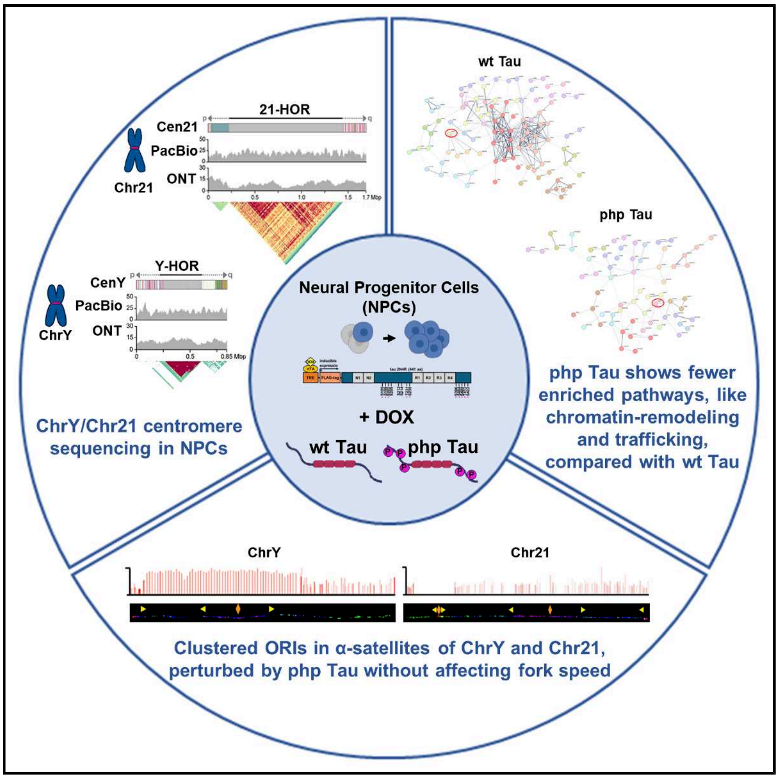
iScience

Article

Impact of Tau overexpression on DNA replication dynamics in centromeres of human neural progenitor cells

Graphical abstract



Highlights

- SMARD with long-read sequencing enables single-molecule replication analysis of centromeres
- Wt and php Tau differentially modulate centromeric replication of chromosomes Y and 21
- Php Tau has fewer nuclear interactors, mainly affecting chromatin remodeling and trafficking

Authors

Elisa Balzano, Shyam Twayana, Settapong T. Kosiyatrakul, ..., Kendra Hoekzema, Mirit I. Aladjem, Carl L. Schildkraut

Correspondence

elisa.balzano@genetics.utah.edu (E.B.), carl.schildkraut@einsteinmed.edu (C.L.S.)

In brief

Molecular biology; Epigenetics; Molecular neuroscience

Balzano et al., 2026, iScience *29*, 114707
February 20, 2026 © 2026 The Author(s). Published by Elsevier Inc.
https://doi.org/10.1016/j.isci.2026.114707

iScience

CellPress
OPEN ACCESS

Article

Impact of Tau overexpression on DNA replication dynamics in centromeres of human neural progenitor cells

Elisa Balzano,[1,10,*] Shyam Twayana,[1] Settapong T. Kosiyatrakul,[1] Glennis A. Logsdon,[2,11] Bhushan L. Thakur,[3] Evan E. Eichler,[2,4] Bettina Bohl,[5,6,7,8,12] Philipp Koch,[5,6,7,8] Simone Sidoli,[9] Anupama Kumari,[1] Katherine M. Munson,[2] Kendra Hoekzema,[2] Mirit I. Aladjem,[3] and Carl L. Schildkraut[1,13,*]
[1]Department of Cell Biology, Albert Einstein College of Medicine, Bronx, NY, USA
[2]Department of Genome Sciences, University of Washington School of Medicine, Seattle, WA, USA
[3]Developmen of Therapeutics Branch, Center for Cancer Research, NIH, Bethesda, MD, USA
[4]Howard Hughes Medical Institute, University of Washington, Seattle, WA, USA
[5]Department of Translational Brain Research, Central Institute of Mental Health (ZI), University of Heidelberg/ Medical Faculty Mannheim, Mannheim, Germany
[6]Hector Institute for Translational Brain Research (HITBR gGmbH), Mannheim, Germany
[7]German Cancer Research Center (DKFZ), Heidelberg, Germany
[8]German Center for Mental Health (DZPG), Partner Site Mannheim, Mannheim, Germany
[9]Department of Biochemistry, Albert Einstein College of Medicine, Bronx, NY, USA
[10]Present address: Department of Human Genetics, University of Utah School of Medicine, Salt Lake City, UT, USA
[11]Present address: Department of Genetics, Epigenetics Institute, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA
[12]Present address: Department of Bioengineering, Imperial College London, London, UK
[13]Lead contact
*Correspondence: elisa.balzano@genetics.utah.edu (E.B.), carl.schildkraut@einsteinmed.edu (C.L.S.)
https://doi.org/10.1016/j.isci.2026.114707

SUMMARY

Aging somatic cells are characterized by specific chromosome aneuploidy, particularly involving chromosome Y (ChrY) and chromosome 21 (Chr21), which are associated with Alzheimer's disease (AD) pathology. This study investigates the role of DNA replication within centromeric regions of these chromosomes using human neural progenitor cells engineered to overexpress either wild-type (wt) or pseudo-hyper-phosphorylated (php) Tau protein. We developed a method to analyze replication dynamics in centromeric DNA. Our findings reveal that replication origins and fork pausing events are mainly located within α-satellite sequences of ChrY and Chr21, where wt and php Tau distinctly modulate origin activation and initiation. Mass spectrometry analysis on immunoprecipitated Tau identified nuclear interactors of Tau, particularly in its php form, which might directly influence the chromatin architecture and gene expression. These studies provide critical insights into the molecular mechanisms of aneuploidy in tauopathies.

INTRODUCTION

Chromosome mis-segregation is one of the hallmarks for age-related diseases, such as neurodegenerative and cardiovascular diseases, and cancers.[1,2] The centromere, defined as the primary constriction of a chromosome, is the site where the kinetochore proteins assemble and recruit the microtubules for an accurate chromosome segregation. The centromere region is constituted in part by ∼171-bp tandem repeat units of α (alpha)-satellite monomers organized in layered clusters, called higher-order repeats (HORs).[3–9] The Telomere-to-Telomere (T2T) consortium recently completed the first complete sequence of a human genome, thanks to advances in long-read sequencing technologies and assembly algorithms.[10–12] Despite the crucial function of the centromeric sequences, they can accumulate mutations that, in turn, could be exacerbated in aging disorders.[13–16] Causative factors for the fragility of these centromeric repetitive sequences are still under study.[17,18] DNA replication process is prone to errors, especially for these highly repetitive sequences.[19] Impaired replication can result in DNA damage and, consequentially, in chromosome rearrangements and an improper centromere chromatin condensation and defective disjunction, which results in its misfunction.[13]

Unique aneuploidies, especially involving chromosome Y (ChrY)[20–27] and chromosome 21 (Chr21),[28–37] have been frequently observed in the somatic cells of Alzheimer's disease (AD) patients.[31–34,36]

The centromere of the human Y chromosome is the only functional human centromere that lacks the CENP-B (centromere protein B) box sequence, whereas all other chromosomes contain this motif. The CENP-B box is a 17-bp specific sequence recognized by the CENP-B protein, which stabilizes the centromere together with the constitutive centromere-associated network (CCAN) complex.[38–41] As a result, chromosome Y can mis-segregate at higher frequencies than other chromosomes, generating mosaic cells with loss of chromosome Y (mLOY).[42] Recently, mLOY was shown to be involved in AD pathogenesis and it is considered as a male AD risk factor.[43]

Chr21 mis-segregation is historically related to AD pathology. Indeed, individuals with the trisomy of the Chr21 affected by Down syndrome are exposed to an elevated risk for the early-onset AD-like dementia.[44,45] Possibly, this relationship between Down syndrome and AD lies in the triplicate copy of the amyloid precursor protein (APP) gene on Chr21.[37] Amyloid deposit abundance is considered the most prominent cause for a genetic or familiar AD development.[46]

Here, we investigated a potential cause of two specific-chromosome aneuploidies, modeling the AD background by overexpression of two variants of Tau. Tau is a microtubule-associated protein (MAPT), and its pathological aggregation is present in the brain cells of AD patients.

A higher concentration of nuclear Tau can lead to its interaction with the DNA helix, such as rDNA and centromeric α-satellite DNA.[47,48] Tau was also suggested to have a role in chromatin condensation.[48–50] It has been demonstrated that Tau is involved in chromosome stability, and its malfunction can lead to aneuploidy.[51] Neurons and glial cells can aberrantly re-enter the cell cycle under stress or aging, often leading to aneuploidies and genomic instability.[52] These events are linked to neurodegeneration and suggest a possible connection between early Tau expression, cell cycle dysregulation, and AD progression. However, the molecular mechanism of specific chromosome aneuploidy in AD and other tauopathies remains unclear.[53]

Our aim was to elucidate the effect of the overexpressed Tau protein on the replication dynamics of centromeric α-satellite DNA of two of the most mis-segregated chromosomes in AD, ChrY and Chr21. The goal was to advance our understanding of early and presymptomatic AD risk factors, which can also be beneficial for targeted treatment development.

RESULTS

Neural progenitor cells overexpressing Tau

We generated a human neural progenitor cell (NPC) line, named CIMH-NP-3, from induced pluripotent stem cells of a heathy and young man, to simulate the neurological conditions of AD (see STAR Methods, Figures 1A and 1B). NPCs are in specialized niches of the human adult brain, and they can divide and replace various neural cell types upon brain damage[55,56]; thus, they represent a good controlled cellular model to study neurological diseases.

Two isogenic cell lines were produced by the transduction of CIMH-NP-3 cells with EtO (Ef1a TetOn) lentiviral system for doxycycline (DOX)-inducible expression of FLAG-tagged full-length 2N4R wild-type Tau (wt Tau) protein or 2N4R pseudo-hyper-phosphorylated Tau (php Tau) protein variant, as described previously[54] (Figures 1A, 1C, 1D, and S1A). The php Tau biologically mimics the pathological phosphorylated Tau, found mainly in AD patients, due to the substitution of serine at nine positions (195, 199, 202, 235, 396, 404, 409, 413, and 442) and one tyrosine position (231) by glutamic acid[54] (Figure 1A).

Isogenic cell lines eliminate genetic variability for evaluating the cellular responses under controlled conditions.

The transduced cells were selected by the antibiotics geneticin and puromycin antibiotics (Figure 1B).

We selected a single clonal cell line for both wt Tau and php Tau cells that have only one insertion of the Tau expression cassette (Figures S1C and S1D) and showed the DOX-induced expression of the specific Tau variant (Figures 1C and 1D). The stemness in these neural progenitors was verified by the presence of multipotent neural stem cell markers such as Sox2 (SRY-box transcription factor 2) and Nestin, and the absence of the differentiated neuronal marker Tubb3 (tubulin beta 3 class III) (Figures 1E–1G and S1B).

Since the focus of our study is on both ChrY and Chr21, their presence was verified in metaphase spreads and interphasic nuclei, using chromosome paint (MetaSystems) fluorescence *in situ* hybridization (FISH) probes (Figure S1E: one spot for ChrY, and two spots for Chr21).

Long-read sequencing coupled to SMARD analysis for α-satellite of a specific chromosome and allele

Since we wanted to analyze centromere replication at a high resolution, we used single-molecule analysis of replicated DNA (SMARD). SMARD is a powerful technique developed in our laboratory to visualize the localization of origins of replication (ORIs), replication termination sites, and the replication fork direction in specific loci, using FISH probes on single molecules derived from single cells[57,58] (Figure 2A). The replication forks can be visualized and followed by adding thymidine analogs that are incorporated only in the proliferating cells: 5-iodo-2′-deoxyuridine (IdU) for 4 h and 5-chloro-2′-deoxyuridine (CldU) for an additional 4 h. We chose not to synchronize the cells to ensure the native DNA replication process. We adopted the SMARD approach for the satellite repeats. However, due to the repetitive nature of α-satellite DNA, the replication dynamics of these sequences can be challenging to determine. We used Pacific Biosciences (PacBio) high-fidelity (HiFi) and Oxford Nanopore Technologies (ONT) sequencing methods, developed by the T2T consortium,[8,10–12,59] to sequence and assemble the genome of the neuronal progenitor cells that we used in this study and, in particular, to focus on the centromeres of ChrY and Chr21 (cenY and cen21, respectively). We validated the assemblies of these two centromeres using three orthogonal approaches to ensure their correct construction. First, we aligned native long-read sequencing data from the same source genome to the assemblies and confirmed uniform read depth along the cenY and cen21 regions, which indicated that these regions are free of large structural errors. Next, we applied an algorithm, VerityMap,[60] which identifies discordant *k*-mers between the centromere assemblies and PacBio HiFi reads, and found no evidence of discordance. Finally, we applied a method, GAVISUNK,[61] that compares single unique nucleotide *k*-mers (SUNKs) in the centromere assemblies to those in the ONT reads generated from the same sample and observed support for each SUNK with orthogonal ONT data.

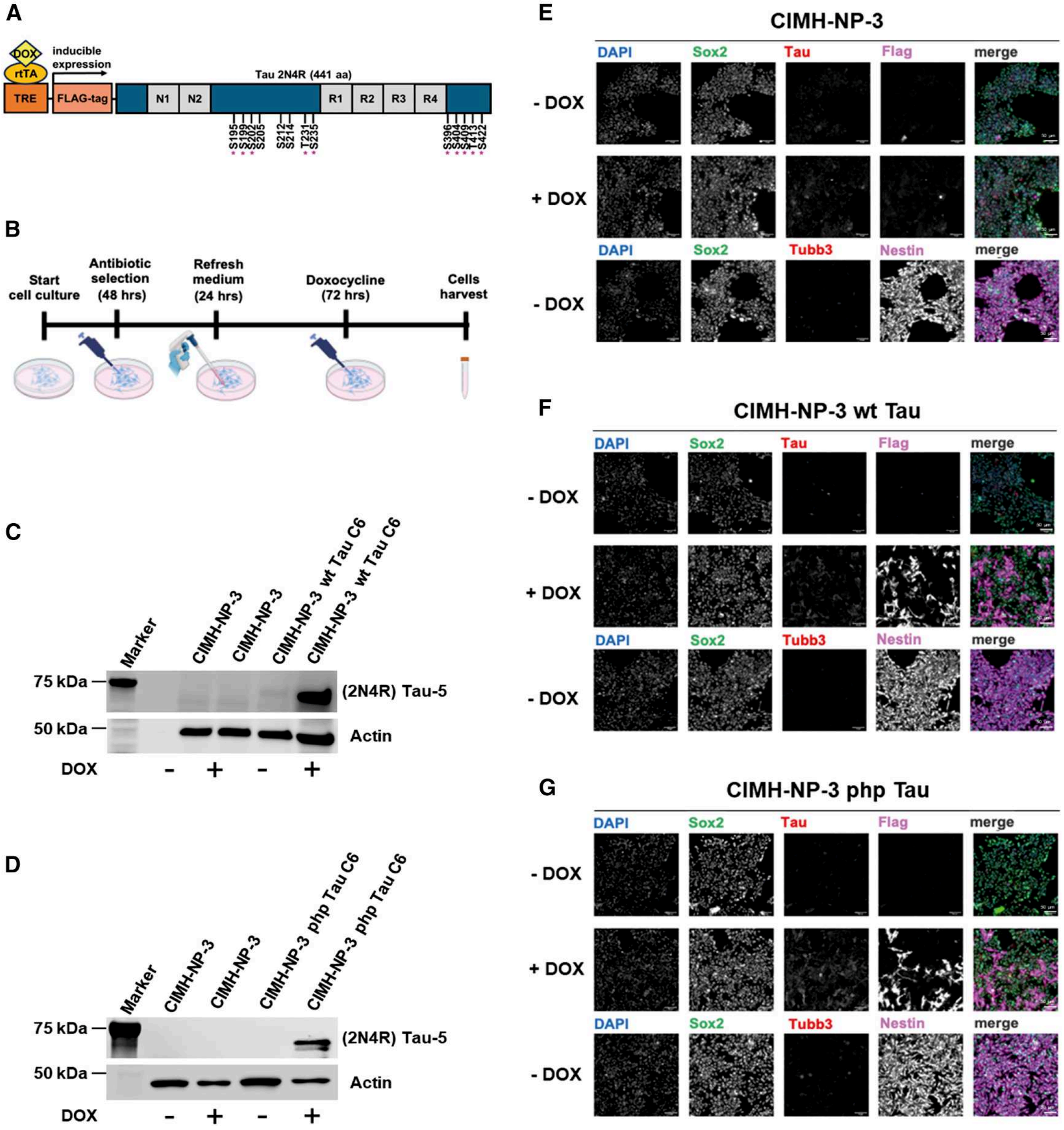


Figure 1. Characterization of CIMH-NP-3 cell lines

(A) A schematic of the pseudo-hyper-phosphorylated Tau (php Tau) in which 10 serine/threonine residues (shown with *) are substituted with glutamic acid compared with wild-type Tau (wt Tau), adapted from a study.[54] Tau 2N4R is composed of 441 amino acids.

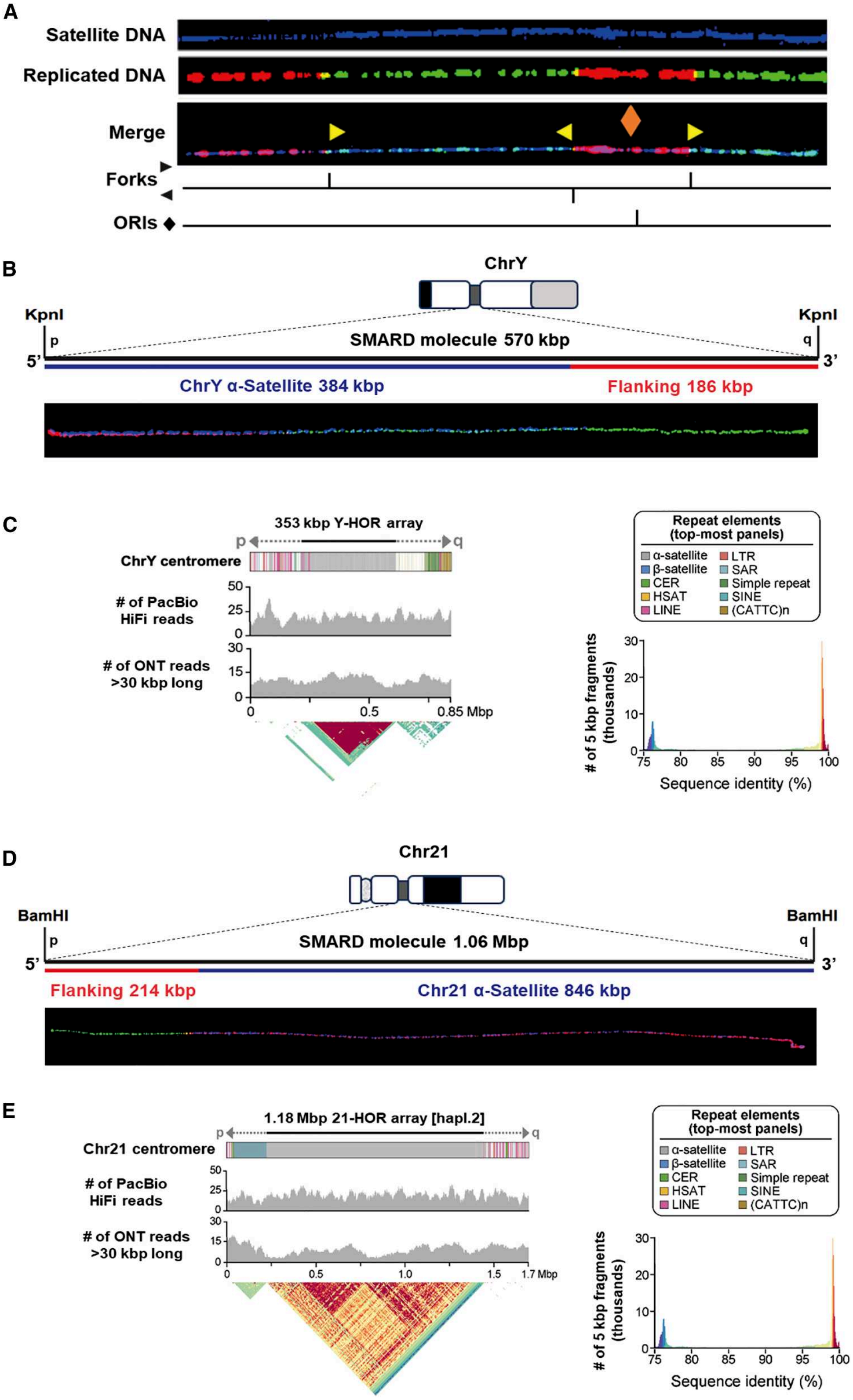
(B) Schematic temporal line from the start of the cell culture to the harvest of cells, including antibiotic selection of transduced cells and Tau induction upon doxycycline treatment for 3 days.

(C and D) Immunoblot for the expression of wt and php Tau in clone #6 of each CIMH-NP-3 cell line using anti-Tau-5 antibody. Actin was used as loading control.

(E–G) CIMH-NP-3 parental cell line (−/+DOX) (E), wt Tau cell line (F), and php Tau cell line (G). Tau protein (red) and FLAG tag (magenta) are shown by immunofluorescence. Moreover, the expression of the neuronal progenitor markers Sox2 (green) and Nestin (magenta) and the absence of the neuronal marker Tubb3 (red) were verified by immunofluorescence (bottom images -DOX). Scale bars, 50 μm.

Obtaining the sequences of cenY and cen21 helped us to set up a strategy for using our locus-specific replication assay, SMARD.

The ChrY SMARD molecule was obtained by cutting ChrY with the KpnI restriction enzyme, which generates a 570-kbp DNA molecule, consisting of a 384-kbp region of Y α-satellite (full Y-HOR) and a 186-kbp region of flanking DNA (Figures 2B and S2A). The comparison of length between the linear sequence and SMARD molecule was in agreement with the sequencing results. The ChrY centromere in CIMH-NP-3 cells consists of a highly homogeneous 353-kbp α-satellite HOR array (Y-HOR), as shown in Figure 2C. In the regions outside of the Y-HOR are other classes of repetitive elements, as described in the panel to the right (Figure 2C). The flanking



(legend on next page)

region provided directional orientation (5′-3′) for our repetitive SMARD segments.

In contrast to the hemizygous condition of ChrY, Chr21 is diploid and consists of two chromosome homologs or haplotypes. Moreover, although the Chr21 α-satellite HORs (21-HOR) are highly similar to those in chromosome 13,[11] we were able to discriminate the two different chromosomes by the molecule size and flanking sequence found *in silico* and by Southern blots (Figures S2B and S2C). For these purposes, we selected only one of the two Chr21 haplotypes to study DNA replication precisely in our SMARD molecules. The haplotype selection was made by assessing the differential length of the two Chr21 SMARD molecules by Southern blots and *in silico* (Figure S2B). To generate the SMARD molecule for the α-satellite HOR array of the Chr21 hapl.2 (haplotype 2), we cut Chr21 with the restriction endonuclease BamHI, which produces a 1.06-Mbp DNA molecule, consisting of a 846-kbp region of Chr21 α-satellite DNA (almost the entire 21-HOR) and a 214-kbp region of flanking DNA (Figures 2D, 2E, and S2E). We noted that the 21-HOR haplotype 2 molecule cut by BamHI is constituted by 98% of α-satellite compared with 88% in the 21-HOR haplotype 1 molecule (Figures S2D and S2E). This optimized SMARD is, thus, chromosome specific and, most importantly, haplotype specific.

Paused replication forks at ChrY α-satellites and compensatory ORI activation upon php Tau overexpression

For a rigorous and reproducible analysis of centromeric repeats on long SMARD molecules, we focused on well-characterized replication events: ORIs, termination, and replication-pausing events (Figure 3A). To quantify replication dynamics, each molecule was divided into 30 kbp intervals or bins numbered according to genomic position (e.g., bin 1 spans 0–30 kbp, and bin 2 spans 30–60 kbp). The number of replication forks and ORIs per interval was measured, and intervals with elevated z-scores (illustrated as * in Figures 3B and 3C) indicated difficult-to-replicate sites, consistent with replication pausing (Table S1).

SMARD analysis of ChrY revealed a distinctive cluster of centromeric replication events (Figures 3B, 3C, and S3A–S3D). Both replication forks in both directions (5′-3′ and 3′-5′) and ORIs were predominantly detected within the α-satellite regions, whereas the adjacent flanking regions showed few or no initiation events (Figures S3A–S3D). Across all cell lines in control condition, replication forks accumulated within the first ∼120 kbp of the Y α-satellites (bin 4) (Figure 3B), suggesting the presence of clustered replication pauses within a short interval. This pattern indicated an intrinsic difficulty in fully replicating these repetitive sequences.

Overexpression of both wt Tau and php Tau further enhanced pausing at around 120 kbp (bin 4). Upon wt Tau overexpression, replication-pausing events were also detected at approximately 150 kbp (bin 5) (Figure 3B).

To assess whether these pausing events were fully compensated, we analyzed the activation of ORIs along the molecules (Figure 3C). Interestingly, we observed increased variability in the location of replication initiation sites compared to control cells, suggesting a shift in ORI usage. In wt Tau + DOX cells, a distinct ORI activation was observed near 180 kbp (bin 6), whereas in php Tau + DOX cells, additional ORIs emerged at ∼120, 270, and 390 kbp (bins 4, 9, and 13), suggesting a compensatory mechanism to overcome the replication stress (Figure 3C). Notably, only php Tau + DOX cells exhibited 68% of molecules with active ORIs distributed throughout the α-satellite regions, rather than being confined to a specific locus, compared to 32% of molecules in the control condition (Figure S3E). Increased pausing was followed by an adaptive response to replication stress, enabling fork rescue facilitated by the initiation of additional replication events along the molecule. Since the replication fork speed (DNA sequence replicated from one single ORI) was not altered (Figure S3F), these observations likely reflect a mechanistic interference with ORI activation or their restart.

Overexpressed php Tau disrupts fork progression at the flanking region of cen21

Chr21 was technically challenging given the long length of the SMARD molecules, both for the molecule stretching and

Figure 2. Validation of the SMARD molecules of the ChrY and Chr21 centromeric regions

(A) Representative SMARD molecule shows different steps of DNA replication including initiation (ORIs) (◆), fork progression from IdU (red) to CldU (green) (▸: 5′-3′; ◂: 3′-5′), and termination (▸◂). Replication initiation events are predicted to be at the centers of the red signals, which are flanked by two green signals. Below the four SMARD molecules, there are two diagrams, which summarize the replication processes captured by the SMARD. The top diagram summarizes the fork progressions: two forks move from 5′ to 3’ (the vertical lines pointing up), while one fork moves from 3′ to 5′ (the vertical line pointing down). The bottom diagram summarizes the fork initiation event (ORI).

(B) A diagram of the ChrY centromeric SMARD molecule. ChrY (57 Mbp) was cut by the KpnI restriction enzyme to generate a 570-kbp DNA segment of ChrY SMARD molecule (the black line), which is composed of a 384-kbp DNA region of ChrY α-satellite DNA (the blue line) and a 186-kbp DNA region of flanking DNA (the red line). A representative ChrY SMARD molecule is also shown. Blue, signals detected using a FISH probe for ChrY α-satellite; red, IdU incorporation; green, CldU incorporation.

(C) Long-read DNA sequencing confirmed the size and molecular composition of the ChrY molecule detected by SMARD. The gray color indicates the α-satellite HOR of ChrY (Y-HOR, 353 kbp). The insert to the right shows the other repeat elements flanking the Y-HOR in different colors. The heatmap visualized by StainedGlass indicates the pairwise sequence identity with the percentage identity of 5-kbp fragments.

(D) A diagram of the Chr21 centromeric SMARD molecule. Chr21 (37 Mbp) was cut by the BamHI restriction enzyme to generate a 1.06-Mbp DNA segment of the hapl.2 (haplotype 2) Chr21 SMARD molecule (the black line), which is composed of an 846-kbp DNA region of Chr21 α-satellite DNA (the blue line) and a 214-kbp DNA region of flanking DNA (the red line). A representative Chr21 SMARD molecule is also shown. Blue, signals detected using a FISH probe for Chr21 α-satellite; red, IdU incorporation; green, CldU incorporation.

(E) Long-read DNA sequencing confirms the size and molecular composition of the hapl.2 Chr21 SMARD molecule. The gray color indicates the α-satellite HOR of Chr21 (21-HOR, 1.18 Mbp). The insert to the right shows the other repeats flanking the 21-HOR haplotype 2 in different colors. The heatmap visualized by StainedGlass indicates the pairwise sequence identity with the percentage identity of 5-kbp fragments.

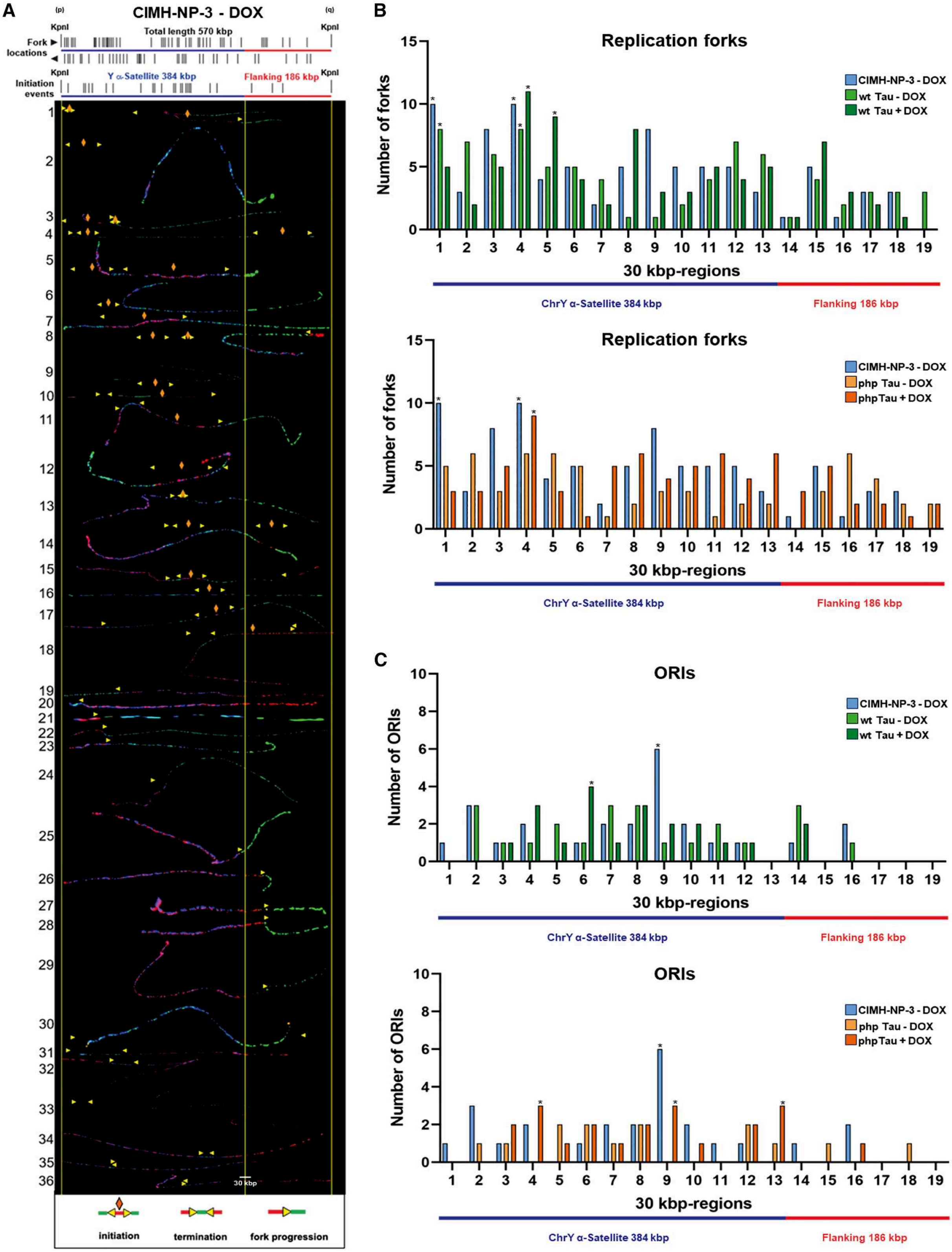


Figure 3. Pausing events at Y α-satellites and secondary ORI activations under Tau overexpression

(A) Representative analysis of the aligned centromeric ChrY SMARD molecules from the CIMH-NP-3-DOX cells ($N = 36$). The region detected by biotin-labeled, Y-specific α-satellite DNA FISH probe is shown in blue, and the flanking sequence is shown in red. On the top, two diagrams indicate the fork locations and

(legend continued on next page)

integrity. However, we were able to collect well-stretched SMARD molecules for cen21 (Figures 4A and S4A).

In CIMH-NP-3 cells, replication fork-pausing events were detected at approximately the first 120 kbp within the α-satellite region of Chr21 (bin 11), extending from the flanking sequence (Figure 4B). In wt Tau cells without DOX, pausing events occurred in the α-satellites further downstream at ∼180, 240, and 330 kbp (bins 13, 15, and 18), while in wt Tau overexpression, pausing sites shifted to a specific site at ∼210 kbp (bin 14). In php Tau cells without DOX, replication pausing was observed within the first 30 kbp of the 21 α-satellite region (bin 8), near the boundary region between the 21 α-satellites and the flanking region, and at 180 and 240 kbp of the α-satellites (bins 15 and 18) under wt Tau overexpression condition; in php Tau overexpression, an increased number of replication forks was observed within the last 30 kbp of the flanking region (bin 7) (Figure 4B). The complete replication of the flanking region depends on the progression of long replication forks originating from the 21 α-satellite region.

Indeed, both the α-satellite and flanking regions show long replication forks originating from active ORIs within Chr21 α-satellites travel toward the flanking sequences, and not vice versa (3′-5′; Figures S4B and S4C). All the CIMH-NP-3 cell lines showed ORI activation specifically around 30–390 kbp of the α-satellite regions (Figure 4C). As for the SMARD molecules of the cenY, replication forks and ORIs were found mostly in the 21 α-satellite rather than the flanking region (Figure S4D).

While the overexpression of wt Tau reduced the activity of constitutive ORIs, this effect did not apply to the ORIs at ∼210 kbp (bin 14), where pausing events were located. In addition, secondary ORIs were observed around 60, 210 and 330 kbp within the α-satellite region (bins 9, 16, and 18). Under php Tau overexpression, constitutive ORIs ∼30 and 150 kbp of the α-satellite region of Chr21 (bins 8 and 12) remained active, while the usage of other constitutive ORIs observed in control condition was diminished (Figure 4C). The reduced activation of constitutive ORIs, together with fork stalling in flanking regions, indicates a compromised replication process that may result in incomplete DNA synthesis.

In both wt Tau and php Tau overexpression, no additional molecules showed either increased α-satellite ORIs compared with the control or any detectable changes in replication fork speed observed (Figures S4E and S4F, Table S1).

Finally, all three cell lines showed similar replication dynamics in untreated condition across cenY and cen21, confirming consistent replication dynamics. Both wt and php Tau expression enhanced replication pausing within α-satellite regions but not in the flanking sequences, indicating an α-satellite-specific effect of Tau. However, the resolution of pausing events was mediated differently between wt Tau and php Tau.

Repetitive sequences are enriched in active ORIs

To have a better understanding of the replication dynamics within tandemly repetitive regions, we analyzed the ribosomal DNA (rDNA) repeat units as controls for the centromeric satellite sequences. Each rDNA repeat consists of 13-kbp transcribed region and 32-kbp intergenic spacer. The 13 kbp-transcribed region encodes 18S, 5.8S, and 28S rRNAs, which are important components of the human ribosomes, while the 32 kbp-intergenic spacer contains regulatory elements of the adjoining rDNA repeat[62] (Figure S5A). We cleaved genomic DNA using the EcoRV restriction endonuclease to produce a 240-kbp segment, which was found to contain five canonical tandem (head-to-tail) rDNA repeats (Figures S5A and S5B). Two asymmetric probes for rDNA repeat, pA and pC, were used to detect and align each rDNA molecule in the correct orientation. We analyzed the replication program of this rDNA segment in CIMH-NP-3 cells by using SMARD (Figure S5B). Among 25 molecules, 11 showed initiation events in the first three rDNA repeats. None of the molecules showed initiation in the last two repeats. This indicated that replication initiations are localized to the first three rDNA repeats of this DNA segment in the CIMH-NP-3 cells, with the evidence for last two repeats being replicated from the origins that are activated outside of the rDNA repeats. We observed more replication forks in the 3′ to 5′ direction, suggesting the presence of further replication initiation events outside the 3′ end of the rDNA segment (Figure S5C).

Tau overexpression decreases the global and centromeric replication activity

From the SMARD results, we showed replication fork-stalling events strongly at the α-satellites in both ChrY and Chr21. We were able to display how many active origins were localized in one molecule derived from a single cell, and termination events were located. Thanks to the visual transitions from red to green, which enabled us to distinguish the direction of the replication forks.

To provide further insights into the results obtained by SMARD, DNA replication dynamics were further analyzed using a high-resolution orthogonal approach, nascent strand sequencing (NS-seq).[63–67,63–68]

This technique measures the specific localization of active ORIs by isolating short, RNA-primed DNA fragments from non-synchronized proliferating cells, quantifying the number of sequencing reads associated with each ORI[68,69] (Figure 5A). Thus, this method provides both the precise localization of the ORIs and quantitative insights into their usage. Using NS-Seq, we first evaluated the effects of overexpressed wt and php Tau on the global activation of the ORIs (Figures 5B–5D). As shown in the zoomed cytogenetic region 21q2 (Figure 5B), distinct

initiation events of the SMARD molecules. The 570-kbp SMARD molecules are aligned to the yellow lines that correspond to the external KpnI cut sites and to the transition region between Y α-satellite and flanking regions (middle yellow line). SMARD molecules are arranged in the following order: initiation events (ORIs), fork progressing, and termination events. The initiation events are marked by orange diamonds (◆); the orientation of the replication forks is shown as the transition from IdU (red) to CldU (green), represented by yellow arrows ▸ 5′-3′ or ◂ 3′-5′; and the termination events (▸◂). A 30-kbp scale bar is illustrated at the bottom of the aligned SMARD molecules; 1 kbp is 0.38 μm.

(B and C) Quantification of the replication forks and ORIs within every 30 kbp (bin) of the SMARD molecules from several replicates for CIMH-NP-3-DOX (n = 36), wt Tau (n = 34,35), and php −/+DOX (n = 31,34). n = number of molecules. Graphs were generated using the same CIMH-NP-3 −DOX dataset.

The *Z*-score was calculated as $z = (x - \mu)/\sigma$, where x is the observed value, μ is the mean, and σ is the standard deviation (one-tailed). Values ≥ 1.645 were considered significant*. Clone #6 for wt Tau and clone #6 for php Tau were used for the SMARD analysis.

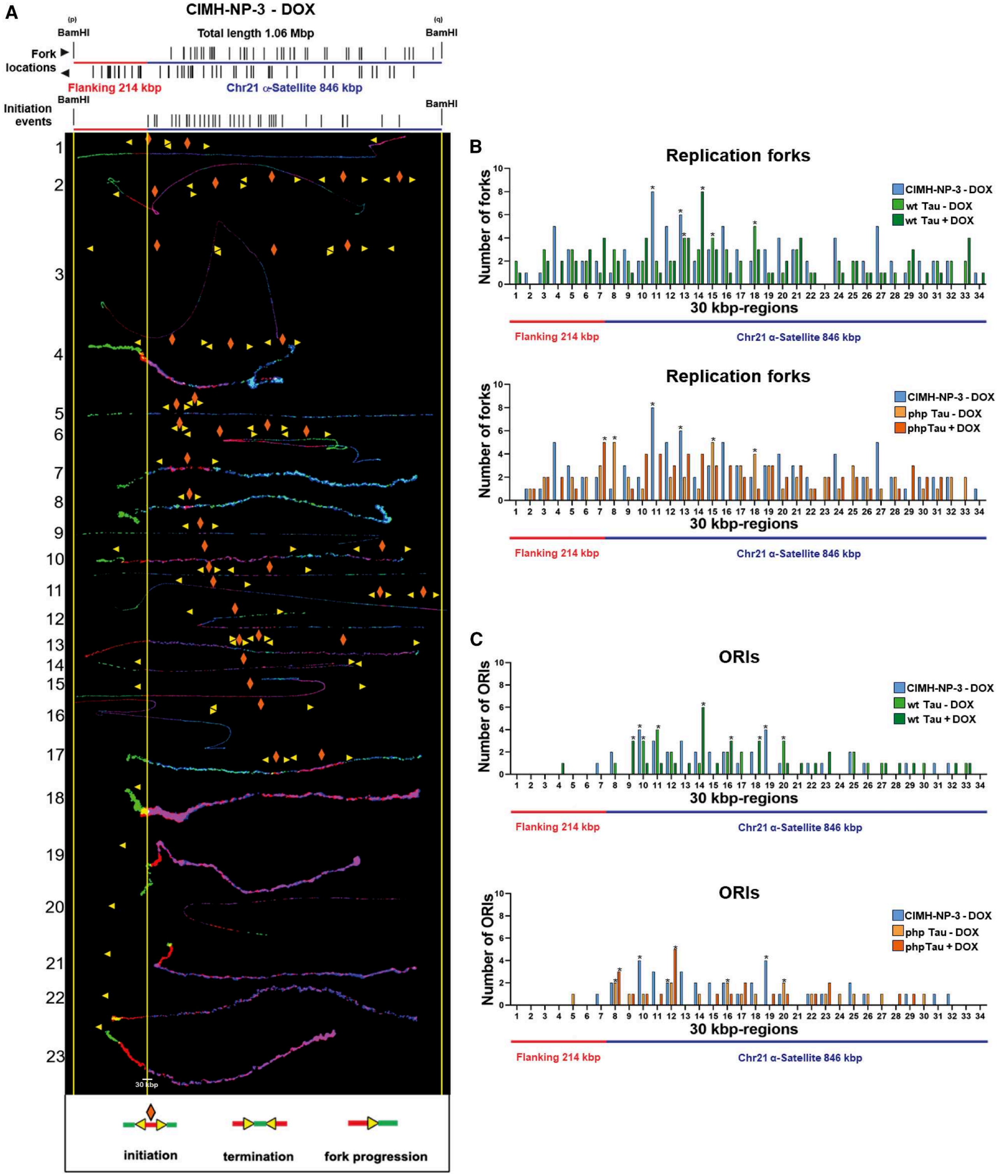


Figure 4. Expression of wt Tau and php Tau affects the replication in the Chr21 α-satellite and flanking regions

(A) Representative analysis of the aligned centromeric Chr21 SMARD molecules from the CIMH-NP-3-DOX cells (*N* = 23). The region detected by biotin-labeled, 21 α-satellite-specific DNA FISH probe is shown in blue, and the flanking sequence is shown in red. On the top, two diagrams indicate the fork locations and initiation events of the SMARD molecules. The 1.06-Mbp SMARD molecules are aligned to the yellow lines that correspond to the external BamHI cut sites and to the transition region between 21 α-satellites and flanking region (middle yellow line). SMARD molecules are arranged in the following order: initiation events

(legend continued on next page)

ORIs can be identified as peaks that can be visualized on a genome browser, with the relative abundance of any particular sequence in the nascent strand population represented as a fraction between 0 and 1. Alignment of NS-seq data among the three cell lines (CIMH-NP3 and the two derived cell lines, expressing either wt or php Tau) with and without DOX could identify common origins (gray arrows), origins that activated only in the presence of DOX in wt Tau and php Tau (blue arrows), and origins that inactivated by DOX treatment (red arrows). As shown in Figures 5C and 5D, the expression of Tau (either wt or php) profoundly altered ORI utilization. Overexpression of either wt or php Tau decreased the total number of active origins by 20%–25%, and about half of the ORIs activated in each of the two Tau-overexpressing cell lines were not utilized in the parental cell lines. Notably, global origin usage in cells overexpressing either wt or php Tau was highly concordant ($\sim$96% of ORIs share the same locations). We then focused on the initiation of DNA replication in the α-satellite and flanking regions of cenY and cen21 (Figures 5E and 5F). The treatment with DOX in the parental cell line did not alter the locations of ORIs in these centromeric regions, whereas the overexpression of either wt or php Tau profoundly reduced the number of active initiation sites in both cell lines (Table S1).

For ChrY, replication initiated within the α-satellite sequences as well as within the flanking region, with a robust preference to origin activation within the α-satellites (Figure 5E). Comparison of replication initiation event frequency under wt and php Tau overexpression revealed that 68%–69% of origins remained stably active relative to CIMH-NP-3 cells. When comparing wt and php Tau conditions, about 25% of origins differed, while 75% were conserved. These results suggest that each variant of Tau modulates a specific subset of ORIs within centromeric regions (Figure S5D). For Chr21, replication initiated mainly within the α-satellite sequences, while no substantial enrichment was observed in the flanking region (Figure 5F).

Replication initiation analysis in cen21 showed that about 28% of origins remained consistently active under wt Tau overexpression and 50% under php Tau overexpression compared with CIMH-NP-3 cells, while about 14% differed between wt and php Tau conditions (Figure S5E). As summarized in Figure 5G, these data suggest that Tau protein overexpression reduced the frequency of ORI utilization globally, and to a markedly higher extent in the α-satellite sequences of cenY and cen21.

Proteomics analysis of Tau and its interactors

Thus, to better understand the interaction landscape of Tau in NPCs, we conducted quantitative mass spectrometry analysis following Tau immunoprecipitation (Figure 6A). We wanted to investigate the Tau nuclear interactors, since it has also been reported to be localized in the nucleus.[70,71] Both wt and php forms of Tau were examined in the nuclear fraction. The proteomics approach aimed to reveal differences in protein binding partners in this cellular compartment under the two different Tau forms. Tau signal intensity was normalized using a data processing procedure described in a previous publication.[72]

Tau abundance data were further refined and presented in $\log_2$-transformed form to facilitate comparisons between conditions and to visualize subtle yet significant differences in protein levels. Here, Tau levels are grouped by ectopic expression of either wt or php Tau (Table S2).

By performing gene set enrichment analysis (GSEA) which identifies biological pathways or protein groups statistically overrepresented in a dataset based on ranked protein levels, we observed distinct enrichment profiles depending on the Tau variant. The normalized enrichment score (NES) quantifies the strength and direction of enrichment, while the false discovery rate (FDR) indicates the significance after correction.

Proteins enriched by the FLAG tag M2 antibody (blue bars) display variant-specific patterns between wt and php Tau, whereas proteins enriched by the IgG control (orange bars) mainly correspond to canonical nuclear components, reflecting background nuclear material (Figures 6B–6D). Using STRING software, we evaluated the proteins significantly enriched by the FLAG tag M2 antibody relative to the corresponding IgG control, focusing on those derived from the nuclear fraction (Figures 6C–6E). The results suggest that several previously characterized Tau interactors are co-enriched in the FLAG tag M2 immunoprecipitation, reinforcing the specificity and relevance of our approach.

Regarding php Tau, fewer significantly enriched pathways were noted compared with wt Tau, which strongly associates with ribosomal biogenesis proteins. In addition, php Tau appears to pull down fewer complexes or proteins belonging to specific pathways, consistent with its less significant enrichment scores. Interestingly, php Tau shows a depletion of protein functions related to chromatin remodeling and intracellular trafficking, whereas wt Tau preferentially enriches trafficking-associated proteins, as expected, such as keratins, and mechanical links between the cytoskeleton and the nucleus.

DISCUSSION

Our study reports how the overexpression of two variants of the longest Tau isoform in human NPCs differentially impacts centromeric DNA replication dynamics in ChrY and Chr21. These findings show how php Tau, in particular, can contribute to chromosomal instability and brain mosaic aneuploidies—unseen drivers of neurodegenerative process—by disturbing the delicate balance of centromeric replication dynamics.

The application of high-throughput sequencing and single-molecule analysis techniques opens a way for examining the

---

(ORIs), fork progressing, and termination events. The initiation events are marked by orange diamonds (◆); the orientation of the replication forks is shown as the transition from IdU (red) to CldU (green), represented by yellow arrows ▶he′-3′ or ◀ 3′-5′; and the termination events (▶◀). A 30-kbp scale bar is illustrated at the bottom of the aligned SMARD molecules. 1 kbp is 0.38 μm.

(B and C) Quantification of the replication forks and ORIs within every 30 kbp (bin) of the SMARD molecules from several replicates for CIMH-NP-3-DOX (n = 23), wt Tau (n = 23,23), and php −/+DOX (n = 21,22). n, number of molecules. Graphs were generated using the same CIMH-NP-3 − DOX dataset.

The *Z*-score was calculated as $z = (x - \mu)/\sigma$, where x is the observed value, μ is the mean, and σ is the standard deviation (one-tailed). Values ≥ 1.645 were considered significant*. Clone #6 for wt Tau and clone #6 for php Tau were used for the SMARD analysis.

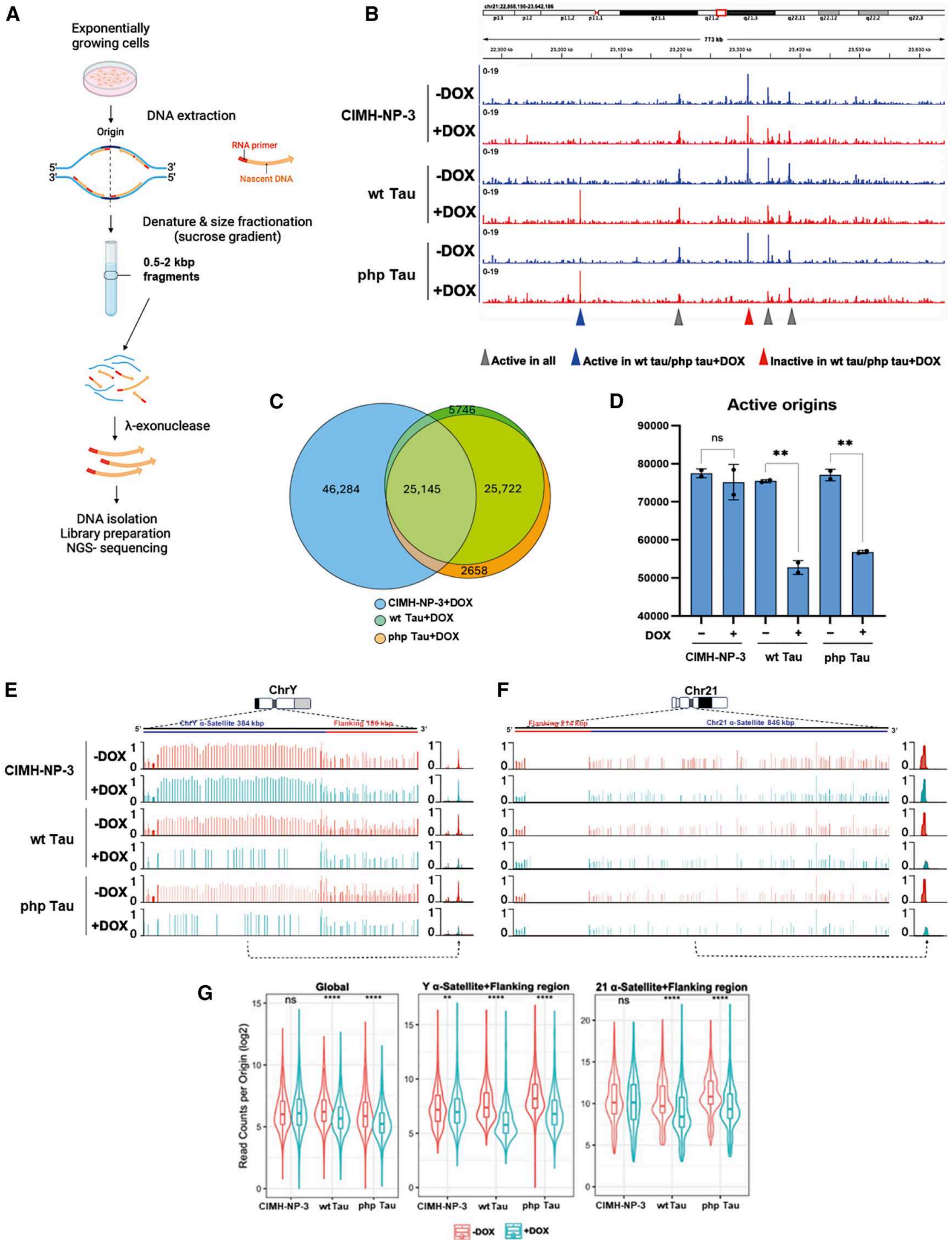


Figure 5. Tau overexpression alters global replication initiation and reduces origin initiation frequency in α-satellites of ChrY and Chr21

(A) A schematic of the preparation of NS-seq. Proliferating cells are used to extract the genomic DNA and the newly synthetized strands of DNA. The newly synthesized strands of DNA are isolated from genomic DNA by sucrose gradient (ranging size 0.5 to 2 kbp, not including the lagging strand Okazaki fragments). After the digestion of DNA without RNA primers at the 5′ by λ exonuclease to select only the RNA-primed DNA, the sequencing library is prepared to sequence and align to a reference genome (T2T_v2 or CIMH-NP-3 parental cell line). Created in BioRender. Balzano, E. (2025) https://BioRender.com/zx7j271.

(B) Genome browser snapshots showing the ORIs mapped by NS-seq in 773 kbp of the q-arm of Chr21, located at 21q21 (aligned using T2T_v.2). Each peak corresponds to an active ORI. Gray arrowheads mark origins active in all three conditions (CIMH-NP-3 parental, wt Tau + DOX, and php Tau + DOX cells). Blue and red arrowheads mark origins active or inactive only in wt/php Tau, respectively.

(legend continued on next page)

replication dynamics in centromeres and in other repetitive genomic regions in chromosome- and haplotype-specific manners (such as rDNA and telomeres), in different tissue and cell types. Combining NS-seq and single-fiber analyses assays characterizes mechanistically the replication dynamics. On one hand, NS-seq provides sequence-specific mapping of the location of replication initiation sites complementing the SMARD data and also provided population-level measurements supplementing the selected number of molecules that can be analyzed by SMARD.[73] On the other hand, SMARD can show in high resolution the replication initiation and termination events in a single cell-derived single molecule, in addition to the replication fork directions.[57,58,74–78]

Genome-wide analysis determines that upon Tau expression, fewer initiation events are activated; nevertheless, it cannot distinguish between a scenario in which the distance between ORIs is longer or the alternative that fewer ORIs are active at a higher frequency. The SMARD analyses could distinguish between these two possibilities, indicating that several replication initiation events on the same fiber have an average distance of ∼50 kbp. The replication dynamics analysis revealed a strong usage of ORIs and long-travelling replication forks within the α-satellite sequences of both cenY and cen21, even under control conditions and in the parental cell line. However, wt Tau overexpression preserves a balanced regulation between replication initiation and pausing within Y α-satellite regions, while php Tau overexpression disrupts this balance, leading to unscheduled or compensatory ORI activation. cen21 exhibits increased sensitivity to replication perturbations induced by php Tau overexpression. Importantly, neither wt nor php Tau affects replication fork speed within centromeric DNA of ChrY and Chr21. These findings suggest that wt and php Tau differentially influence the origin activation and replication initiation, rather than affecting the fork progression within the centromeric DNA of both chromosomes. This distinction implies that the specific form of Tau may modulate the selection or accessibility of ORIs in these repetitive regions, potentially altering replication timing and genome stability.

This replication pattern underscores the unique replication landscape of centromeres, which is essential for the fidelity of chromosome segregation. These observations align with our further studies demonstrating that repetitive sequences often contain multiple replication initiation sites.[78–80] The replication initiation events in the first three repeats of the rDNA locus we studied in this work imply that repetitive sequences possess specific patterns of replication initiation events that span the entire locus.

When this delicate balance is disrupted by Tau overexpression, paused replication events occur at the α-satellites in several locations. NS-seq detected a decrease in the overall number of replication initiation sites at both cenY and cen21, specifically within the α-satellite sequences. At the cell population level, there were fewer potential replication initiation sites, although the frequency of initiation at specific sites increased with wt and php Tau overexpression.

Thus, Tau overexpression changed replication fork pausing and confined initiation activity to fewer sites within the repeats. The ability to resolve these alternatives provides a significant mechanistic advantage, showing how wt Tau and php Tau affect normal replication dynamics within the α-satellites of ChrY and Chr21 by increasing fork pausing and confining initiation activity to fewer loci.

This is further supported by the proteomics analysis that elucidated the role of wt and php Tau, revealing their differential protein interactions in the nucleus. php Tau appears to reshape its interaction landscape, with less and weaker connections with proteins involved in translation, chromatin remodeling, and intracellular transport. These changes likely reflect a shift that disrupts normal scaffolding role of Tau. The loss of ribosomal components and chromatin-associated interactors may compromise the gene regulation and protein synthesis, while reduced trafficking interactions could impair neuronal transport.

In conclusion, our findings underscore the role of Tau protein in modulating centromere replication, which is essential for its integrity and function in specific chromosome aneuploidies. Even if the replication dynamics observed vary across different chromosomes, focusing on ChrY and Chr21 presents significant findings relevant to AD. Brain mosaic aneuploidies can lead to a pathological physiology or death of the neural cells, making them reliable aging biomarkers for diagnosis and targeted therapies. Applying these techniques could support future studies of repetitive sequences in a chromosome- and allele-specific manner. This study offers a potential foundation for analyses aimed at identifying preventive targets in tauopathies and related disorders.

Limitations of the study

While this study provides mechanistic insight into how Tau overexpression alters centromeric replication in chromosomes 21 and Y, the findings are constrained by the focus on a single

---

(C) Venn diagram comparing ORI locations in doxycycline-treated CIMH-NP-3 parental, wt Tau, and php Tau cells. Cells overexpressing either wt Tau or php Tau showed fewer origins, with 30% of same origins shared under all conditions.

(D) Bar plot displaying the number of active ORIs in doxycycline-treated and untreated CIMH-NP-3 parental and wt Tau or php Tau cells, showing a 20%–25% reduction in ORIs after wt Tau or php Tau overexpression. Unpaired *t* test was used for statistical significance; not significant (ns), $^{**}p \leq 0.001$; $n = 2$ independent experiments; error bars represent the standard deviation (SD).

(E and F) Genome browser snapshots showing ORIs mapped on the centromeric regions (α-satellite and flanking regions used in the SMARD assay) of ChrY (E) and Chr21 (F). NS-seq reads were aligned to a custom reference generated from PacBio long-read sequencing (Figure 2) of CIMH-NP-3 ChrY (KpnI-digested) and Chr21 (BamHI-digested, haplotype 2). At the top of each chromosome ideogram, the blue lines mark the *in silico* probe location of Y α-satellites using DYZ3 Y-A and B primers and the 21 α-satellites using D13/21 ZL-3A and 1B primers. The right side shows a zoomed-in view of a single origin within the centromeric region.

(G) Violin plot of initiation strength expressed in $\log_2$ of read counts, measured by normalized reads per origin to genomic controls shown in (E) and (F). Statistical analysis was calculated using Wilcoxon signed-rank test; not significant (ns), $^{****}p \leq 0.0001$; $n = 2$ independent experiments. ORIs, origins of replication; NS-seq, nascent strand sequencing.

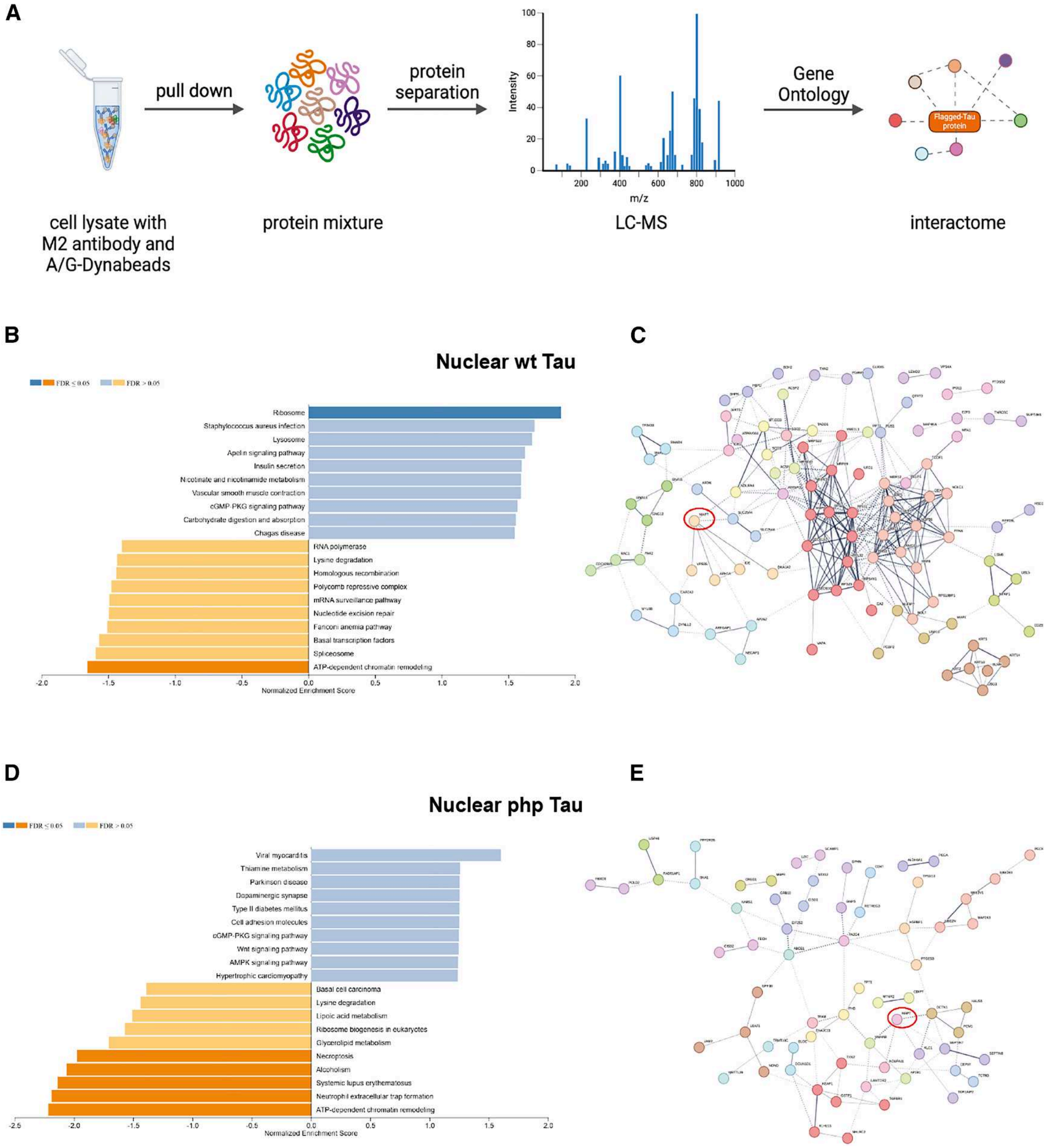


Figure 6. Quantitative proteomics of Tau and its interactors

(A) Procedure for the detection of Tau interactors: Tau-pull down, protein separation, LC-MS, and gene ontology search. Created in BioRender. Balzano, E. (2025) https://BioRender.com/c53t849.

(B–E) GSEA of the proteins enriched by the FLAG tag M2 antibody compared to the background IgG, performed using WebGestalt software. (B) Enrichment using cells with overexpressed wt Tau. (C) String analysis shows the wt Tau (MAPT) interactors in the nucleus, based on the molecular function. (D) Enrichment using cells with overexpressed php Tau. (E) String analysis shows the php Tau interactors in the nucleus, based on the molecular function. Statistical significance was assessed using heteroscedastic *t* test (if $p < 0.05$); $n = 2$ independent experiments. LC-MS, liquid chromatography-mass spectrometry; GSEA, gene set enrichment analysis.

cell type and two centromeres. The resolution of replication initiation site mapping was limited by the inherent complexity of repetitive DNA and the coverage achievable with single-molecule analysis. Although SMARD and NS-seq offer complementary strengths, the combined approaches cannot definitively capture all dynamic replication events at genome-wide level. The analysis of long SMARD molecules requires numerous and rigorous observations because only two-color molecules (IdU and CldU) with defined replication events were included in the study. The small number of labeled molecules (red and green) may present a challenge, but this limitation could be addressed by implementing automated approaches for image acquisition of SMARD molecules. Moreover, proteomic analyses, while informative, cannot distinguish direct from indirect interactions.

Further studies across additional genomic loci, cell types, and using orthogonal approaches will be needed to generalize these findings and fully elucidate the nuclear role of Tau. Finally, because the study used only male-derived cell lines, we could not determine whether sex-related biological differences influence Tau-dependent effects on centromeric replication.

RESOURCE AVAILABILITY

Lead contact

Requests for resources, reagents, and additional information should be directed to the lead contact, Carl L. Schildkraut (carl.schildkraut@einsteinmed.edu).

Materials availability

Further information regarding materials will be provided upon request by the lead contact, Carl L. Schildkraut (carl.schildkraut@einsteinmed.edu).

Data and code availability

- PacBio HiFi and ONT data generated from the CIMH-NP-3, CIMH-NP-3 wt Tau, and CIMH-NP-3 php Tau cell lines have been deposited to SRA in NCBI: PRJNA1181812.
- Whole-genome assemblies generated from the CIMH-NP-3, CIMH-NP-3 wt Tau, and CIMH-NP-3 php Tau cell lines and containing the complete sequences of the chromosome 21 and Y centromeres have been deposited to the NCBI:: PRJNA1181812, PRJNA1229460, PRJNA1229461, PRJNA1229462, and PRJNA1229463.
- Cell line details and original Western blot, Southern blot, and microscopy images have been deposited in Mendeley Data: https://doi.org/10.17632/6sfpcymkdk.1, https://doi.org/10.17632/v9fn28tvps.1, https://doi.org/10.17632/srwvmmyd58.1, and https://doi.org/10.17632/gmz4mrbyz4.1.
- Nascent strand sequencing data have been deposited to GEO in the NCBI, with the accession number GEO: GSE286931.
- Proteomics raw files have been deposited in PRIDE: PXD058752.
- Any data reported in this paper and any additional information required to reanalyze the data reported in this paper will be available from the lead contact upon request.

ACKNOWLEDGMENTS

This work was supported by National Institutes of Health (NIH) grants R01-GM045751 and R01-CA085344 to C.S.L. and to use of the core facility at AECOM, NIH Training Grant T-32 NIH T32AG023475 to S.T., NIH K99 Training Grant (GM147352) to G.A.L., NIH grant R01MH101221 to E.E.E., Intramural Research Program of the NIH, Center for Cancer Research, NCI (project ZIA BC010411) to M.I.A., Hevolution Foundation (AFAR), the Einstein-Mount Sinai Diabetes center, and the NIH Office of the Director (S10OD030286) to S.S., and Hector Stiftung II to P.K. and B.B. E.E.E. is an investigator of the Howard Hughes Medical Institute. Dr. Elisa Balzano carried out her studies for this publication while she was a postdoctoral fellow in the laboratory of Dr. C.L Schildkraut at Albert Einstein College of Medicine. The deconvoluted images were taken using the equipment of the AECOM Microscope Core. The mass spectrometry protein analysis was carried out in collaboration with the AECOM Proteomics Core Facility. We extend our gratitude to the Stengel laboratory for their invaluable assistance with equipment, reagents, and guidance. We also acknowledge Dr. Eric Bouhassira, Dr Arthur I. Skoultchi, and Dr Dong Zhang for their pivotal discussions; Tonia Brown for proofreading the manuscript; and the Capecchi laboratory for the support.

AUTHOR CONTRIBUTIONS

Conceptualization, C.L.S., S.T., and E.B.; methodology and investigation, S.T., S.T.K., E.B., G.A.L., B.B, K.M.M., K.H., A.K., B.L.T., and S.S.; writing – original draft, E.B.; writing – review & editing, all authors; funding acquisition, C.L.S., G.A.L., E.E.E., M.I.A., S.S., and P.K.; resources, C.L.S., G.A.L., E.E.E., M.I.A., S.S., and P.K.; supervision, C.L.S. and E.B.

DECLARATION OF INTERESTS

E.E.E. is a scientific advisory board (SAB) member of Variant Bio, Inc.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
  - ○ Cell lines
- METHOD DETAILS
  - ○ Plasmid vector design and lentiviral transduction construction
  - ○ Cell culture
  - ○ Immunoblot for verification of CIMH-NP-3 cell lines
  - ○ Immunofluorescence verification of CIMH-NP-3 cell lines
  - ○ Monoclonal cell population stabilization and Immunoblotting
  - ○ DNA extraction, library preparation, and sequencing
  - ○ Sequence assembly and validation of centromeric regions
  - ○ Single Molecule Analysis of Replicated DNA (SMARD)
  - ○ DNA fibers SMARD analysis
  - ○ Two color fiber-FISH
  - ○ Photomicrograph image acquisition for rDNA
  - ○ Fluorescence *in situ* hybridization with chromosome paint probes
  - ○ Nascent strand data analysis sequencing
  - ○ Nascent strand sequence analyses
  - ○ S-trap protein digestion
  - ○ Sample desalting
  - ○ LC-MS/MS acquisition and analysis
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2026.114707.

Received: April 7, 2025
Revised: November 8, 2025
Accepted: January 12, 2026
Published: January 15, 2026

REFERENCES

1. Iourov, I.Y., Yurov, Y.B., Vorsanova, S.G., and Kutsev, S.I. (2021). Chromosome Instability, Aging and Brain Diseases. Cells *10*, 1256. https://doi.org/10.3390/cells10051256.
2. Sdeor, E., Okada, H., Saad, R., Ben-Yishay, T., and Ben-David, U. (2024). Aneuploidy as a driver of human cancer. Nat. Genet. *56*, 2014–2026. https://doi.org/10.1038/s41588-024-01916-2.
3. Willard, H.F., and Waye, J.S. (1987). Chromosome-specific subsets of human alpha satellite DNA: analysis of sequence divergence within and between chromosomal subsets and evidence for an ancestral pentameric repeat. J. Mol. Evol. *25*, 207–214. https://doi.org/10.1007/BF02100014.
4. Willard, H.F. (1989). The genomics of long tandem arrays of satellite DNA in the human genome. Genome *31*, 737–744. https://doi.org/10.1139/g89-132.
5. Wu, J.S., Giuffra, L.A., Goodfellow, P.J., Myers, S., Carson, N.L., Anderson, L., Hoyle, L.S., Simpson, N.E., and Kidd, K.K. (1989). The beta subunit locus of the human fibronectin receptor: DNA restriction fragment length polymorphism and linkage mapping studies. Hum. Genet. *83*, 383–390. https://doi.org/10.1007/BF00291386.
6. Devilee, P., Kievits, T., Waye, J.S., Pearson, P.L., and Willard, H.F. (1988). Chromosome-specific alpha satellite DNA: Isolation and mapping of a polymorphic alphoid repeat from human chromosome 10. Genomics *3*, 1–7. https://doi.org/10.1016/0888-7543(88)90151-6.

7. Balzano, E., and Giunta, S. (2020). Centromeres under Pressure: Evolutionary Innovation in Conflict with Conserved Function. Genes *11*, 912. https://doi.org/10.3390/genes11080912.
8. Logsdon, G.A., Rozanski, A.N., Ryabov, F., Potapova, T., Shepelev, V.A., Catacchio, C.R., Porubsky, D., Mao, Y., Yoo, D., Rautiainen, M., et al. (2024). The variation and evolution of complete human centromeres. Nature *629*, 136–145. https://doi.org/10.1038/s41586-024-07278-3.
9. Janssen, A., Colmenares, S.U., and Karpen, G.H. (2018). Heterochromatin: Guardian of the Genome. Annu. Rev. Cell Dev. Biol. *34*, 265–288. https://doi.org/10.1146/annurev-cellbio-100617-062653.
10. Miga, K.H., Koren, S., Rhie, A., Vollger, M.R., Gershman, A., Bzikadze, A., Brooks, S., Howe, E., Porubsky, D., Logsdon, G.A., et al. (2020). Telomere-to-telomere assembly of a complete human X chromosome. Nature *585*, 79–84. https://doi.org/10.1038/s41586-020-2547-7.
11. Nurk, S., Koren, S., Rhie, A., Rautiainen, M., Bzikadze, A.V., Mikheenko, A., Vollger, M.R., Altemose, N., Uralsky, L., Gershman, A., et al. (2022). The complete sequence of a human genome. Science *376*, 44–53. https://doi.org/10.1126/science.abj6987.
12. Logsdon, G.A., Vollger, M.R., Hsieh, P., Mao, Y., Liskovykh, M.A., Koren, S., Nurk, S., Mercuri, L., Dishuck, P.C., Rhie, A., et al. (2021). The structure, function and evolution of a complete human chromosome 8. Nature *593*, 101–107. https://doi.org/10.1038/s41586-021-03420-7.
13. Barra, V., and Fachinetti, D. (2018). The dark side of centromeres: types, causes and consequences of structural abnormalities implicating centromeric DNA. Nat. Commun. *9*, 4340. https://doi.org/10.1038/s41467-018-06545-y.
14. Giunta, S., and Funabiki, H. (2017). Integrity of the human centromere DNA repeats is protected by CENP-A, CENP-C, and CENP-T. Proc. Natl. Acad. Sci. USA *114*, 1928–1933. https://doi.org/10.1073/pnas.1615133114.
15. Naylor, R.M., and van Deursen, J.M. (2016). Aneuploidy in Cancer and Aging. Annu. Rev. Genet. *50*, 45–66. https://doi.org/10.1146/annurev-genet-120215-035303.
16. Chiang, T., Duncan, F.E., Schindler, K., Schultz, R.M., and Lampson, M.A. (2010). Evidence that weakened centromere cohesion is a leading cause of age-related aneuploidy in oocytes. Curr. Biol. *20*, 1522–1528. https://doi.org/10.1016/j.cub.2010.06.069.
17. Miga, K.H., and Sullivan, B.A. (2021). Expanding studies of chromosome structure and function in the era of T2T genomics. Hum. Mol. Genet. *30*, R198–R205. https://doi.org/10.1093/hmg/ddab214.
18. Sullivan, L.L., and Sullivan, B.A. (2020). Genomic and functional variation of human centromeres. Exp. Cell Res. *389*, 111896. https://doi.org/10.1016/j.yexcr.2020.111896.
19. Neil, A.J., Kim, J.C., and Mirkin, S.M. (2017). Precarious maintenance of simple DNA repeats in eukaryotes. Bioessays *39*, 1700077. https://doi.org/10.1002/bies.201700077.
20. Danielsson, M., Halvardson, J., Davies, H., Torabi Moghadam, B., Mattisson, J., Rychlicka-Buniowska, E., Jaszczyński, J., Heintz, J., Lannfelt, L., Giedraitis, V., et al. (2020). Longitudinal changes in the frequency of mosaic chromosome Y loss in peripheral blood cells of aging men varies profoundly between individuals. Eur. J. Hum. Genet. *28*, 349–357. https://doi.org/10.1038/s41431-019-0533-z.
21. Mendivil-Perez, M., Velez-Pardo, C., Kosik, K.S., Lopera, F., and Jimenez-Del-Rio, M. (2019). iPSCs-derived nerve-like cells from familial Alzheimer's disease PSEN 1 E280A reveal increased amyloid-beta levels and loss of the Y chromosome. Neurosci. Lett. *703*, 111–118. https://doi.org/10.1016/j.neulet.2019.03.032.
22. Guo, X., Dai, X., Zhou, T., Wang, H., Ni, J., Xue, J., and Wang, X. (2020). Mosaic loss of human Y chromosome: what, how and why. Hum. Genet. *139*, 421–446. https://doi.org/10.1007/s00439-020-02114-w.
23. Cáceres, A., Jene, A., Esko, T., Pérez-Jurado, L.A., and González, J.R. (2020). Extreme Downregulation of Chromosome Y and Cancer Risk in Men. J. Natl. Cancer Inst. *112*, 913–920. https://doi.org/10.1093/jnci/djz232.
24. Vermeulen, M.C., Pearse, R., Young-Pearse, T., and Mostafavi, S. (2022). Mosaic loss of Chromosome Y in aged human microglia. Genome Res. *32*, 1795–1807. https://doi.org/10.1101/gr.276409.121.
25. García-González, P., de Rojas, I., Moreno-Grau, S., Montrreal, L., Puerta, R., Alarcón-Martín, E., Quintela, I., Orellana, A., Andrade, V., Adami, P.V.M., et al. (2023). Mendelian Randomisation Confirms the Role of Y-Chromosome Loss in Alzheimer's Disease Aetiopathogenesis in Men. Int. J. Mol. Sci. *24*, 898. https://doi.org/10.3390/ijms24020898.
26. Lopez-Lee, C., Torres, E.R.S., Carling, G., and Gan, L. (2024). Mechanisms of sex differences in Alzheimer's disease. Neuron *112*, 1208–1221. https://doi.org/10.1016/j.neuron.2024.01.024.
27. Dumanski, J.P., Lambert, J.-C., Rasi, C., Giedraitis, V., Davies, H., Grenier-Boley, B., Lindgren, C.M., Campion, D., Dufouil, C., Florence, P., et al. (2016). Mosaic Loss of Chromosome Y in Blood Is Associated with Alzheimer Disease. Am. J. Hum. Genet. *98*, 1208–1219. https://doi.org/10.1016/j.ajhg.2016.05.014.
28. Potter, H. (1991). Review and hypothesis: Alzheimer disease and Down syndrome–chromosome 21 nondisjunction may underlie both disorders. Am. J. Hum. Genet. *48*, 1192–1200.
29. Geller, L.N., and Potter, H. (1999). Chromosome missegregation and trisomy 21 mosaicism in Alzheimer's disease. Neurobiol. Dis. *6*, 167–179. https://doi.org/10.1006/nbdi.1999.0236.
30. Yang, Y., Geldmacher, D.S., and Herrup, K. (2001). DNA replication precedes neuronal cell death in Alzheimer's disease. J. Neurosci. *21*, 2661–2668. https://doi.org/10.1523/JNEUROSCI.21-08-02661.2001.
31. Kingsbury, M.A., Yung, Y.C., Peterson, S.E., Westra, J.W., and Chun, J. (2006). Aneuploidy in the normal and diseased brain. Cell. Mol. Life Sci. *63*, 2626–2641. https://doi.org/10.1007/s00018-006-6169-5.
32. Mosch, B., Morawski, M., Mittag, A., Lenz, D., Tarnok, A., and Arendt, T. (2007). Aneuploidy and DNA replication in the normal human brain and Alzheimer's disease. J. Neurosci. *27*, 6859–6867. https://doi.org/10.1523/JNEUROSCI.0379-07.2007.
33. Thomas, P., and Fenech, M. (2008). Chromosome 17 and 21 aneuploidy in buccal cells is increased with ageing and in Alzheimer's disease. Mutagenesis *23*, 57–65. https://doi.org/10.1093/mutage/gem044.
34. Iourov, I.Y., Vorsanova, S.G., Liehr, T., and Yurov, Y.B. (2009). Aneuploidy in the normal, Alzheimer's disease and ataxia-telangiectasia brain: differential expression and pathological meaning. Neurobiol. Dis. *34*, 212–220. https://doi.org/10.1016/j.nbd.2009.01.003.
35. Arendt, T., Brückner, M.K., Mosch, B., and Lösche, A. (2010). Selective cell death of hyperploid neurons in Alzheimer's disease. Am. J. Pathol. *177*, 15–20. https://doi.org/10.2353/ajpath.2010.090955.
36. Iourov, I.Y., Vorsanova, S.G., and Yurov, Y.B. (2011). Genomic Landscape of the Alzheimer's Disease Brain: Chromosome Instability – Aneuploidy, but Not Tetraploidy – Mediates Neurodegeneration. Neurodegener. Dis. *8*, 35–40. https://doi.org/10.1159/000315398.
37. Granic, A., Padmanabhan, J., Norden, M., and Potter, H. (2010). Alzheimer Aβ Peptide Induces Chromosome Mis-Segregation and Aneuploidy, Including Trisomy 21: Requirement for Tau and APP. Mol. Biol. Cell *21*, 511–520. https://doi.org/10.1091/mbc.e09-10-0850.
38. Kixmoeller, K., Allu, P.K., and Black, B.E. (2020). The centromere comes into focus: from CENP-A nucleosomes to kinetochore connections with the spindle. Open Biol. *10*, 200051. https://doi.org/10.1098/rsob.200051.
39. Hoffmann, S., Dumont, M., Barra, V., Ly, P., Nechemia-Arbely, Y., McMahon, M.A., Hervé, S., Cleveland, D.W., and Fachinetti, D. (2016). CENP-A Is Dispensable for Mitotic Centromere Function after Initial Centromere/Kinetochore Assembly. Cell Rep. *17*, 2394–2404. https://doi.org/10.1016/j.celrep.2016.10.084.
40. Earnshaw, W.C., and Rothfield, N. (1985). Identification of a family of human centromere proteins using autoimmune sera from patients with scleroderma. Chromosoma *91*, 313–321. https://doi.org/10.1007/BF00328227.

41. Masumoto, H., Masukata, H., Muro, Y., Nozaki, N., and Okazaki, T. (1989). A human centromere antigen (CENP-B) interacts with a short specific sequence in alphoid DNA, a human centromeric satellite. J. Cell Biol. *109*, 1963–1973. https://doi.org/10.1083/jcb.109.5.1963.
42. Fachinetti, D., Han, J.S., McMahon, M.A., Ly, P., Abdullah, A., Wong, A.J., and Cleveland, D.W. (2015). DNA Sequence-Specific Binding of CENP-B Enhances the Fidelity of Human Centromere Function. Dev. Cell *33*, 314–327. https://doi.org/10.1016/j.devcel.2015.03.020.
43. Gutiérrez-Hurtado, I.A., Sánchez-Méndez, A.D., Becerra-Loaiza, D.S., Rangel-Villalobos, H., Torres-Carrillo, N., Gallegos-Arreola, M.P., and Aguilar-Velázquez, J.A. (2024). Loss of the Y Chromosome: A Review of Molecular Mechanisms, Age Inference, and Implications for Men's Health. Int. J. Mol. Sci. *25*, 4230. https://doi.org/10.3390/ijms25084230.
44. Oliver, C., and Holland, A.J. (1986). Down's syndrome and Alzheimer's disease: a review. Psychol. Med. *16*, 307–322. https://doi.org/10.1017/s0033291700009120.
45. Holland, A.J., Hon, J., Huppert, F.A., Stevens, F., and Watson, P. (1998). Population-based study of the prevalence and presentation of dementia in adults with Down's syndrome. Br. J. Psychiatry *172*, 493–498. https://doi.org/10.1192/bjp.172.6.493.
46. Hampel, H., Hardy, J., Blennow, K., Chen, C., Perry, G., Kim, S.H., Villemagne, V.L., Aisen, P., Vendruscolo, M., Iwatsubo, T., et al. (2021). The Amyloid-β Pathway in Alzheimer's Disease. Mol. Psychiatry *26*, 5481–5503. https://doi.org/10.1038/s41380-021-01249-0.
47. Sjöberg, M.K., Shestakova, E., Mansuroglu, Z., Maccioni, R.B., and Bonnefoy, E. (2006). Tau protein binds to pericentromeric DNA: a putative role for nuclear tau in nucleolar organization. J. Cell Sci. *119*, 2025–2034. https://doi.org/10.1242/jcs.02907.
48. Park, C., Jung, J., Hong, Y., Lee-Eom, C., Rah, S.-H., Yang, K., Shin, J., Jeong, A., Hong, S., Lee, J.-B., et al. (2025). BPS2025 - Tau condensation on DNA and localization on centromeres: A potential link to cell division. Biophys. J. *124*, 86a. https://doi.org/10.1016/j.bpj.2024.11.526.
49. Rico, T., Gilles, M., Chauderlier, A., Comptdaer, T., Magnez, R., Chwastyniak, M., Drobecq, H., Pinet, F., Thuru, X., Buée, L., et al. (2021). Tau Stabilizes Chromatin Compaction. Front. Cell Dev. Biol. *9*, 740550. https://doi.org/10.3389/fcell.2021.740550.
50. Mansuroglu, Z., Benhelli-Mokrani, H., Marcato, V., Sultan, A., Violet, M., Chauderlier, A., Delattre, L., Loyens, A., Talahari, S., Bégard, S., et al. (2016). Loss of Tau protein affects the structure, transcription and repair of neuronal pericentromeric heterochromatin. Sci. Rep. *6*, 33047. https://doi.org/10.1038/srep33047.
51. Rossi, G., Dalprà, L., Crosti, F., Lissoni, S., Sciacca, F.L., Catania, M., Di Fede, G., Mangieri, M., Giaccone, G., Croci, D., and Tagliavini, F. (2008). A new function of microtubule-associated protein tau: Involvement in chromosome stability. Cell Cycle *7*, 1788–1794. https://doi.org/10.4161/cc.7.12.6012.
52. Arendt, T., Stieler, J.T., and Holzer, M. (2016). Tau and tauopathies. Brain Res. Bull. *126*, 238–292. https://doi.org/10.1016/j.brainresbull.2016.08.018.
53. Potter, H., Chial, H.J., Caneus, J., Elos, M., Elder, N., Borysov, S., and Granic, A. (2019). Chromosome Instability and Mosaic Aneuploidy in Neurodegenerative and Neurodevelopmental Disorders. Front. Genet. *10*, 1092. https://doi.org/10.3389/fgene.2019.01092.
54. Mertens, J., Stüber, K., Poppe, D., Doerr, J., Ladewig, J., Brüstle, O., and Koch, P. (2013). Embryonic stem cell-based modeling of tau pathology in human neurons. Am. J. Pathol. *182*, 1769–1779. https://doi.org/10.1016/j.ajpath.2013.01.043.
55. Moreno-Jiménez, E.P., Terreros-Roncal, J., Flor-García, M., Rábano, A., and Llorens-Martín, M. (2021). Evidences for Adult Hippocampal Neurogenesis in Humans. J. Neurosci. *41*, 2541–2553. https://doi.org/10.1523/JNEUROSCI.0675-20.2020.
56. Dumitru, I., Paterlini, M., Zamboni, M., Ziegenhain, C., Giatrellis, S., Saghaleyni, R., Björklund, Å., Alkass, K., Tata, M., Druid, H., et al. (2025). Identification of proliferating neural progenitors in the adult human hippocampus. Science *389*, 58–63. https://doi.org/10.1126/science.adu9575.
57. Norio, P., and Schildkraut, C.L. (2004). Plasticity of DNA replication initiation in Epstein-Barr virus episomes. PLoS Biol. *2*, e152. https://doi.org/10.1371/journal.pbio.0020152.
58. Norio, P., and Schildkraut, C.L. (2001). Visualization of DNA Replication on Individual Epstein-Barr Virus Episomes. Science *294*, 2361–2364. https://doi.org/10.1126/science.1064603.
59. Rhie, A., Nurk, S., Cechova, M., Hoyt, S.J., Taylor, D.J., Altemose, N., Hook, P.W., Koren, S., Rautiainen, M., Alexandrov, I.A., et al. (2023). The complete sequence of a human Y chromosome. Nature *621*, 344–354. https://doi.org/10.1038/s41586-023-06457-y.
60. Bzikadze, A.V., Mikheenko, A., and Pevzner, P.A. (2022). Fast and accurate mapping of long reads to complete genome assemblies with VerityMap. Genome Res. *32*, 2107–2118. https://doi.org/10.1101/gr.276871.122.
61. Dishuck, P.C., Rozanski, A.N., Logsdon, G.A., Porubsky, D., and Eichler, E.E. (2023). GAVISUNK: genome assembly validation via inter-SUNK distances in Oxford Nanopore reads. Bioinformatics *39*, btac714. https://doi.org/10.1093/bioinformatics/btac714.
62. Agrawal, S., and Ganley, A.R.D. (2018). The conservation landscape of the human ribosomal RNA gene repeats. PLoS One *13*, e0207531. https://doi.org/10.1371/journal.pone.0207531.
63. Giacca, M., Zentilin, L., Norio, P., Diviacco, S., Dimitrova, D., Contreas, G., Biamonti, G., Perini, G., Weighardt, F., and Riva, S. (1994). Fine mapping of a replication origin of human DNA. Proc. Natl. Acad. Sci. USA *91*, 7119–7123. https://doi.org/10.1073/pnas.91.15.7119.
64. Yoon, Y., Sanchez, J.A., Brun, C., and Huberman, J.A. (1995). Mapping of Replication Initiation Sites in Human Ribosomal DNA by Nascent-Strand Abundance Analysis. Mol. Cell Biol. *15*, 2482–2489. https://doi.org/10.1128/MCB.15.5.2482.
65. Bielinsky, A.-K., and Gerbi, S.A. (1998). Discrete Start Sites for DNA Synthesis in the Yeast *ARS1* Origin. Science *279*, 95–98. https://doi.org/10.1126/science.279.5347.95.
66. Aladjem, M.I., Groudine, M., Brody, L.L., Dieken, E.S., Fournier, R.E., Wahl, G.M., and Epner, E.M. (1995). Participation of the Human β-Globin Locus Control Region in Initiation of DNA Replication. Science *270*, 815–819. https://doi.org/10.1126/science.270.5237.815.
67. Martin, M.M., Ryan, M., Kim, R., Zakas, A.L., Fu, H., Lin, C.M., Reinhold, W.C., Davis, S.R., Bilke, S., Liu, H., et al. (2011). Genome-wide depletion of replication initiation events in highly transcribed regions. Genome Res. *21*, 1822–1832. https://doi.org/10.1101/gr.124644.111.
68. Thakur, B.L., Baris, A.M., Fu, H., Redon, C.E., Pongor, L.S., Mosavarpour, S., Gross, J.M., Jang, S.-M., Sebastian, R., Utani, K., et al. (2022). Convergence of SIRT1 and ATR signaling to modulate replication origin dormancy. Nucleic Acids Res. *50*, 5111–5128. https://doi.org/10.1093/nar/gkac299.
69. Thakur, B.L., Redon, C.E., Fu, H., Sebastian, R., Kusi, N.A., Zhuang, S.Z., Pongor, L.S., Bohr, V.A., and Aladjem, M.I. (2025). Selective interactions at pre-replication complexes categorize baseline and dormant origins. Nat. Commun. *16*, 4140. https://doi.org/10.1038/s41467-025-59509-4.
70. Mueller, R.L., Combs, B., Alhadidy, M.M., Brady, S.T., Morfini, G.A., and Kanaan, N.M. (2021). Tau: A Signaling Hub Protein. Front. Mol. Neurosci. *14*, 647054. https://doi.org/10.3389/fnmol.2021.647054.
71. Parra Bravo, C., Naguib, S.A., and Gan, L. (2024). Cellular and pathological functions of tau. Nat. Rev. Mol. Cell Biol. *25*, 845–864. https://doi.org/10.1038/s41580-024-00753-9.
72. Aguilan, J.T., Kulej, K., and Sidoli, S. (2020). Guide for protein fold change and p-value calculation for non-experts in proteomics. Mol. Omics *16*, 573–582. https://doi.org/10.1039/d0mo00087f.
73. Aladjem, M.I., and Redon, C.E. (2017). Order from clutter: selective interactions at mammalian replication origins. Nat. Rev. Genet. *18*, 101–116. https://doi.org/10.1038/nrg.2016.141.

74. Twayana, S., Bacolla, A., Barreto-Galvez, A., De-Paula, R.B., Drosopoulos, W.C., Kosiyatrakul, S.T., Bouhassira, E.E., Tainer, J.A., Madireddy, A., and Schildkraut, C.L. (2021). Translesion polymerase eta both facilitates DNA replication and promotes increased human genetic variation at common fragile sites. Proc. Natl. Acad. Sci. USA *118*, e2106477118. https://doi.org/10.1073/pnas.2106477118.

75. Madireddy, A., Kosiyatrakul, S.T., Boisvert, R.A., Herrera-Moyano, E., García-Rubio, M.L., Gerhardt, J., Vuono, E.A., Owen, N., Yan, Z., Olson, S., et al. (2016). FANCD2 Facilitates Replication through Common Fragile Sites. Mol. Cell *64*, 388–404. https://doi.org/10.1016/j.molcel.2016.09.017.

76. Drosopoulos, W.C., Deng, Z., Twayana, S., Kosiyatrakul, S.T., Vladimirova, O., Lieberman, P.M., and Schildkraut, C.L. (2020). TRF2 Mediates Replication Initiation within Human Telomeres to Prevent Telomere Dysfunction. Cell Rep. *33*, 108379. https://doi.org/10.1016/j.celrep.2020.108379.

77. Schultz, S.S., Desbordes, S.C., Du, Z., Kosiyatrakul, S., Lipchina, I., Studer, L., and Schildkraut, C.L. (2010). Single-Molecule Analysis Reveals Changes in the DNA Replication Program for the POU5F1 Locus upon Human Embryonic Stem Cell Differentiation. Mol. Cell Biol. *30*, 4521–4534. https://doi.org/10.1128/MCB.00380-10.

78. Little, R.D., Platt, T.H., and Schildkraut, C.L. (1993). Initiation and termination of DNA replication in human rRNA genes. Mol. Cell Biol. *13*, 6600–6613. https://doi.org/10.1128/mcb.13.10.6600-6613.1993.

79. Thakur, B.L., Kusi, N.A., Mosavarpour, S., Zhu, R., Redon, C.E., Fu, H., Dhall, A., Pongor, L.S., Sebastian, R., Indig, F.E., and Aladjem, M.I. (2023). SIRT1 Prevents R-Loops during Chronological Aging by Modulating DNA Replication at rDNA Loci. Cells *12*, 2630. https://doi.org/10.3390/cells12222630.

80. Erliandri, I., Fu, H., Nakano, M., Kim, J.-H., Miga, K.H., Liskovykh, M., Earnshaw, W.C., Masumoto, H., Kouprina, N., Aladjem, M.I., et al. (2014). Replication of alpha-satellite DNA arrays in endogenous human centromeric regions and in human artificial chromosome. Nucleic Acids Res. *42*, 11502–11516. https://doi.org/10.1093/nar/gku835.

81. Warburton, P.E., Greig, G.M., Haaf, T., and Willard, H.F. (1991). PCR amplification of chromosome-specific alpha satellite DNA: Definition of centromeric STS markers and polymorphic analysis. Genomics *11*, 324–333. https://doi.org/10.1016/0888-7543(91)90139-6.

82. Rautiainen, M., Nurk, S., Walenz, B.P., Logsdon, G.A., Porubsky, D., Rhie, A., Eichler, E.E., Phillippy, A.M., and Koren, S. (2023). Telomere-to-telomere assembly of diploid chromosomes with Verkko. Nat. Biotechnol. *41*, 1474–1482. https://doi.org/10.1038/s41587-023-01662-6.

83. Cheng, H., Concepcion, G.T., Feng, X., Zhang, H., and Li, H. (2021). Haplotype-resolved *de novo* assembly using phased assembly graphs with hifiasm. Nat. Methods *18*, 170–175. https://doi.org/10.1038/s41592-020-01056-5.

84. Li, H. (2018). Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics *34*, 3094–3100. https://doi.org/10.1093/bioinformatics/bty191.

85. Marçais, G., and Kingsford, C. (2011). A fast, lock-free approach for efficient parallel counting of occurrences of *k* -mers. Bioinformatics *27*, 764–770. https://doi.org/10.1093/bioinformatics/btr011.

86. Jain, C., Rhie, A., Zhang, H., Chu, C., Walenz, B.P., Koren, S., and Phillippy, A.M. (2020). Weighted minimizer sampling improves long read mapping. Bioinformatics *36*, i111–i118. https://doi.org/10.1093/bioinformatics/btaa435.

87. Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., and Mesirov, J.P. (2011). Integrative genomics viewer. Nat. Biotechnol. *29*, 24–26. https://doi.org/10.1038/nbt.1754.

88. Vollger, M.R., Dishuck, P.C., Sorensen, M., Welch, A.E., Dang, V., Dougherty, M.L., Graves-Lindsay, T.A., Wilson, R.K., Chaisson, M.J.P., and Eichler, E.E. (2019). Long-read sequence and assembly of segmental duplications. Nat. Methods *16*, 88–94. https://doi.org/10.1038/s41592-018-0236-3.

89. Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics *30*, 2114–2120. https://doi.org/10.1093/bioinformatics/btu170.

90. Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics *25*, 1754–1760. https://doi.org/10.1093/bioinformatics/btp324.

91. Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841–842. https://doi.org/10.1093/bioinformatics/btq033.

92. Ramírez, F., Ryan, D.P., Grüning, B., Bhardwaj, V., Kilpert, F., Richter, A.S., Heyne, S., Dündar, F., and Manke, T. (2016). deepTools2: a next generation web server for deep-sequencing data analysis. Nucleic Acids Res. *44*, W160–W165. https://doi.org/10.1093/nar/gkw257.

93. Wickham, H. (2016). ggplot2 (Springer International Publishing).

94. Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., et al. (2005). Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl. Acad. Sci. USA *102*, 15545–15550. https://doi.org/10.1073/pnas.0506580102.

95. Mootha, V.K., Lindgren, C.M., Eriksson, K.-F., Subramanian, A., Sihag, S., Lehar, J., Puigserver, P., Carlsson, E., Ridderstråle, M., Laurila, E., et al. (2003). PGC-1α-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. Nat. Genet. *34*, 267–273. https://doi.org/10.1038/ng1180.

96. Dirks, W.G., MacLeod, R.A.F., Nakamura, Y., Kohara, A., Reid, Y., Milch, H., Drexler, H.G., and Mizusawa, H. (2010). Cell line cross-contamination initiative: An interactive reference database of STR profiles covering common cancer cell lines. Int. J. Cancer *126*, 303–304. https://doi.org/10.1002/ijc.24999.

97. Reinhardt, P., Glatza, M., Hemmer, K., Tsytsyura, Y., Thiel, C.S., Höing, S., Moritz, S., Parga, J.A., Wagner, L., Bruder, J.M., et al. (2013). Derivation and Expansion Using Only Small Molecules of Human Neural Progenitors for Neurodegenerative Disease Modeling. PLoS One *8*, e59252. https://doi.org/10.1371/journal.pone.0059252.

98. Jain, M., Koren, S., Miga, K.H., Quick, J., Rand, A.C., Sasani, T.A., Tyson, J.R., Beggs, A.D., Dilthey, A.T., Fiddes, I.T., et al. (2018). Nanopore sequencing and assembly of a human genome with ultra-long reads. Nat. Biotechnol. *36*, 338–345. https://doi.org/10.1038/nbt.4060.

99. Warburton, A., Redmond, C.J., Dooley, K.E., Fu, H., Gillison, M.L., Akagi, K., Symer, D.E., Aladjem, M.I., and McBride, A.A. (2018). HPV integration hijacks and multimerizes a cellular enhancer to generate a viral-cellular super-enhancer that drives high viral oncogene expression. PLoS Genet. *14*, e1007179. https://doi.org/10.1371/journal.pgen.1007179.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
mouse monoclonal FLAG	Sigma-Aldrich	Cat#F1804; RRID: AB_262044
rabbit FLAG	Cell Signaling Technology	Cat#2368S; RRID: AB_2217020
mouse Nestin	R&D Systems	Cat#MAB-1259; RRID: B_2251304
rabbit Nestin	Abcam	Cat#ab27952; RRID: AB_776698
rabbit Sox2	Cell Signaling Technology	Cat#3579; RRID: AB_2195767
mouse Actin	Cell Signaling Technology	Cat#3700; RRID: AB_2242334
rabbit polyclonal Actin	Sigma-Aldrich	Cat#A2066; RRID: AB_476693
guinea pig Tau	Synaptic Systems	Cat#314 004; RRID: AB_1547385
human monoclonal Tau	Santa Cruz	Cat#Sc-58860; RRID: AB_785931
guinea pig Tubb3	Synaptic Systems	Cat#302 304; RRID: AB_10805138
Avidin Alexa Flour 568	Invitrogen	Cat#S11226; RRID: AB_2315774
Anti-digoxigenin, FITC conjugated	Roche	Cat#11207741910; RRID: AB_514498
Secondary antibodies: Biotinylated anti-avidin D	Vector Laboratories	Cat#BA-0300; RRID: AB_2336108
Secondary antibodies: anti-FITC	Invitrogen	Cat#71-1900; RRID: AB_2533978
Secondary antibodies: anti-rabbit FITC conjugates	Rockland	Cat#611-102-122; RRID: AB_2340612
Secondary antibodies: anti-rabbit DyLight800	Cell Signaling Technology	Cat#5151; RRID: AB_2687826
Secondary antibodies: anti-mouse DyLight680	Cell Signaling Technology	Cat#5470; RRID: AB_10696895
Secondary antibodies: AF488-conjugated (rabbit)	Invitrogen	Cat#A-11008; RRID: AB_143165
Secondary antibodies: AF555-conjugated (guinea pig)	Invitrogen	Cat#A-21435; RRID: AB_2535856
Secondary antibodies: AF647-conjugated (mouse)	Invitrogen	Cat#A-21236; RRID: AB_2535805
Secondary antibodies: Goat Anti-rabbit IRDye800CW	LI-COR Biosciences	Cat#926-32211; RRID: AB_1850025
Secondary antibodies: Goat Anti-human IRDye800CW	LI-COR Biosciences	Cat#926-32232; RRID: AB_6858
Bacterial and virus strains		
Sendai virus (CytoTune-iPS Sendai Reprogramming Kit)	Invitrogen	Cat#A16517
Biological samples		
Dermal fibroblasts (healthy male donor)	University of Bonn, Central Institute of Mental Health in Mannheim	N/A
Chemicals, peptides, and recombinant proteins		
Doxycycline (DOX)	Sigma-Aldrich	Cat#D9891
LDN-193189	REPROCELL / Stemgent	Cat#04-0074-10
A83	REPROCELL / Stemgent	Cat#04-0014-10
Geltrex	Invitrogen	Cat#A1413201
DMEM/F12	Invitrogen	Cat#11320074
N2 supplement	Invitrogen	Cat#17502048
B27 supplement	Invitrogen	Cat#17504044; 12587010
GlutaMAX	Invitrogen	Cat#35050038

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Glucose	Roth & Sigma-Aldrich	Cat#HN06.2; G7021
NEAA (non-essential amino acids)	Invitrogen	Cat#11140035
CHIR99021 (GSK3β inhibitor)	Cell guidance systems	Cat#SM13
L-glutamine	Invitrogen	Cat#25030081
L-ascorbic acid	Sigma-Aldrich	Cat#A4544
G418	Invitrogen	Cat#10131035
Puromycin	Sigma-Aldrich	Cat#P7255
Purmorphamine (Smo activator)	Cell guidance systems	Cat#SM30
Fetal bovine serum	Invitrogen	Cat#10270106
Prolong gold antifade reagent	Invitrogen	Cat#P36930
3-aminopropyltriethoxysilane	Sigma-Aldrich	Cat#99 919-30-2
RNase A	Qiagen	Cat#19101
Proteinase K	Qiagen	Cat#19131
DAPI ProLong^{TM} Gold Antifade Mountant	ThermoFisher	Cat#P36941
Precision Plus Protein Dual Color Standards	ThermoFisher	Cat#1610374
Critical commercial assays		
Pierce BCA Protein Assay Kit	Thermo Fisher	Cat#23225
Monarch HMW DNA Extraction Kit	New England Biolabs	Cat#T3050L
SMRTbell Prep Kit 3.0	PacBio	Cat#102-141-700
ONT Ultra-Long DNA Sequencing Kit	Oxford Nanopore Technologies	Cat#SQK-ULK001
Nanobind UL Library Prep Kit	Circulomics	Cat#NB-900-601-01
RadPrime DNA Labeling System	Invitrogen	Cat#18428-011
BioPrime DNA Labeling System	N/A	Cat#18094-011
DIG DNA Labeling Mix	Roche	Cat#11277065910
BioNick Labeling System	Invitrogen	Cat#18247-015
Paint probe-green labeled Chromosome Y	MetaSystems	Cat#D-0324-100-FI
Paint probe-orange labeled Chromosome 21	MetaSystems	Cat#D-0321-100-OR
KpnI-HF	NEB	Cat#R3142
BamHI	NEB	Cat#R0136S
EcoRV	NEB	Cat#R0195S
Deposited data		
PacBio HiFi and ONT data generated from the CIMH-NP-3, CIMH-NP-3 wt Tau, and CIMH-NP-3 php Tau cell lines	This study, NCBI	NCBI: PRJNA1181812
Whole-genome assemblies generated from the CIMH-NP-3, CIMH-NP-3 wt Tau, and CIMH-NP-3 php Tau cell lines and containing the complete sequences of the chromosome 21 and Y centromeres	This study, NCBI	NCBI: PRJNA1181812, PRJNA1229460, PRJNA1229461, PRJNA1229462, PRJNA1229463
Genome assembly T2T-CHM13v2.0	Telomere-to-Telomer Consortium, NCBI	NCBI: PRJNA559484
Cell lines, Western blot, Southern blot and microscopy data	This study, Mendeley	Mendeley Data: https://doi.org/10.17632/6sfpcymkdk.1 https://doi.org/10.17632/v9fn28tvps.1 https://doi.org/10.17632/srwvmmyd58.1 https://doi.org/10.17632/gmz4mrbyz4.1 https://doi.org/10.17632/6sfpcymkdk.1; https://doi.org/10.17632/v9fn28tvps.1; https://doi.org/10.17632/srwvmmyd58.1; https://doi.org/10.17632/gmz4mrbyz4.1

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Nascent Strand-Sequencing data	This study, NCBI	NCBI: GSE286931
Proteomics data	This study, PRIDE	PRIDE: PXD058752
Experimental models: Cell lines		
CIMH-NP-3 cells Induced pluripotent stem cells (iPSC) derived from healthy male donor fibroblasts. CIMH-NP-3 wt Tau and php Tau cell lines.	Laboratory of Philipp Koch	N/A
Oligonucleotides		
Primers 13/21-3A and 13/21-1B 5′-CTTCTGCTCAGTTATTAGA-3′ 5′-CATAGAGTAGACATGG-3′	Warburton et al.[81]	N/A
Primers C_{HB} and C_{EB} 5′-GCTTTCGCTCCACCAACTA-3′ 5′-GCTTTCGCTCCACCAACTA-3′	Little et al.[78]	N/A
Recombinant DNA		
pLVX-EtO vector (modified Lenti-X Tet-On with CMV promoter replaced with EF1 promoter)	Mertens et al.[54]	https://doi.org/10.17632/6sfpcymkdk.1
cDNA 2N4R wt Tau and php Tau	Mertens et al.[54]	https://doi.org/10.17632/6sfpcymkdk.1
Software and algorithms		
Verkko v1.1	Rautiainen et al.[82]	https://github.com/marbl/verkko
hifiasm v0.16.1	Cheng et al., 2021[83]	https://github.com/chhylp123/hifiasm
Minimap2 v2.24	Li et al.[84]	https://github.com/lh3/minimap2
pbmm2 v1.1.0	PacificBiosciences/pbmm2	https://github.com/PacificBiosciences/pbmm2
Jellyfish	Marcais and Kingsford[85]	https://github.com/zippav/Jellyfish-2
Winnowmap v1.0	Jain et al.[86]	https://github.com/marbl/Winnowmap
IGV	Robinson et al.[87]	https://igv.org/
NucFreq	Vollger et al.[88]	https://github.com/mrvollger/NucFreq
Guppy v6.3.7	Oxford Nanopore Technologies	https://nanoporetech.com/software/other/guppy/history?version=6-3-7
Trimmomatic v0.36	Bolger et al.[89]	https://github.com/usadellab/Trimmomatic
Trim Galore v0.4.5	Babraham Bioinformatics, Felix Krueger	https://github.com/FelixKrueger/TrimGalore/releases/tag/0.4.5
FastQC v0.11.5	Babraham Bioinformatics	https://www.bioinformatics.babraham.ac.uk/projects/fastqc/
BWA v0.7.17	Li and Durbin[90]	https://github.com/sghignone/bwa
bedtools	Quinlan and Hall[91]	https://bedtools.readthedocs.io/
deepTools (multiBigwigSummary) v3.5.6	Ramírez et al.[92]	https://deeptools.readthedocs.io/en/develop/content/tools/multiBigwigSummary.html
R Project (ggplot2) v3.4.0	Wickham et al.[93]	https://cran.r-project.org/package=ggplot2
SeqBuilder Pro	DNASTAR Lasergene	https://www.dnastar.com/software/lasergene/seqbuilder-pro/
Adobe Photoshop v22.3.1	Adobe	https://www.adobe.com/products/photoshop.html
Adobe Illustrator v25.2.1	Adobe	https://www.adobe.com/products/illustrator.html
GraphPad prism v9; v10	GraphPad Software	https://www.graphpad.com/scientific-software/prism/
Proteome Discoverer software (SEQUEST) v2.5	Thermo Scientific	https://www.thermofisher.com/order/catalog/product/OPTON-30945
SwissProt human database	UniProtKB	https://www.expasy.org/resources/uniprotkb-swiss-prot
Gene Set Enrichment Analysis	Subramanian et al.[94]; Mootha et al.[95]	https://www.gsea-msigdb.org/gsea/index.jsp

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
WebGestalt	Zhang Lab	https://www.webgestalt.org/
String	STRING Consortium	https://string-db.org/

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Cell lines

The CIMH-NP-3 cell lines were provided by Philipp Koch laboratory at Zentralinstitut fÜr Seelische Gesundheit, Mannheim, Germany. The name was suggested by the International Cell Line Authentication Committee (ICLAC). All the cell lines have been authenticated by STR (short tandem repeat) profiling[96] and tested for mycoplasma contamination.

Induced pluripotent stem cells were obtained from dermal fibroblasts of a male healthy donor (age 33) using reprogramming by Sendai virus.

All procedures were reviewed and approved by Ethics Committee II, Medical Faculty Mannheim, University of Heidelberg (Approval No. 2009–350 N-MA). All experiments using human donor-derived primary cells were approved and in line with the Helsinki Declaration. Written informed consent was obtained from the donor prior to sample collection. A male donor was specifically selected to enable investigation of the Y chromosome.

IPSCs were differentiated into smNPCs (small molecule-Neural Progenitor Cells) as described previously,[97] precursor cells that require small molecules for self-renewal and expansion, using Wingless/Integrated (WNT) and Sonic hedgehog (Shh) pathways. Briefly, neuroectodermal fate was induced by dual-Smad inhibition for 10 days: 200 nM LDN-193189 (REPROCELL/Stemgent, 04-0074-10), and 500 nM A83 (REPROCELL/Stemgent, 04-0014-10) in DMEM/F12 (Invitrogen, 11320074), 0.5% N2 (Invitrogen, 17502048), 1% B27 (Invitrogen, 17504044), 1% GlutaMAX (Invitrogen, 35050038), 1% nonessential amino acid (NEAA) (Invitrogen, 11140035), 4.44 mM glucose (Roth, HN06.2).

Thereafter, cells were passed in low density on Geltrex (Invitrogen, A1413201) coated plastic dishes at 37°C and 5% CO_2 in smNPC medium (see cell culture).

METHOD DETAILS

Plasmid vector design and lentiviral transduction construction

Lenti-X Tet-On Advanced System (Clontech, Mountain View, CA) was modified by replacing the cytomegalovirus promoter of the pLVX-Tet-On plasmid with the elongation factor 1 (EF1) a promoter, resulting in the pLVX-EtO vector. The inserted coding DNA (cDNA) sequence are the longest 2N4R wt Tau for the normal human Tau protein and php Tau for the pathogenic variant (S195, 199, 202, 235, 396, 404, 409, 413, 442 and one Y231 substituted with the glutamic acid).[54]

The overexpression of Tau protein was induced by adding 3 μg/mL doxycycline (DOX) (Sigma Aldrich, D9891) for 72 h.

Cell culture

CIMH-NP-3 cells were grown on Geltrex (Invitrogen, A1413201) coated plastic dishes at 37°C and 5% CO_2 in smNPCs medium.

smNPCs medium is composed by Dulbecco’s Modified Eagle Medium/Nutrient Mixture F-12 (DMEM/F-12, Invitrogen, 11320074 (EU) or 11320033 (US)) containing 1% penicillin/streptomycin (Invitrogen, 15140122), 1% N2 (Invitrogen, 17502048) and 0.5% B27 (Invitrogen, 12587010) supplements, glucose 2.2 M (Sigma Aldrich, G7021), 1% L-glutamine (Invitrogen, 25030081), L-ascorbic acid 200 mM (Sigma Aldrich, A4544), CHIR99021 10 mM (Cell guidance, SM13), Purmorphamine 5 mM (Cell guidance, SM30). The CHIR99021 small molecule is a GSK3b inhibitor to stimulate the canonical WNT signaling pathway, while Purmorphamine small molecule is a tri-substituted purine derivative that activates the receptor Smoothened (Smo) inhibiting BODIPY-cyclopamine binding to Smo receptor.

The CIMH-NP-3 wt Tau and CIMH-NP-3 php Tau transformed cell lines were grown in smNPCs medium and selected for 48 h with 100 μg/mL G418 (Invitrogen, 10131035) and 1 μg/mL Puromycin (Sigma Aldrich, P7255-25MG).

Immunoblot for verification of CIMH-NP-3 cell lines

For Immunoblot analysis, transduced and untransduced CIMH-NP-3 were treated with 3 μg/ml doxycycline for 3 days. Cells were lysed in 50 mM tris-HCl (pH 7.4), 150 mM NaCl, 0.2% Triton X-100, 25 mM EDTA, 0.2% SDS, and 1× protease inhibitor for 1 hour on ice and sonicated. Protein content in the supernatant was measured by Pierce BCA Protein Assay (Thermo Fisher Scientific, 23227) and 25 μg of the total protein lysates were separated by SDS–polyacrylamide gel electrophoresis for 2 to 3 hours at 110 V. After blotting to 0.2-μm nitrocellulose membranes using the Trans-Blot TurboTM Transfer System (Bio-Rad) for 45 min at 1 A, membranes were blocked for 1 h in 5% milk powder in TBS-T and incubated with primary antibodies diluted in 5% milk powder in TBS-T overnight at 4°C as follows: FLAG (rabbit, 1:1000, Cell Signaling Technology, 2368S), Nestin (rabbit, 1:500, Abcam, ab27952), Sox2 (rabbit, 1:1000, Cell Signaling Technology, 3579), Actin (mouse, 1:10000, Cell Signaling Technology, 3700). Membranes were washed thrice in TBS-T, incubated with the following secondary antibodies in TBS-T for 1 h at RT: anti-rabbit DyeLight800 and anti-mouse DyeLight680 (both 1:15,000; both Cell Signaling Technology, rabbit 5151 and mouse 5470). Membranes were washed thrice, before imaging using an Odyssey imager (Li-COR).

Immunofluorescence verification of CIMH-NP-3 cell lines

For immunofluorescence analysis, transduced and non-transduced CIMH-NP-3 were seeded on Geltrex-coated glass coverslips. After treatment with 3 μg/ml doxycycline for 3 days, cells were fixed with 4% PFA for 10 min and washed thrice in 1xPBS before blocking and permeabilization in 10% fetal bovine serum (Invitrogen, 10270106), 0.3% Triton X-100 in PBS. Primary antibodies were diluted in blocking solution as follows and incubated overnight at 4°C: Sox2 (rabbit, 1:1000, Cell Signaling Technology, 3579), Nestin (mouse, 1:600, R&D Systems, MAB-1259), Tau (guinea pig, 1:300, Synaptic Systems, 314 004), Tubb3 (guinea pig, 1:1000, Synaptic Systems, 302 304), FLAG (mouse, 1:1000, Sigma Aldrich, F1804). Coverslips were washed thrice in 1xPBS and incubated with the following secondary antibodies for 1 h: rabbit-AF488, guinea pig-AF555 and mouse-AF647 (all 1:1000, all Invitrogen, rabbit A11008, guinea pig A21435 and mouse A21236). DNA was counterstained with DAPI and coverslips were mounted in Mowiol. Images were acquired using a Leica DM6 B microscope equipped with a DFC9000 GT camera and an HC PL APO 20× dry objective (Leica).

Monoclonal cell population stabilization and Immunoblotting

For CIMH-NP-3 wt and php Tau clone validation, the cells were grown in single cell clone to isolate a monoclonal cell population by limiting dilution method. Monoclonal cells were harvested by trypsinization, suspended in complete smNPCs medium, washed with 1xPBS, then pelleted and flash frozen in liquid N_2 and stored at -80°C. For SDS-PAGE, pellets were thawed on ice and lysed by resuspending in Laemmli Buffer (60 mM Tris-HCl pH 6.8, 400 mM 2-mercaptoethanol, 2% SDS, 10% glycerol, 0.01% bromophenol blue) to a final concentration of 106 cells ml-1. Lysates were denatured for 5 min at 100°C and passed through a 25-gauge needle (5x) then spun for 2 min at full speed in a microfuge. Aliquots of lysate corresponding to 10^5 cells were resolved on 4%-15% gradient SDS-PAGE gels (BioRad), proteins transferred to nitrocellulose membrane and blocked in 1xPBS with 5% Blotting-grade Blocker (BioRad) and 0.1% Tween-20. Membranes were then incubated with primary antibodies diluted in 1xPBS with 5% Blotting-grade Blocker. Primary antibodies used were anti-actin tag (rabbit polyclonal, Sigma, A2066), anti-human Tau (human monoclonal, Santa Cruz, sc-58860). Following incubation with primary antibodies membranes were washed with 0.1% Tween-20/1xPBS. Membranes were then incubated with fluorescently labeled Goat Anti-rabbit and Anti-human IRDye800CW (Li-Cor) secondary antibodies then washed in 0.1% Tween-20/1xPBS. Antibody signal was detected using Luminata HRP detection reagent (Millipore). Immunoblots were then imaged on an Odyssey Lc Infrared scanner (Li-Cor) or Luminescent Imager 680 (Amersham Biosciences).

DNA extraction, library preparation, and sequencing

Pacific Biosciences (PacBio) high-fidelity (HiFi) long-read sequencing data were generated from the CIMH-NP-3, CIMH-NP-3 wt Tau, and CIMH-NP-3 php Tau genomes as previously described,[12] with some modifications. Briefly, high-molecular-weight (HMW) DNA was extracted from cells using the Monarch HMW DNA Extraction Kit for Cells and Blood (New England Biolabs P/N T3050L). Then, the HMW DNA was used to generate PacBio HiFi libraries via the SMRTbell prep kit 3.0 and Barcoded Adapter Plate 3.0 (PacBio, 102-182-700 & 102-009-200). Size selection was performed with the PippinHT (Sage Science, HTP0001 & HPE7510) with a 15- or 17-kbp high-pass cutoff. After mass and length quantitation (DeNovix DS-11 FX, Thermo Fisher Qubit dsDNA High Sensitivity Q32851, Agilent FEMTO Pulse M5330AA and FP-1002-0275), the libraries were sequenced on the Sequel II platform (SMRT Link versions 11.0 or 11.1) with three to five SMRT Cells 8M (PacBio, 101-389-001) per sample using Sequel II Sequencing Chemistry 3.2 (PacBio, 101-333-300) with a 2-hour pre-extension and 30-hour movies, aiming for a minimum estimated coverage of 30X in PacBio HiFi reads (assuming a genome size of 3.1 Gbp). Raw sequencing data was processed using the CCS algorithm (v6.3.0 or 7.0.0) with the following parameters: –all –all-kinetics –subread-fallback. After filtering for reads that passed an estimated quality of Q20 (Phred scale), barcode sequences were removed with lima v2.5.1 or 2.7.1.

Oxford Nanopore Technologies (ONT) long-read sequencing data were generated from the CIMH-NP-3 genome according to a previously published protocol. Briefly, 3-5 x 10^7 cells were lysed in a buffer containing 10 mM Tris-Cl (pH 8.0), 0.1 M EDTA (pH 8.0), 0.5% w/v SDS, and 20 ug/mL RNase A (Qiagen, 19101) for 1 hour at 37°C. 200 ug/mL Proteinase K (Qiagen, 19131) was added, and the solution was incubated at 50°C for 2 hours. DNA was purified via two rounds of 25:24:1 phenol-chloroform-isoamyl alcohol extraction followed by ethanol precipitation. Precipitated DNA was solubilized in 10 mM Tris (pH 8.0) containing 0.02% Triton X-100 at 4°C for two days. Libraries were constructed using the Ultra-Long DNA Sequencing Kit (ONT, SQK-ULK001) with modifications to the manufacturer's protocol. Specifically, ~40 ug of DNA was mixed with FRA enzyme and FDB buffer as described in the protocol and incubated for 5 minutes at RT, followed by a 5-minute heat-inactivation at 75°C. RAP enzyme was mixed with the DNA solution and incubated at RT for 1 hour before the clean-up step. Clean-up was performed using the Nanobind UL Library Prep Kit (Circulomics, NB-900-601-01) and eluted in 225 uL EB. 75 uL of library was loaded onto a primed FLO-PRO002 R9.4.1 flow cell for sequencing on the PromethION, with two nuclease washes and reloads after 24 and 48 hours of sequencing. All ONT data were base-called with Guppy (v6.3.7) with the SUP model.

Sequence assembly and validation of centromeric regions

To generate complete sequence assemblies of the chromosome 21 and Y centromeric regions from the CIMH-NP-3 genome, we first assembled the CIMH-NP-3 PacBio HiFi data and ONT data using Verkko (v1.1). We aligned the resulting contigs to the T2T-CHM13 reference genome3 (v2.0) via minimap2 (v2.24) with the following parameters: -I 12G -a –eqx -x asm20 -s 5000. We identified the contigs that span the chromosome 21 and Y centromeres in the T2T-CHM13 reference genome, which were contiguous from p-to- q-arm.

To generate complete sequence assemblies of the chromosome 21 and Y centromeric regions from the CIMH-NP-3 wt Tau and CIMH-NP-3 php Tau genomes, we first assembled the CIMH-NP-3 wt Tau or CIMH-NP-3 php Tau PacBio HiFi data using both hifiasm (v0.16.1) and Verkko (v1.1). We aligned the contigs from each genome assembly to the T2T-CHM13 reference genome3 (v2.0) via minimap2 (v2.24) with the following parameters: -I 12G -a –eqx -x asm20 -s 5000. We identified the contigs that spanned the chromosome 21 and Y centromeres in the T2T-CHM13 reference genome, which were fragmented in all genome assemblies. We scaffolded the fragmented centromeric contigs from the hifiasm assembly with centromeric contigs from the Verkko assembly from the same genome via a method that takes advantage of SUNKs.[8] Briefly, SUNKs (k=20 bp) were identified from the hifiasm whole-genome assembly via Jellyfish (v2.2.4) and barcoded on the hifiasm centromeric contigs as well as the Verkko centromeric contigs. Contigs sharing a SUNK barcode were subsequently joined together to generate contiguous assemblies that traverse each centromeric region.

We validated the construction of each centromere assembly by aligning native PacBio HiFi and ONT data (when available) from the same source genome to each whole-genome assembly using pbmm2 (v1.1.0) (for PacBio HiFi data) or Winnowmap (v1.0) (for ONT data) and assessed the assemblies for uniform read depth across the centromeric regions via IGV and NucFreq. We confirmed that each chromosome 21 and Y centromere assembly lacked evidence of misassembles, such as mis-joins, sequence collapses, or false duplications.

Single Molecule Analysis of Replicated DNA (SMARD)

SMARD was performed as described in[58,74,76] with optimizations to detect α-satellite DNA and rDNA using specific enzymes and probes.

The probe for Y α-Satellite DNA was obtained by performing PCR on genomic DNA from human CIMH-NP-3 cells using a pair of primers, Y-A and Y-B, described in.[81,98] To detect Y α-Satellite DNA in a southern blot, the PCR product of 1 kbp was labeled with [α-32P] dCTP using RadPrime DNA Labeling system (Invitrogen, 18428-011). To detect Y α-Satellite DNA in SMARD, the PCR product was labeled with biotin using BioNick Labeling System (Invitrogen, 18247-015). The ChrY SMARD segment is obtained by digesting the genomic DNA with the restriction enzyme KpnI at GGTAC ˆC sequence (NEB, R3142).

The probe for Chr21 α-Satellite DNA was obtained by performing PCR on genomic DNA from human CIMH-NP-3 cells using a pair of primers, 13/21-3A and 13/21-1B (described in[81]). The PCR product of 600 bp was labeled as described above to use in Southern blot and SMARD. This probe detects satellite DNA in chromosomes 13 and 21. We distinguish Chr21 SMARD segment by the length of flanking sequence generated upon cutting genomic DNA with BamHI at G ˆGATCC sequence (NEB, R3136).

C_{HB} and C_{EB} rDNA segments described in[78] were labeled with [α-32P] dCTP to detect rDNA band in a Southern blot. pA and pC rDNA segments, were labeled with biotin to detect rDNA molecules in SMARD.

DNA fibers SMARD analysis

Replication analysis, i.e., determination of locations of origins and direction of fork movement and termination sites, by SMARD was based on analysis of the mixed red and green (IdU/CldU) dual labeled molecules. Only the dual labeled IdU/CldU molecules, which have red-green transitions, contain unambiguous information on the temporal order of replication in the DNA segment. The transitions indicate that red (IdU) labeled regions (which were labeled during the first pulse) are replicated before green (CldU) labeled regions (which were labeled during the second pulse). The red and green labeling patterns reveal and map the position, direction, and density of the replication forks in the replicated molecules. Although not used for replication analysis, molecules fully labeled in only red (IdU) or green (CldU) provide a valuable internal control. They provide a relative measure of DNA replication during each labeling pulse. Equal numbers of fully red and fully green molecules were obtained from different slides or SMARD replicates. Images were processed with Photoshop (Adobe) and aligned according to the FISH probe pattern using Illustrator software (Adobe).

Two color fiber-FISH

Two color fiber-FISH was performed as described in[99] with modifications to detect rDNA repeats. pB plasmid described in[78] was digested with EcoRV (NEB, R0195S) and labeled with digoxigenin using BioPrime DNA Labeling System (Ref# 18094-011). The biotinylated nucleotide mix from this kit was replaced with DIG DNA Labeling Mix (Roche, 11277065910). pA plasmid described in[78] is labeled with biotin using BioNick Labeling System (Invitrogen, 18247-015). Labeled pA and pB probes were hybridized to the DNA stretched and denatured on a microscopic slide coated with 3-aminopropyltriethoxysilane (Sigma Aldrich, 99 919-30-2) at 37°C overnight. Following hybridization, slides were blocked with 3% BSA/1xPBS for at least 20 min and incubated with following layers of antibodies prepared in 3% BSA/1xPBS: Layer 1: 1:100 avidin Alexa Fluor 568 (Invitrogen, S11226) and 1:100 anti-digoxigenin, FITC conjugated (Roche, 11207741910) for 1 h. Layer 2: 1: 200 biotinylated anti-avidin D (Vector laboratories, BA-0300) and 1:200 anti-FITC (Invitrogen, 71-1900) for 45 min. Layer 3: 1:200 avidin Alexa Fluor 568 for 20 min. Layer 4: 1:200 biotinylated anti-avidin D for 20 min. Layer 5: 1:200 avidin Alexa Fluor 568 for 20 min. Layer 6: 1: 200 biotinylated anti-avidin D and 1:200 anti-rabbit FITC conjugates (Rockland, 611-102-122) for 1 hr. Layer 7: 1:200 avidin Alexa Fluor 568 for 1 hr. Slides were rinsed in 0.03% IGEPAL CA-630/1xPBS after each incubation. After the last rinse in 0.03% IGEPAL CA-630/1xPBS, coverslips were mounted on the slides with Prolong gold antifade reagent (Invitrogen, P36930). Images were processed with Photoshop (Adobe) and aligned according to the FISH probe pattern using Illustrator software (Adobe).

Photomicrograph image acquisition for rDNA

Images of immunostained molecules were acquired at RT using a Zeiss Axioskop 2 fluorescence microscope equipped with a Plan Apochromatic 63X 1.4 NA oil objective and a charge-coupled device camera (CoolSNAP HQ; Photometrics) using 4.8 IPLab software (BD). Images were processed with Photoshop (Adobe) and aligned according to the FISH probe pattern using Illustrator software (Adobe). Only images of molecules that were fully labeled with IdU, CldU or both IdU and CldU, and displayed signals of both FISH probes and rDNA.

Fluorescence *in situ* hybridization with chromosome paint probes

The MetaSystem protocol was carried out for the hybridization of the paint probes. The green labeled-chromosome Y (MetaSystems, D-0324-100-FI) and orange labeled-chromosome 21 (MetaSystems, D-0321-100-OR) paint probes were used. The samples and paint FISH probes were denatured at the same time at 75°C for 2 min to be incubated at 37°C overnight in a humidified chamber. The next day, the slides were washed with 0.4xSSC at 72°C for 2 min, for then washed with 0.05% Tween-20/2xSSC at RT for 30 sec. The slides were wash with distilled water, before the closing of the slides with DAPI ProLong^{TM} Gold Antifade Mountant (ThermoFisher, P36941). The slides were stored at 4°C.

Nascent strand data analysis sequencing

Nascent strand sequencing (NS-Seq) analyses were performed as originally described in[67] with minor modifications (as described in[69]). In brief, RNA-primed DNA fragments were purified by isolating DNA from asynchronously growing, untreated cells under RNAse-free conditions, fractionation of denatured DNA samples on a neutral sucrose gradient followed by collection of fragments of 0.5–2 kb. Nascent DNA was further enriched in the population of size-selected single stranded DNA fragments (which contained a mixture of nascent strand DNA and broken genomic DNA) by λ exonuclease, which removes remove non-RNA-primed broken genomic DNA. λ exonuclease resistant, 0.5-2kbp DNA fragments were used as templates for the BioPrime DNA Labeling System (ThermoFisher, cat# 18094011) to create double-stranded fragments corresponding to nascent DNA sequences. A sample (1 μg) of the nascent DNA preparation was used to create a library. NS-seq data were collected in duplicates from two biological replicates for each sample.

Nascent strand sequence analyses

Libraries were sequenced on the Illumina Genome Analyzer II platform (Solexa), and genomic DNA processed without size fractionation served as control datasets for peak calling. Raw FASTQ reads were quality-trimmed using Trimmomatic (v0.36) and Trim Galore (v0.4.5), assessed with FastQC (v0.11.5), and aligned to the T2T genome assembly and a custom genome build (PRJNA1181812) using BWA (v0.7.17). BAM files were then converted into RPM-normalized BigWig files using deeptools for downstream analyses. Nascent strand peaks were called using exonuclease-treated genomic DNA controls, with peaks common to two independent biological replicates retained for subsequent analyses. Coverage data were extracted using the deeptools multiBigwigSummary function, and visualizations including violin plots, bar plots, and genome tracks were generated using R (ggplot2, v3.4.0) and GraphPad Prism 10. Peak overlap analysis between experimental conditions was carried out with bedtools intersect and subtract commands, and Venn diagrams were generated to highlight shared and unique peaks across conditions.

S-trap protein digestion

Immunoprecipitated proteins were eluted in a buffer containing 5% SDS, 5 mM DTT and 50 mM ammonium bicarbonate (pH = 8) and left at RT for about 1 h for disulfide bond reduction. Samples were then alkylated with 20 mM iodoacetamide in the dark for 30 min. Afterward, phosphoric acid was added to the sample at a final concentration of 1.2%. Samples were diluted in six volumes of binding buffer (90% methanol and 10 mM ammonium bicarbonate, pH 8.0). After gentle mixing, the protein solution was loaded to an S-trap filter (Protifi) and spun at 500 g for 30 sec. The sample was washed twice with binding buffer. Finally, 1 μg of sequencing grade trypsin (Promega), diluted in 50 mM ammonium bicarbonate, was added into the S-trap filter and samples were digested at 37°C for 18 h. Peptides were eluted in three steps: (I) 40 μl of 50 mM ammonium bicarbonate, (II) 40 μl of 0.1% TFA (Trifluoroacetic acid) and (III) 40 μl of 60% acetonitrile and 0.1% TFA. The peptide solution was pooled, spun at 1,000 g for 30 sec and dried in a vacuum centrifuge.

Sample desalting

Prior to mass spectrometry analysis, samples were desalted using a 96-well plate filter (Orochem) packed with 1 mg of Oasis HLB C-18 resin (Waters). Briefly, the samples were resuspended in 100 μl of 0.1% TFA and loaded onto the HLB resin, which was previously equilibrated using 100 μl of the same buffer. After washing with 100 μl of 0.1% TFA, the samples were eluted with a buffer containing 70 μl of 60% acetonitrile and 0.1% TFA and then dried in a vacuum centrifuge.

LC-MS/MS acquisition and analysis

Samples were resuspended in 10 μl of 0.1% TFA and loaded onto a Dionex RSLC Ultimate 300 (Thermo Scientific), coupled online with an Orbitrap Fusion Lumos (Thermo Scientific). Chromatographic separation was performed with a two-column system, consisting of a C-18 trap cartridge (300 μm ID, 5 mm length) and a picofrit analytical column (75 μm ID, 25 cm length) packed in-house with reversed-phase Repro-Sil Pur C18-AQ 3 μm resin. Peptides were separated using a 60 min gradient from 4-30% buffer B (buffer A: 0.1% formic acid, buffer B: 80% acetonitrile + 0.1% formic acid) at a flow rate of 300 nl/min. The mass spectrometer was set to acquire spectra in a data-dependent acquisition (DDA) mode. Briefly, the full MS scan was set to 300-1200 m/z in the orbitrap with a resolution of 120,000 (at 200 m/z) and an AGC target of $5\text{x}10^{e5}$. MS/MS was performed in the ion trap using the top speed mode (2 secs), an AGC target of $1\text{x}10^{e4}$ and an HCD collision energy of 35.

Proteome raw files were searched using Proteome Discoverer software (v2.5, Thermo Scientific) using SEQUEST search engine and the SwissProt human database (updated March 2024). The search for total proteome included variable modification of N-terminal acetylation, and fixed modification of carbamidomethyl cysteine. Trypsin was specified as the digestive enzyme with up to 2 missed cleavages allowed. Mass tolerance was set to 10 pm for precursor ions and 0.2 Da for product ions. Peptide and protein false discovery rate was set to 1%. Following the search, data was processed as previously described. Briefly, proteins were log2-transformed, normalized by the average value of each sample and missing values were imputed using a normal distribution of 2 standard deviations lower than the mean. Statistical regulation was assessed using heteroscedastic T-test (if $p\text{-value} < 0.05$). Data distribution was assumed to be normal, but this was not formally tested. Gene Set Enrichment Analysis, WebGestalt and String were used to log2-transformed data.

QUANTIFICATION AND STATISTICAL ANALYSIS

All statistical analyses were calculated using GraphPad prism and are indicated in the figure legends. For replication dynamics quantified from SMARD, the number of replication forks along the SMARD molecules of chromosomes Y and 21 was quantified per 30 kbp intervals, and z-scores were calculated as $z = (x - \mu)/\sigma$, where x is the specific value, μ is the mean, and σ is the standard deviation (one-tailed); values ≥ 1.645 were considered significant. Fisher's test was used to compare the number of molecules with or without replication forks and ORIs. Differences in replicon fork speed were assessed using the Mann-Whitney U test. For all SMARD analyses, the number of molecules examined from several replicates were: ChrY, CIMH-NP-3-DOX (n=36), wt Tau (n=34,35) and php −/+DOX (n=31,34); Chr21, CIMH-NP-3-DOX (n=23), wt Tau (n=23,23) and php −/+DOX (n=21,22). n= number of molecules. For replication dynamics quantified from NS-seq datasets (n=2 independent experiments),statistical significance between experimental conditions was defined as $P \leq 0.003$ for unpaired t-tests and $P \leq 0.0001$ for Wilcoxon signed-rank tests. Data are presented as mean ± SEM unless otherwise noted. Proteomics data (n=2 independent experiments) were $\log_2$-transformed for normalization, and $\log_2$ fold changes in protein abundance were calculated to compare IgG and FLAG tag M2 samples for both wt Tau and php Tau samples pulled down in nuclear fractions.

Non-significant differences are reported as "ns.", $^{*}P \leq 0.05$, $^{**}P \leq 0.01$, $^{***}P \leq 0.001$, and $^{****}P \leq 0.0001$.