csankovszki, G., Nagy, A. & Jaenisch, R. Synergism of Xist RNA, DNA methylation, and histone hypoacetylation in maintaining X chromosome inactivation. *J. Cell Biol.* **153**, 773–784. https://doi.org/10.1083/jcb.153.4.773 (2001).
11. Lyon, M. F. X-chromosome inactivation and developmental patterns in mammals. *Biol. Rev. Camb. Philos. Soc.* **47**, 1–35. https://doi.org/10.1111/j.1469-185x.1972.tb00969.x (1972).
12. Monkhorst, K., Jonkers, I., Rentmeester, E., Grosveld, F. & Gribnau, J. X inactivation counting and choice is a stochastic process: Evidence for involvement of an X-linked activator. *Cell* **132**, 410–421. https://doi.org/10.1016/j.cell.2007.12.036 (2008).
13. Migeon, B. R. Non-random X chromosome inactivation in mammalian cells. *Cytogenet. Cell Genet.* **80**, 142–148. https://doi.org/10.1159/000014971 (1998).
14. Muers, M. R. *et al.* Defining the cause of skewed X-chromosome inactivation in X-linked mental retardation by use of a mouse model. *Am. J. Hum. Genet.* **80**, 1138–1149. https://doi.org/10.1086/518369 (2007).
15. Plenge, R. M., Stevenson, R. A., Lubs, H. A., Schwartz, C. E. & Willard, H. F. Skewed X-chromosome inactivation is a common feature of X-linked mental retardation disorders. *Am. J. Hum. Genet.* **71**, 168–173. https://doi.org/10.1086/341123 (2002).
16. Furlan, G. & Galupa, R. Mechanisms of choice in X-chromosome inactivation. *Cells* https://doi.org/10.3390/cells11030535 (2022).
17. Fieremans, N. *et al.* Identification of intellectual disability genes in female patients with a skewed X-inactivation pattern. *Hum. Mutat.* **37**, 804–811. https://doi.org/10.1002/humu.23012 (2016).
18. Vianna, E. Q. *et al.* Understanding the landscape of X-linked variants causing intellectual disability in females through extreme X chromosome inactivation skewing. *Mol. Neurobiol.* **57**, 3671–3684. https://doi.org/10.1007/s12035-020-01981-8 (2020).
19. Gribnau, J. & Stefan Barakat, T. X-chromosome inactivation and its implications for human disease. *bioRxiv* https://doi.org/10.1101/076950 (2017).
20. Fieremans, N. *et al.* De novo MECP2 duplications in two females with intellectual disability and unfavorable complete skewed X-inactivation. *Hum. Genet.* **133**, 1359–1367. https://doi.org/10.1007/s00439-014-1469-6 (2014).
21. Brand, B. A., Blesson, A. E. & Smith-Hicks, C. L. The impact of X-chromosome inactivation on phenotypic expression of X-linked neurodevelopmental disorders. *Brain Sci.* https://doi.org/10.3390/brainsci11070904 (2021).
22. Renault, N. K. *et al.* Heritable skewed X-chromosome inactivation leads to haemophilia A expression in heterozygous females. *Eur. J. Hum. Genet.* **15**, 628–637. https://doi.org/10.1038/sj.ejhg.5201799 (2007).
23. Leehey, M. A. *et al.* FMR1 CGG repeat length predicts motor dysfunction in premutation carriers. *Neurology* **70**, 1397–1402. https://doi.org/10.1212/01.wnl.0000281692.98200.f5 (2008).
24. Hall, D. A. *et al.* X-inactivation in the clinical phenotype of fragile X premutation carrier sisters. *Neurol. Genet.* **2**, e45. https://doi.org/10.1212/nxg.0000000000000045 (2016).
25. Berry-Kravis, E., Potanos, K., Weinberg, D., Zhou, L. & Goetz, C. G. Fragile X-associated tremor/ataxia syndrome in sisters related to X-inactivation. *Ann. Neurol.* **57**, 144–147. https://doi.org/10.1002/ana.20360 (2005).
26. Yano, S. *et al.* Familial Simpson-Golabi-Behmel syndrome: studies of X-chromosome inactivation and clinical phenotypes in two female individuals with GPC3 mutations. *Clin. Genet.* **80**, 466–471. https://doi.org/10.1111/j.1399-0004.2010.01554.x (2011).
27. Shimojima, K. *et al.* Loss-of-function mutations and global rearrangements in GPC3 in patients with Simpson–Golabi–Behmel syndrome. *Hum. Genome Var.* **3**, 16033. https://doi.org/10.1038/hgv.2016.33 (2016).
28. Schirwani, S. *et al.* Duplications of GPC3 and GPC4 genes in symptomatic female carriers of Simpson-Golabi-Behmel syndrome type 1. *Eur. J. Med. Genet.* **62**, 243–247. https://doi.org/10.1016/j.ejmg.2018.07.022 (2019).
29. Simonetti, L. *et al.* Intelligence quotient variability in klinefelter syndrome is associated with GTPBP6 expression under regulation of X-chromosome inactivation pattern. *Front. Genet.* **12**, 724625. https://doi.org/10.3389/fgene.2021.724625 (2021).
30. Rheault, M. N. *et al.* X-inactivation modifies disease severity in female carriers of murine X-linked Alport syndrome. *Nephrol. Dial. Transpl.* **25**, 764–769. https://doi.org/10.1093/ndt/gfp551 (2009).
31. Juan-Mateu, J. *et al.* Prognostic value of X-chromosome inactivation in symptomatic female carriers of dystrophinopathy. *Orphanet J. Rare Dis.* **7**, 82. https://doi.org/10.1186/1750-1172-7-82 (2012).
32. Minamikawa, S. *et al.* Development of ultra-deep targeted RNA sequencing for analyzing X-chromosome inactivation in female Dent disease. *J. Hum. Genet.* **63**, 589–595. https://doi.org/10.1038/s10038-018-0415-1 (2018).
33. Zhang, X. *et al.* A new sex-specific underlying mechanism for female schizophrenia: Accelerated skewed X chromosome inactivation. *Biol. Sex Differ.* **11**, 39. https://doi.org/10.1186/s13293-020-00315-6 (2020).
34. Garagiola, I. *et al.* X Chromosome inactivation: A modifier of factor VIII and IX plasma levels and bleeding phenotype in Haemophilia carriers. *Eur. J. Hum. Genet.* **29**, 241–249. https://doi.org/10.1038/s41431-020-00742-4 (2021).
35. Fahim, A. T. *et al.* X-chromosome inactivation is a biomarker of clinical severity in female carriers of RPGR-associated X-linked retinitis pigmentosa. *Ophthalmol. Retina* **4**, 510–520. https://doi.org/10.1016/j.oret.2019.11.010 (2020).
36. Zhao, Y. *et al.* X chromosome inactivation pattern and pregnancy outcome of female carriers of pathogenic heterozygous X-linked deletions. *Front. Genet.* https://doi.org/10.3389/fgene.2021.782629 (2021).

37. Amos-Landgraf, J. M. *et al.* X chromosome-inactivation patterns of 1005 phenotypically unaffected females. *Am. J. Hum. Genet.* **79**, 493–499. https://doi.org/10.1086/507565 (2006).
38. Shvetsova, E. *et al.* Skewed X-inactivation is common in the general female population. *Eur. J. Hum. Genet.* **27**, 455–465. https://doi.org/10.1038/s41431-018-0291-3 (2019).
39. Knudsen, G. P., Pedersen, J., Klingenberg, O., Lygren, I. & Ørstavik, K. H. Increased skewing of X chromosome inactivation with age in both blood and buccal cells. *Cytogenet. Genome Res.* **116**, 24–28. https://doi.org/10.1159/000097414 (2007).
40. Busque, L. *et al.* Nonrandom X-inactivation patterns in normal females: Lyonization ratios vary with age. *Blood* **88**, 59–65 (1996).
41. Zito, A. *et al.* Heritability of skewed X-inactivation in female twins is tissue-specific and associated with age. *Nat. Commun.* **10**, 5339. https://doi.org/10.1038/s41467-019-13340-w (2019).
42. Kristiansen, M. *et al.* Twin study of genetic and aging effects on X chromosome inactivation. *Eur. J. Hum. Genet.* **13**, 599–606. https://doi.org/10.1038/sj.ejhg.5201398 (2005).
43. Vickers, M. A., McLeod, E., Spector, T. D. & Wilson, I. J. Assessment of mechanism of acquired skewed X inactivation by analysis of twins. *Blood* **97**, 1274–1281. https://doi.org/10.1182/blood.v97.5.1274 (2001).
44. Gao, X., Jia, M., Zhang, Y., Breitling, L. P. & Brenner, H. DNA methylation changes of whole blood cells in response to active smoking exposure in adults: A systematic review of DNA methylation studies. *Clin. Epigenet.* **7**, 113. https://doi.org/10.1186/s13148-015-0148-3 (2015).
45. Allen, R. C., Zoghbi, H. Y., Moseley, A. B., Rosenblatt, H. M. & Belmont, J. W. Methylation of HpaII and HhaI sites near the polymorphic CAG repeat in the human androgen-receptor gene correlates with X chromosome inactivation. *Am. J. Hum. Genet.* **51**, 1229–1239 (1992).
46. Machado, F. B. *et al.* 5meCpG epigenetic marks neighboring a primate-conserved core promoter short tandem repeat indicate X-chromosome inactivation. *PLoS One* **9**, e103714. https://doi.org/10.1371/journal.pone.0103714 (2014).
47. Wolf, S. F., Jolly, D. J., Lunnen, K. D., Friedmann, T. & Migeon, B. R. Methylation of the hypoxanthine phosphoribosyltransferase locus on the human X chromosome: Implications for X-chromosome inactivation. *Proc. Natl. Acad. Sci. U. S. A.* **81**, 2806–2810. https://doi.org/10.1073/pnas.81.9.2806 (1984).
48. Keith, D. H., Singer-Sam, J. & Riggs, A. D. Active X chromosome DNA is unmethylated at eight CCGG sites clustered in a guanine-plus-cytosine-rich island at the 5' end of the gene for phosphoglycerate kinase. *Mol. Cell. Biol.* **6**, 4122–4125. https://doi.org/10.1128/mcb.6.11.4122 (1986).
49. Multer, G. L. & Boynton, K. A. PCR bias in amplification of androgen receptor alleles, a trinucleotide repeat marker used in clonality studies. *Nucleic Acids Res.* **23**, 1411–1418. https://doi.org/10.1093/nar/23.8.1411 (1995).
50. Yang, Y. *et al.* Quantitative and multiplexed DNA methylation analysis using long-read single-molecule real-time bisulfite sequencing (SMRT-BS). *BMC Genom.* **16**, 350. https://doi.org/10.1186/s12864-015-1572-7 (2015).
51. Rupert, J. L., Brown, C. J. & Willard, H. F. Direct detection of non-random X chromosome inactivation by use of a transcribed polymorphism in the XIST gene. *Eur. J. Hum. Genet.* **3**, 333–343. https://doi.org/10.1159/000472322 (1995).
52. Höijer, I. *et al.* Detailed analysis of HTT repeat elements in human blood using targeted amplification-free long-read sequencing. *Hum. Mutat.* **39**, 1262–1272. https://doi.org/10.1002/humu.23580 (2018).
53. Gilpatrick, T. *et al.* Targeted nanopore sequencing with Cas9-guided adapter ligation. *Nat. Biotechnol.* **38**, 433–438. https://doi.org/10.1038/s41587-020-0407-5 (2020).
54. López-Girona, E. *et al.* CRISPR-Cas9 enrichment and long read sequencing for fine mapping in plants. *Plant Methods* **16**, 121. https://doi.org/10.1186/s13007-020-00661-x (2020).
55. McDonald, T. L. *et al.* Cas9 targeted enrichment of mobile elements using nanopore sequencing. *Nat. Commun.* **12**, 3586. https://doi.org/10.1038/s41467-021-23918-y (2021).
56. Giesselmann, P. *et al.* Analysis of short tandem repeat expansions and their methylation state with nanopore sequencing. *Nat. Biotechnol.* **37**, 1478–1481. https://doi.org/10.1038/s41587-019-0293-x (2019).
57. Gudmundsson, S. *et al.* TAF1, associated with intellectual disability in humans, is essential for embryogenesis and regulates neurodevelopmental processes in zebrafish. *Sci. Rep.* **9**, 10730. https://doi.org/10.1038/s41598-019-46632-8 (2019).
58. Simpson, J. T. *et al.* Detecting DNA cytosine methylation using nanopore sequencing. *Nat. Methods* **14**, 407–410. https://doi.org/10.1038/nmeth.4184 (2017).
59. Carrette, L. L. G., Blum, R., Ma, W., Kelleher, R. J. & Lee, J. T. Tsix–Mecp2 female mouse model for Rett syndrome reveals that low-level MECP2 expression extends life and improves neuromotor function. *Proc. Natl. Acad. Sci.* **115**, 8185–8190. https://doi.org/10.1073/pnas.1800931115 (2018).
60. Bhatnagar, S. *et al.* Genetic and pharmacological reactivation of the mammalian inactive X chromosome. *Proc. Natl. Acad. Sci.* **111**, 12591–12598. https://doi.org/10.1073/pnas.1413620111 (2014).
61. Carrette, L. L. G. *et al.* A mixed modality approach towards Xi reactivation for Rett syndrome and other X-linked disorders. *Proc. Natl. Acad. Sci.* **115**, E668–E675. https://doi.org/10.1073/pnas.1715124115 (2018).
62. Aguilar, R. *et al.* Targeting Xist with compounds that disrupt RNA structure and X inactivation. *Nature* **604**, 160–166. https://doi.org/10.1038/s41586-022-04537-z (2022).
63. Schneider, C. A., Rasband, W. S. & Eliceiri, K. W. NIH image to imageJ: 25 years of image analysis. *Nat. Methods* **9**, 671–675. https://doi.org/10.1038/nmeth.2089 (2012).
64. Labun, K. *et al.* CHOPCHOP v3: Expanding the CRISPR web toolbox beyond genome editing. *Nucleic Acids Res.* **47**, W171-w174. https://doi.org/10.1093/nar/gkz365 (2019).
65. Andrew, S. E., Goldberg, Y. P. & Hayden, M. R. Rethinking genotype and phenotype correlations in polyglutamine expansion disorders. *Hum. Mol. Genet.* **6**, 2005–2010. https://doi.org/10.1093/hmg/6.12.2005 (1997).
66. Lee, I. *et al.* Simultaneous profiling of chromatin accessibility and methylation on human cell lines with nanopore sequencing. *Nat. Methods* **17**, 1191–1199. https://doi.org/10.1038/s41592-020-01000-7 (2020).

Acknowledgements

Author contributions

J.J., M.-L.B. and M.W. conceived the study. J.J., M.W., M.-L.B. and A.A. contributed with experimental and analytic design. M.-L.B. and G.A. contributed with patient material. J.J., S.L., I.H. and S.G. performed the experiments. J.J., S.L., M.E. and M.W. performed the analysis. J.J. illustrated and prepared the figures. J.J. and M.W. drafted the manuscript. All authors read and approved the final manuscript.

Funding

Competing interests

J.J, M.W and M-L.B: Patent application pending "QUANTIFICATION OF X-CHROMOSOME INACTIVATION" (Application number 2250254-6). S.L, G.A, S.G, A.A, M.E, I.H: No competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-023-34413-3.

Correspondence and requests for materials should be addressed to M.W.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.