Quantitative assessment reveals the dominance of duplicated sequences in germline-derived extrachromosomal circular DNA

Lila Mouakkad-Montoya[a,1], Michael M. Murata[a,1], Arvis Sulovari[b,c], Ryusuke Suzuki[a], Beth Osia[d], Anna Malkova[d], Makoto Katsumata[e], Armando E. Giuliano[a,e,f], Evan E. Eichler[b,c], and Hisashi Tanaka[a,e,f,2]

[a]Department of Surgery, Cedars-Sinai Medical Center, Los Angeles, CA 90048; [b]Department of Genome Sciences, University of Washington School of Medicine, Seattle, WA 98195; [c]HHMI, University of Washington School of Medicine, Seattle, WA 98195; [d]Department of Biology, University of Iowa, Iowa City, IA 52242; [e]Department of Biomedical Sciences, Cedars-Sinai Medical Center, Los Angeles, CA 90048; and [f]Samuel Oschin Comprehensive Cancer Institute, Cedars-Sinai Medical Center, Los Angeles, CA 90048

Edited by Scott Keeney, Memorial Sloan Kettering Cancer Center, New York, NY, and approved October 12, 2021 (received for review February 15, 2021)

Extrachromosomal circular DNA (eccDNA) originates from linear chromosomal DNA in various human tissues under physiological and disease conditions. The genomic origins of eccDNA have largely been investigated using in vitro–amplified DNA. However, in vitro amplification obscures quantitative information by skewing the total population stoichiometry. In addition, the analyses have focused on eccDNA stemming from single-copy genomic regions, leaving eccDNA from multicopy regions unexamined. To address these issues, we isolated eccDNA without in vitro amplification (naïve small circular DNA, nscDNA) and assessed the populations quantitatively by integrated genomic, molecular, and cytogenetic approaches. nscDNA of up to tens of kilobases were successfully enriched by our approach and were predominantly derived from multicopy genomic regions including segmental duplications (SDs). SDs, which account for 5% of the human genome and are hotspots for copy number variations, were significantly overrepresented in sperm nscDNA, with three times more sequencing reads derived from SDs than from the entire single-copy regions. SDs were also overrepresented in mouse sperm nscDNA, which we estimated to comprise 0.2% of nuclear DNA. Considering that eccDNA can be integrated into chromosomes, germline-derived nscDNA may be a mediator of genome diversity.

extrachromosomal circular DNA | sperm | segmental duplications | multi-mapped reads

In eukaryotes, the vast majority of genetic information is encoded in the linear chromosomes of the nucleus where circular DNA elements, or extrachromosomal circular DNA (eccDNA), are also present (1). eccDNA could modulate the genetic information of linear chromosomes and, as a result, phenotypes (2–5). This is the case with the most extensively studied extrachromosomal circular chromosomes, double minutes (DMs), which are several megabases (Mb) in size and are heavily implicated in cancer progression (2). DMs carry amplified oncogenes and DNA regulatory elements such as enhancers (3, 4) that promote the therapeutic resistance and progression of tumors (5). This phenotypic plasticity is associated with the remodeling of cancer genomes, as DMs can emerge from linear chromosomes, circularize, and reintegrate back into the genome (6, 7). In contrast, the biological significance of small-sized eccDNA populations found in both cancer and noncancer genomes remains unclear despite the recent surge of nucleotide-level data from next-generation sequencing (NGS)-based studies (8–15).

Uncovering the expanding biological roles of eccDNA necessitates a delineation of their characteristics, including their chromosomal compositions and prevalence in cells. A barrier to fully understanding eccDNA biology could be associated with eccDNA enrichment approaches that rely on multiple displacement amplification (MDA), a non-PCR technique that amplifies DNA with a highly processive Φ29 DNA polymerase (16). When the template DNA is circular, DNA synthesis continues indefinitely, which is known as rolling-circle amplification (RCA). While RCA amplifies circular DNA, it often overrepresents small-sized populations of eccDNA (17, 18), rendering the analysis of in vitro–amplified DNA qualitative and not quantitative. RCA also eliminates epigenetic information associated with template DNA, leaving crucial biological information unaddressed. Furthermore, calling eccDNA is based on identifying sequencing reads that span the fusion points of circular DNA, which requires that discordant reads are confidently mapped to single-copy genomic regions (19).

To overcome these issues and gain a broader scope of eccDNA characteristics, we developed a strategy to enrich endogenous eccDNA populations at the native state without RCA, which we designated as naive small circular DNA (nscDNA). The purity of nscDNA populations was rigorously validated by both internal (mitochondrial DNA) and exogenous (plasmid) controls. We integrated molecular, cytogenetic, and genomic approaches to further validate and examine the physical properties and genomic origins of nscDNA. We found that human nscDNA ranges from ~0.5 to greater than 10 kilobases

Significance

Extrachromosomal circular DNA (eccDNA) plays a role in human diseases such as cancer, but little is known about the impact of eccDNA in healthy human biology. Since eccDNA is a tiny fraction of nuclear DNA, artificial amplification has been employed to increase eccDNA amounts, resulting in the loss of native compositions. We developed an approach to enrich eccDNA populations at the native state (naïve small circular DNA, nscDNA) and investigated their origins in the human genome. We found that, in human sperm, the vast majority of nscDNA came from high-copy genomic regions, including the most variable regions between individuals. Because eccDNA can be incorporated back into chromosomes, eccDNA may promote human genetic variation.

Author contributions: L.M.-M., M.M.M., A.M., E.E.E., and H.T. designed research; L.M.-M., M.M.M., R.S., M.K., and H.T. performed research; L.M.-M., M.M.M., A.S., R.S., B.O., and H.T. analyzed data; L.M.-M., M.M.M., R.S., and H.T. wrote the paper; A.G. contributed new reagents/analytic tools; and A.G. provided funding support.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

[1]L.M.-M. and M.M.M. contributed equally to this work.

[2]To whom correspondence may be addressed. Email: Hisashi.Tanaka@cshs.org.

This article contains supporting information online at http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2102842118/-/DCSupplemental.

Published November 17, 2021.

(kb), encompassing relatively unexplored size populations of circular DNA between microDNA (<400 base pairs [bp]) and DMs. We rigorously examined the genomic origins of nscDNA beyond single-copy regions and found that nscDNA is predominantly derived from duplicated sequences of the human genome including segmental duplications (SDs). SDs represent recently duplicated genomic regions and are considered to be actively evolving areas of the human genome, as they are strongly associated with copy number variations and contribute more to genetic diversity in humans than single nucleotide polymorphisms (20–22). The strong association of duplicated sequences with genetic diversity prompted us to study germline-derived nscDNA in human and mouse sperm and tissues, which we found to be overrepresented with SDs.

Results

Enrichment of eccDNA without In Vitro Amplification. We developed a strategy to enrich nscDNA populations without in vitro amplification (Fig. 1*A*). In brief, we used alkaline conditions for cell lysis and neutral conditions to recover circular DNA, which were captured through an anion exchange column typically used for isolating bacterial plasmids. Following column purification, nscDNA-enriched samples were treated with excess amounts of exonuclease (Plasmid-Safe DNase), which efficiently hydrolyzes linear double-stranded DNA (dsDNA) (23). As a measure of nscDNA quality control, we analyzed the relative quantity of mitochondrial DNA (mtDNA) to chromosomal DNA after exonuclease digestion using qPCR in human normal (IMR-90, GM12878) and cancer (HeLa S3, Colo320DM) cell lines (Fig. 1*B*). mtDNA served as an internal control for circular DNA, while a chr17 locus represented the population of linear DNA. The mtDNA:chr17 ratio, after normalization to genomic DNA (gDNA), was at least 10,000:1 in all cases, indicating extensive enrichment of circular DNA and depletion of linear DNA. The enrichment of mtDNA was reproducible when we used another commercially available exonuclease, *Escherichia coli* Exonuclease V (Exo V) (*SI Appendix*, Fig. S1*A*).

We further examined physical properties of HeLa S3 nscDNA by pulsed-field gel electrophoresis (PFGE) and sonication and found that nscDNA exhibited very distinct mobility and sonication response (Fig. 1*C*). A significant fraction of nscDNA was trapped in the wells, a characteristic of large circular DNA (24, 25). gDNA was very sensitive to sonication, and high–molecular weight linear DNA disappeared with 45 s of sonication. In contrast, nscDNA was relatively resistant to sonication, indicating another property of circular DNA (26). Finally, Southern blotting showed that mtDNA was extremely enriched in nscDNA over gDNA samples.

The enrichment of mtDNA in our nscDNA pools shows the efficient recovery of circular DNA of large sizes (>10 kb), including very high–molecular weight eccDNA that have not been rigorously investigated (27). We sequenced cell line–derived nscDNA and aligned the reads to hg38 to investigate their compositions (Datasets S1–S5). In all nscDNA libraries, over 90% of reads mapped to chrM (Fig. 1*D* and *SI Appendix*, Fig. S1*B*). As Colo320DM has numerous large extrachromosomal, double minute chromosomes (DMs) covering the 1.6-Mb region at 8q24.1, in addition to chromosomally integrated amplicons (28), we assessed the region for the recovery of large nscDNA of nuclear origin. At 8q24.1, nscDNA coverage increased compared to the 500-kb flanking regions on both sides of the locus (Fig. 1*E*) to a similar degree as gDNA (Fig. 1*E*, *Right*). In contrast, nscDNA from Colo320HSR, a sister cell line in which amplicons are mostly within a single chromosome, forming a homogenously staining region (HSR) (29), contained lower coverage than gDNA, showing the reproducible purification of circular DNA molecules from cell lines known to contain DMs.

A common method to enrich eccDNA involves RCA (8–12, 15). However, RCA seemed to deplete large circular DNA and, on occasion, created artificial eccDNA. We compared HeLa S3 nscDNA pools amplified by RCA to nonamplified nscDNA and found that circular mtDNA quantities decreased 10-fold (*SI Appendix*, Fig. S1*C*). We also observed that the copy number of chr17 was extremely higher in one of the replicates over others. Phi29 polymerase often switches templates (30). Switching templates within the same DNA fragment would initiate RCA and lead to amplification. Furthermore, while nscDNA retained DNA methylation, nscDNA amplified by RCA did not, as RCA copies methylated cytosine to cytosine. We measured the methylation of long interspersed nuclear elements (LINE-1) in GM12878 and HeLa S3 DNA and found that amplified DNA exhibited a dramatic reduction of 5% methylcytosine to background levels, whereas nonamplified nscDNA retained methylation higher than background levels (*SI Appendix*, Fig. S1*D*).

Overrepresentation of Segmental Duplications in Cell Line-Derived nscDNA. The very high fractions of mtDNA from cell lines limited the sequencing depth of nscDNA of nuclear origin; however, we proceeded to characterize the non-mtDNA reads using the bowtie2 aligner (31). We first confirmed the very high mappability (>95%) of nscDNA to hg38, with <5% of unmapped reads from all cell lines tested (*SI Appendix*, Fig. S2*A*). We then examined the reads that aligned with hg38 exactly once (aligned = 1, uniquely mapped reads, UMR) and more than once (aligned > 1, multimapped reads, MMR). UMR were defined by a mapping quality Phred score (MAPQ ≥ 40) in which the probability of erroneous mapping was <0.0001 (32). Subtracting the number of UMR from the total number of mapped reads returned the number of MMR. Noticeably, nscDNA reads from all cell lines except Colo320DM were dominantly MMR (Fig. 2*A*), whereas at least 75% of gDNA reads were UMR. Colo320DM nscDNA was an outlier in its higher composition of UMR (59.3 ± 3.9%). There are several high-coverage, uniquely mappable areas in this cell line including the DM locus (Fig. 1*E*), which would contribute to the higher coverage of UMR.

Since the majority of nscDNA reads were multimapped, we questioned whether repetitive elements (short interspersed nuclear element [SINE]/LINE/Simple/Satellite) were major constituents. Interestingly, the dominance of MMR did not correlate with a greater composition of repetitive elements in nscDNA (Fig. 2*B* and *SI Appendix*, Fig. S2*B*). Another major source of MMR are nearly identical, recently duplicated sequences, or SDs (also called low copy repeats). SDs are DNA segments (>1 kb) that occur more than once in the human genome and share a very high sequence identity (>90%) with each other (33). Initially, we found multiple high-coverage peaks throughout the genome corresponding to SDs in nscDNA (Fig. 2*C*). We cautioned this observation because ∼13 kb of mtDNA segments are nearly identical with SDs in chr1, chr3, chr5, chr11, and chr14 (*SI Appendix*, Fig. S2*C*).

We utilized a defined set of SD tracks to quantify the percentage of SDs in nscDNA and gDNA reads with analyses including (*SI Appendix*, Fig. S2*D*) or excluding (Fig. 2*D*) the SD tracks associated with mtDNA. Although the exclusion would most likely underestimate the SD fraction in nscDNA, SD enrichment was still observed in nscDNA. GM12878 and HeLa S3 carried the highest percentages of nscDNA reads attributed to SDs at 12.6 ± 2.0 and 13.2 ± 1.6%, which was more than twice the amount of SDs in corresponding gDNA samples (5.5 and 5.8%, respectively).

We confirmed that SD-containing nscDNA were indeed extrachromosomal by metaphase fluorescence in situ hybridization (FISH) on HeLa S3 and GM12878 cells (Fig. 2*E* and *SI Appendix*, Fig. S2*E*) using two SD-harboring fosmid clones

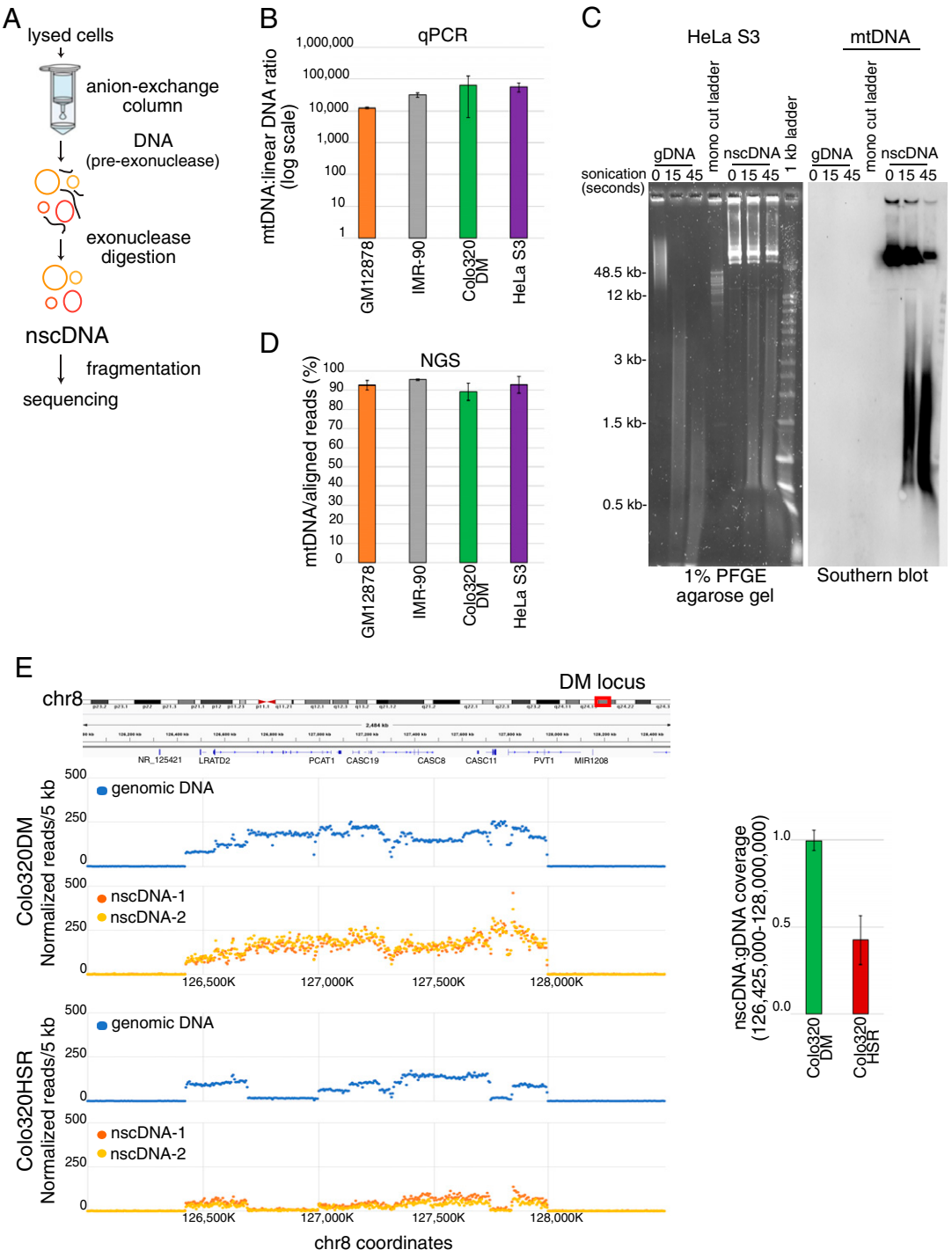

Fig. 1. Purification and enrichment of nscDNA without in vitro amplification. (*A*) A schematic overview of the purification and enrichment of nscDNA. Cell lysis was followed by circular DNA isolation through an anion-exchange resin typically used for isolating plasmids. Contaminating linear DNA was digested by Plasmid-Safe DNase (Lucigen). Following qPCR to confirm the enrichment of exonuclease-resistant nscDNA, nscDNA libraries were constructed, sequenced, and analyzed through our NGS pipeline. gDNA samples in which nscDNA was not enriched were also analyzed. (*B*) qPCR results displaying the mtDNA:chr17 ratio (log scale) of enriched nscDNA from cell lines preceding library construction. The relative quantities for both targets were normalized to gDNA samples. $n = 1$ for all gDNA. $n = 3$ for Colo320DM/HeLa S3 nscDNA. $n = 2$ for IMR-90/GM12878 nscDNA. qPCR reactions were run in quadruplicates. The data represent the mean ± SEM. (*C*) A total of 60 ng HeLa S3 gDNA and nscDNA (both sonicated by Bioruptor Pico for 0s/15s/45s) were size fractionated by PFGE (*Left*), stained with SYBR Gold, and analyzed by Southern blotting with hybridization to a DIG-labeled mtDNA probe (*Right*). (*D*) Percentages of paired reads in nscDNA from cell lines mapped to chrM. $n = 1$ for all gDNA. $n = 3$ for Colo320DM/HeLa S3 nscDNA. $n = 2$ for IMR-90/GM12878 nscDNA. The data represent the mean ± SD. (*E*) The read coverages per 5-kb bin (the number of reads divided by per-million scaling factor excluding reads mapped to chrM) of the gDNA (blue) and nscDNA (yellow/orange) for the amplified DM locus in Colo320DM (*Top*) and for the chromosomally integrated amplicons in the sister cell line, Colo320HSR (*Bottom*). The DM locus in chr8 is displayed along with 500-kb surrounding regions. The nscDNA to gDNA coverage ratio is displayed (*Right*) for the 126,425,000- to 128,000,000-bp region. $n = 1$ for gDNA and $n = 2$ for nscDNA. The data represent the mean ± SD.

https://doi.org/10.1073/pnas.2102842118

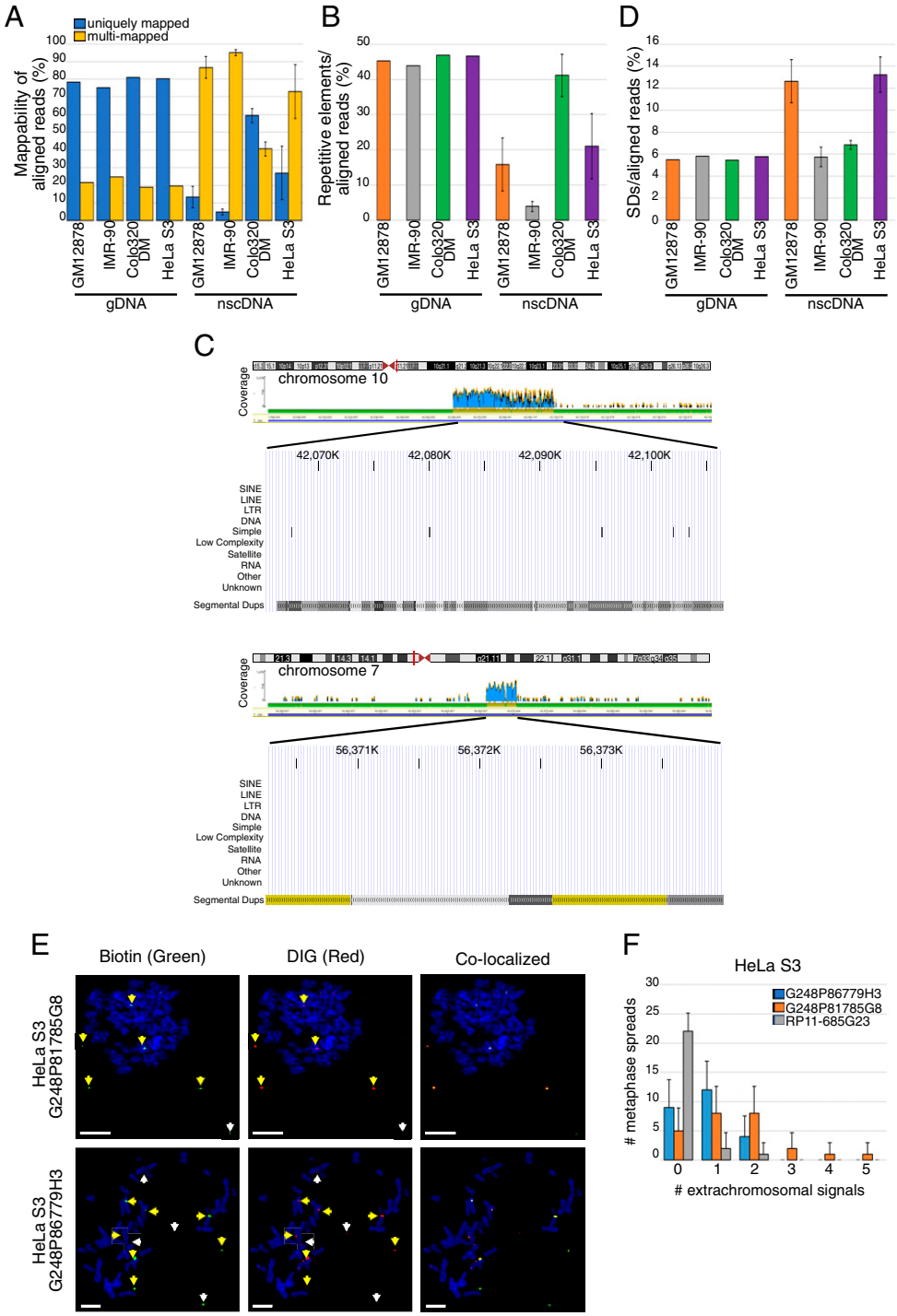

Fig. 2. Genomic, molecular, and cytogenetic analysis of cell line–derived nscDNA. (*A*) Percentages of total mappable unique (blue) and multimapped (yellow) reads in gDNA and nscDNA to hg38 excluding reads mapped to chrM. $n = 1$ for all gDNA. $n = 3$ for Colo320DM/HeLa S3 nscDNA. $n = 2$ for IMR-90/GM12878 nscDNA. The data represent the mean ± SD. (*B*) Percentages of total mappable reads aligned with repetitive elements (SINE/LINE/Simple/Satellite) in gDNA and nscDNA as defined by RepeatMasker in hg38 excluding reads mapped to chrM. $n = 1$ for all gDNA. $n = 3$ for Colo320DM/HeLa S3 nscDNA. $n = 2$ for IMR-90/GM12878 nscDNA. The data represent the mean ± SD. (*C*) Geneious software snapshot of GM12878 nscDNA coverage peaks (blue) found in chr10q11.21 (*Top*) and chr7p11.2 (*Bottom*) after paired read alignment to hg38. The repetitive elements and SDs associated with these regions are displayed using the UCSC Genome Browser. Light to dark gray and light to dark yellow SD tracks indicate 90 to 98% and 98 to 99% sequence similarity between SDs, respectively. (*D*) Percentages of total mappable reads in gDNA and nscDNA reads aligned with SD tracks in hg38. The reads mapped to SD tracks in mtDNA were excluded. $n = 1$ for all gDNA. $n = 3$ for Colo320DM/HeLa S3 nscDNA. $n = 2$ for IMR-90/GM12878 nscDNA. The data represent the mean ± SD. (*E*) Representative metaphase FISH spreads of HeLa S3 chromosomes stained with DAPI (blue) and hybridized with two different SD-harboring fosmid clones, G248P81785G8 (*Top*) and G248P86779H3 (*Bottom*), as probes. The probes were labeled with two different haptens, DIG (red) or Biotin (green). The yellow signals (yellow arrows) in which probes colocalized were considered positive. White arrows indicate localization of only one probe. (Scale bars: 10 μm.) (*F*) The number of colocalized extrachromosomal signals in HeLa S3 metaphase spreads with two SD-harboring probes (G248P81785G8/G248P86779H3). The number of spreads (*y*-axis) and the number of signals (0 to 5) counted (*x*-axis) are displayed. A probe (RP11-685G23) representing a single-copy genomic region was used as a negative control. The error bars represent 95% CIs. The *P* values (Fisher's exact test, two-tailed) were calculated comparing two outcomes (signal versus no signal) for G248P81785G8 ($P = 0.0001$) and G248P86779H3 ($P = 0.0003$) against RP11-685G23.

(G248P86779H3/G248P81785G8) as probes. To exclude false-positive signals, we applied very stringent criteria and labeled the probes with either DIG (red) or Biotin (green) and searched for signals in which colocalization (yellow) occurred. We observed signal colocalization in multiple chromosomes as well as outside of chromosomes. In summary, 16 (G248P86779H3) and 20 (G248P81785G8) out of 25 metaphases in HeLa S3 cells exhibited extrachromosomal signals (Fig. 2*F*), while 15 (G248P86779H3) and 19 (G248P81785G8) out of 25 showed signals in GM12878 (*SI Appendix*, Fig. S2*F*). Many of the metaphases had multiple extrachromosomal signals. In contrast, extrachromosomal signals were very rare in metaphases hybridized to a BAC clone (RP11-685G23) from a single-copy genomic region that was not enriched with eccDNA. Despite covering a much larger genomic region (163 kb) than fosmid clones (~40 kb), only three metaphases showed extrachromosomal signals with RP11-685G23.

Dominance of SDs in Human and Mouse Sperm nscDNA. Since SDs are known to contribute disproportionately to copy number variations within humans and between primates (21, 22), we hypothesized that nscDNA may act as mediators in these events given their high SD composition in cell lines. We approached this question by enriching for nscDNA from pooled frozen human semen (0.5 mL/replicate) and mature mouse sperm extracted from the epididymis of four C57BL/6 mice (two mice/replicate). Because mtDNA could possibly be degraded technically during semen preservation (34) or biologically during the process of eliminating paternal mtDNA (35), we developed an exogenous plasmid-based quality control method to quantify the enrichment of nscDNA by spiking circular pUC18 and linearized pEGFP-C1 (double digested) into lysed sperm cells. Following enrichment, we obtained circular:linear plasmid ratios of 738 ± 149 and 1,418 ± 254 for human and mouse nscDNA, respectively (Fig. 3*A*). Following quality control, we directly examined nscDNA from human sperm using transmission electron microscopy (*SI Appendix*, Fig. S3*A*). Unlike high–molecular weight gDNA, we observed a great degree of negative staining in nscDNA, possibly because of the high density of DNA in large supercoiled structures (36).

Importantly, our approach enabled us to estimate the amount of nscDNA per sperm cell, which has not been possible in RCA-amplified DNA. We recovered 18 to 58 ng of nscDNA from mouse sperm after exonuclease treatment, which amounts to 0.9 to 6.5 fg of nscDNA/cell (*SI Appendix*, Table S1). Since a single mouse haploid nucleus contains 3 pg of DNA (37), we could infer that nscDNA accounts for up to 0.2% of nuclear DNA in sperm. This figure could be underestimated given that the loss of nscDNA at each purification step was not considered.

We sequenced human and mouse sperm nscDNA and, as we anticipated, mtDNA was not enriched (0.0006 ± 0.00002% in human and 0.02 ± 0.004% in mouse). Instead, we found that nscDNA reads were dominantly MMR (human: 82.2 ± 1.4%, mouse: 96.6 ± 0.9%) (Fig. 3*B*). We noticed that, unlike cell line–derived nscDNA, sperm nscDNA carried greater total repetitive elements than gDNA because of the profound dominance of satellite DNA in both human (52.9 ± 1.7%) and mouse (81.6 ± 0.6%) samples (Fig. 3*C*). Significantly, human nscDNA was four times more enriched with SDs than gDNA (nscDNA: 21.2 ± 1.1%, gDNA: 5.7%) (Fig. 3*D*). The presence of SDs in human sperm nscDNA was also confirmed through Southern blotting (*SI Appendix*, Fig. S3*B*). Consistent with the PFGE for HeLa S3 nscDNA (Fig. 1*C*), sperm nscDNA showed very different mobility in response to sonication. gDNA was readily sonicated and became smaller in length, whereas nscDNA was trapped in wells with little changes by sonication. In corroboration with the NGS data (Fig. 3*C*), both SDs and satellite DNA were enriched in nscDNA relative to gDNA.

SD enrichments were exemplified by high coverage peaks of nscDNA in the genomic regions with multiple SDs (Fig. 3*E*). Some of the high coverage regions overlapped between SDs and satellite DNA (*SI Appendix*, Fig. S4*A*) or were only attributed to satellites (*SI Appendix*, Fig. S4*B*). Given that 1) SD tracks are not repeat masked and 2) SDs overlap with satellite DNA in pericentromeric regions (38), we cautioned that the overrepresentation of SDs in nscDNA would be as a result of the dominance of satellite DNA. After removing the intersection between SDs and satellite DNA, however, we found that SDs remained overrepresented in human sperm nscDNA (13.9 ± 0.1%) (Fig. 3*F*). With the dominance of SDs and satellite DNA, UMR remained a minor fraction (Fig. 3*B*). Reads mapped to SDs were more abundant than total UMR in both human (3.4 times) and mouse (5.9 times). The ratio of SD-mapped reads to UMR was 42-fold and 67-fold higher in nscDNA than in gDNA in human and mouse, respectively (Fig. 3*G*).

These results indicate that circular DNA molecules in sperm, small or large, would arise predominantly from duplicated regions of the genome. This notion was further validated for human nscDNA by mapping to the most recent telomere-to-telomere assembly of the haploid CHM13 cell line genome (39). This assembly fills many of the gaps in hg38, which significantly coincide with SDs (40). The mappability of sperm nscDNA to hg38 was low; on average, only 62.0 ± 4.8% of reads were mapped to the hg38 genome (Fig. 3*H*). With the CHM13 genome, we observed a dramatic increase (23.4%) in the mappability of human sperm nscDNA, while gDNA mappability remained almost the same between hg38 (96.1%) and CHM13 (96.9%). Thus, the nearly identical SDs and other sequences in hg38 gaps are the significant source of nscDNA.

Fusion Junctions of nscDNA. Highly identical sequences could promote the formation of nscDNA by homology-directed rearrangements. Understanding the mechanism of nscDNA formation requires precise, nucleotide-level information at fusion points. Previous studies of eccDNA have relied on the analysis of discordantly aligned read pairs with outward orientation (in the opposite direction) as evidence of circularization (8, 9, 12). Examples of this were observed in the *TWIST2* gene (chr2) and *NSUN6* gene (chr10) in human sperm nscDNA, which contained peaks with outwardly oriented read pairs (outward pairs, red bars), suggesting hotspots for nscDNA formation (Fig. 4*A*). In mouse sperm nscDNA, paired reads of outward orientation encompassed part of the *ASMT* gene in chrX (*SI Appendix*, Fig. S5*A*). We systematically investigated read pairs (excluding chrM reads) in sequencing libraries from normal cell lines and sperm samples and calculated the percentage of outward pairs (*SI Appendix*, Fig. S5*B*). As we expected, outward pairs were rare (<0.5%) in the gDNA libraries of IMR-90, GM12878, and human sperm nscDNA. In the nscDNA libraries of these samples, outward pairs encompassed >1.5% of reads. In contrast, in mouse sperm, the nscDNA and gDNA samples carried approximately equal percentages of outward pairs. Therefore, although the analyses of unpaired reads revealed similar characteristics of nscDNA between human and mouse sperm, such as the enrichment of satellite DNA and SDs, paired read analyses did not. Further investigation is needed on this matter.

To confidently identify specific sequences contributing to nscDNA fusion points, UMR are necessary. To do so, we turned to Colo320DM nscDNA data and investigated fusion points in 8q24.1, a locus surrounding the *MYC* gene that contains very few SDs. The deep coverage of nscDNA in the amplified locus in Colo320DM (Fig. 1*E*) facilitated our survey for fusion points involved in circular DNA formation. In most of the region (126,425,000 to 128,000,000), the nscDNA:gDNA coverage ratio per 1 kb was between 0.6 to 2.0 (Fig. 4*B*). We noticed that the 6-kb region within the long noncoding enhancer RNA

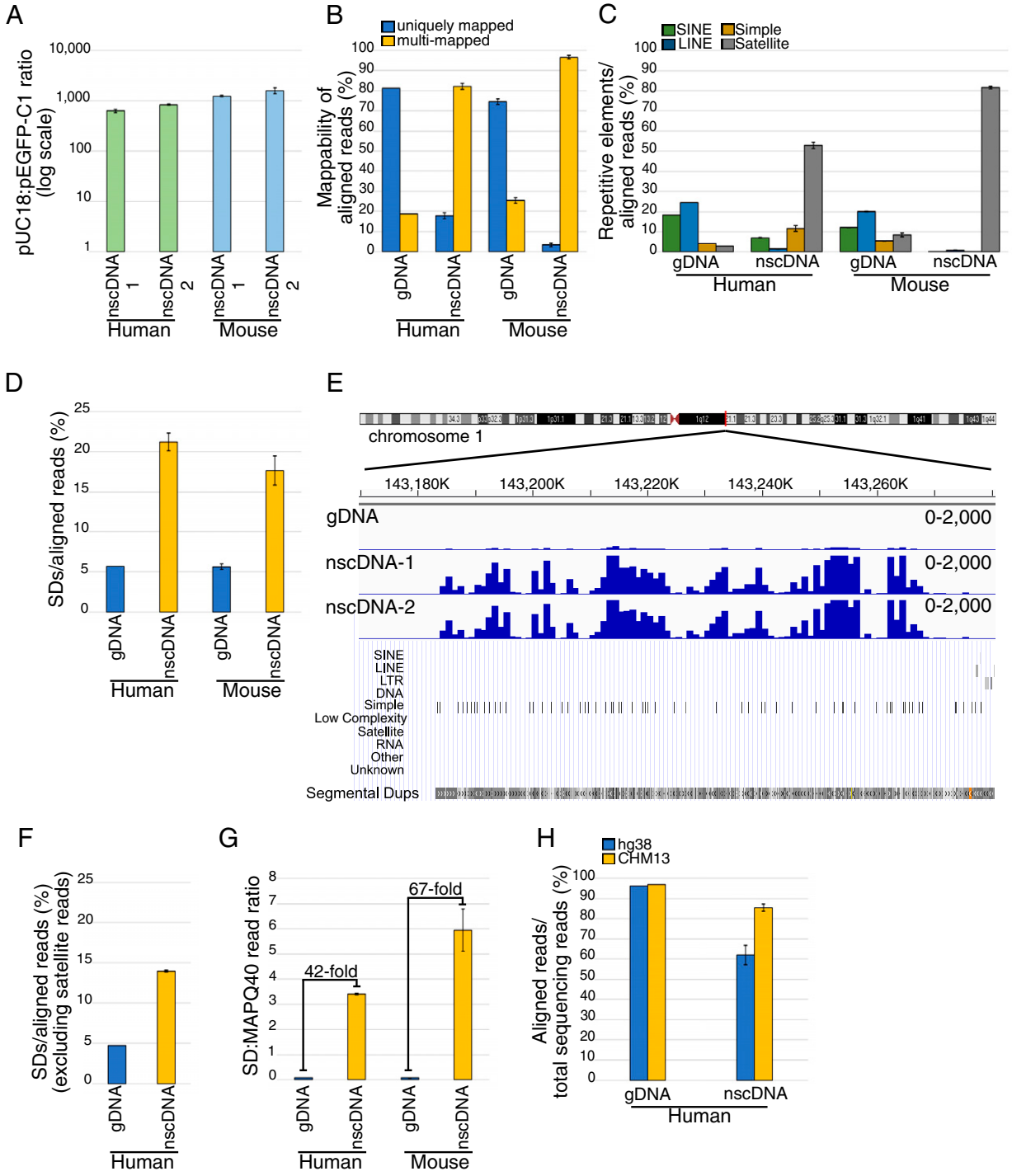

Fig. 3. Genomic compositions of sperm nscDNA. (*A*) qPCR results of the pUC18:pEGFP-C1 ratio (log scale) of enriched human/mouse sperm nscDNA preceding library construction. During sperm lysis, nscDNA and gDNA sample preparations were spiked with equal amounts of circular pUC18 and linearized pEGFP-C1 (*EcoRI*-*BamHI*). Relative quantities for both targets were normalized to gDNA samples. $n = 1$ for human sperm gDNA. $n = 2$ for human/mouse sperm nscDNA and mouse sperm gDNA. qPCR reactions were run in quadruplicates. The data represent the mean ± SEM. (*B*) Percentages of total mappable unique (blue) and multimapped (yellow) reads in sperm gDNA and nscDNA aligned with hg38 or mm10. $n = 1$ for human sperm gDNA. $n = 2$ for human/mouse sperm nscDNA and mouse sperm gDNA. The data represent the mean ± SD. (*C*) Percentages of total mappable reads in sperm gDNA and nscDNA samples aligned with SINE (green), LINE (blue), simple (gold), and satellite (gray) repetitive elements. $n = 1$ for human sperm gDNA. $n = 2$ for human/mouse sperm nscDNA and mouse sperm gDNA. The data represent the mean ± SD. (*D*) Percentages of total mappable sperm gDNA and nscDNA reads aligned with SD tracks in hg38/mm10. $n = 1$ for human sperm gDNA. $n = 2$ for human/mouse sperm nscDNA and mouse sperm gDNA. The data represent the mean ± SD. (*E*) An Integrative Genomics Viewer snapshot of human sperm gDNA and nscDNA coverage peaks (blue) found in chr1 following alignment to hg38. The repetitive elements and SDs associated with this region are displayed using the UCSC Genome Browser. Light to dark gray SD tracks indicate 90 to 98% sequence similarity between SDs. (*F*) Percentages of total mappable human sperm gDNA and nscDNA reads aligned with SD tracks in hg38 after removing the intersections between satellite repeats and SDs. $n = 1$ for human sperm gDNA. $n = 2$ for human sperm nscDNA. The data represent the mean ± SD. (*G*) The ratio of gDNA and nscDNA SD reads to uniquely mapped reads (MAPQ40). The fold changes in the nscDNA to gDNA ratios are indicated. $n = 1$ for human sperm gDNA. $n = 2$ for human/mouse sperm nscDNA and mouse sperm gDNA. The data represent the mean ± SD. (*H*) Percentages of total mapped reads in human sperm gDNA and nscDNA to hg38 (blue) or CHM13 (yellow) human reference genomes. $n = 1$ for human sperm gDNA. $n = 2$ for human sperm nscDNA. The data represent the mean ± SD.

(lncRNA) *PVT1* harbored the highest nscDNA:gDNA coverage ratio (1.3 to 2.0) in the region, which may indicate a hotspot for the origin of small-sized nscDNA.

We extracted the split reads from the *PVT1* region for further analysis. We assumed that a single split read in which both halves map to the same strand harbors a single fusion point of circular DNA (19), and the distance between the coordinates of each end provides the size of the nscDNA (Fig. 4*C*). Nucleotide-level analysis at these fusion points would provide information on how two ends fuse. A total of 52% of fusions were mediated by 1 to 10 bp of microhomology, whereas 40% of fusion events lacked homology. Importantly, there were three events with more than 50 bp segments (82, 114, and 140 bp) exhibiting 77 to 93% homologies. Thus, nscDNA formed using long homology, indicating that duplicated sequences could promote the formation of nscDNA. Since the majority of nscDNA stemming from the surveyed region were less than 3 kb, the region may be a hotspot for small-sized nscDNA formation.

Discussion

Other studies laying the groundwork for eccDNA enrichment have relied on exonucleases for linear DNA depletion and RCA for the amplification of circular DNA (8–12, 15) with a few exceptions (13, 14). eccDNA characterization and classification has been largely based on NGS data analyses of eccDNA, which were focused on investigating uniquely mapped regions of the genome. We developed our method of nscDNA enrichment without RCA to reduce potential limitations caused by in vitro amplification (*SI Appendix*, Fig. S1 *C* and *D*). The purity of our preparations was validated by quality controls (mtDNA/spiked plasmids) and PFGE, which provided us a unique opportunity to uncover the characteristics of eccDNA in the naıve state (nscDNA) in a cell. Importantly, our approach allowed us to quantify the amount of nscDNA in a cell (sperm). By molecular analysis, we estimate that the nscDNA populations we obtained range up to tens of kilobases, the size of which the biological significance has been unclear (27).

The most notable finding of our study is the overrepresentation of SDs in nscDNA, the segments in the human genome that are strongly associated with copy number variations. In human and mouse sperm, the overrepresentation was fourfold and threefold, respectively (human nscDNA: 21.2 ± 1.1% and gDNA: 5.7%, mouse nscDNA:17.6 ± 1.8% and gDNA: 5.6 ± 0.3%) (Fig. 3*D*). This enrichment was not a mapping artifact of the enrichment of satellite DNA, which overlaps with SDs in pericentromeric regions (38). Even after accounting for overlaps with satellite DNA, there was still a threefold enrichment of SDs in nscDNA (human nscDNA: 13.9 ± 0.1% and gDNA: 4.8%) (Fig. 3*F*). With the overrepresentation of these nearly identical SDs and satellite DNA, UMR were depleted. With such dominance of MMR, it is difficult to explicitly define where SD-derived nscDNA originates in the genome based on our current, short sequencing reads–based approaches. A more complete reference genome, such as the CHM13 genome assembly (39), along with ultra-long read sequencing technologies (40–42) could help to define the origins of nscDNA in these structurally diverse areas of the human genome.

Previous studies describe repetitive DNA elements as a significant component of eccDNA (27). As defined by RepeatMasker, we found that satellite DNA was the major constituent of both human and mouse sperm nscDNA at 52.9 ± 1.7 and 81.6 ± 0.6%, respectively (Fig. 3*C*). Satellite DNA accounts for 10% of the human genome and forms the centromeric locus and pericentromeric heterochromatin (43). Satellite DNA is also observed in eccDNA from plants (44, 45), fruit flies (46, 47), mice (48), and humans (49, 50). Most of these findings were obtained by two-dimensional agarose gel electrophoresis, a molecular technique that directly characterizes eccDNA based on size and topology by incorporating labeled probes to identify specific genomic elements. Our studies confirmed these observations by genomic technologies.

In contrast to satellite DNA, SINEs and LINEs were depleted in nscDNA reads from both human (SINEs: 7.1 ± 0.3%, LINEs: 1.5 ± 0.04%) and mouse sperm (SINEs: 0.4 ± 0.05%, LINEs: 0.9 ± 0.2%) (Fig. 3*C*). SINEs and LINEs are groups of transposable elements (TEs) making up 34% (SINEs: 13% and LINEs: 21%) of the human genome (51). Low distributions of Alus, the most common SINE elements (52), were also noted in the sequencing data of microDNA from human maternal plasma that were not in vitro amplified (14). C57BL/6 mouse sperm nscDNA composition differs from that of in vitro–amplified microDNA, which is preferentially derived from full-length LINE-1 retrotransposons compared to various other tissues (9). This suggests that repetitive DNA composition may differ between eccDNA size populations. The depletion of these interspersed elements in nscDNA may be due to the depletion of single-copy regions where SINE and LINE elements scatter. We found an exception to this when analyzing nscDNA from Colo320DM (Fig. 1*E*), which retained high coverage of single-copy genomic regions including 8q24.1, where there are few duplicated segments but many repetitive elements. This suggests that eccDNA may arise from different mechanisms in cancer cells versus normal cells. In cancer cells, large eccDNA species may be generated as a result of genome instability throughout the genome (2).

nscDNA populations from normal human lung and mouse liver tissues were also found to be dominant with SDs (*SI Appendix*, Fig. S6). The presence of nscDNA with similar features in several organs hints that nscDNA likely arises from a general cellular process such as DNA replication in normal cells. The majority of SDs in humans initially formed ∼35 to 40 million years ago during the surge of a younger Alu family of transposable elements, and once formed, SDs became the core of additional rearrangements, leading to regions of complex genome architecture (53). Complex architecture could promote the formation of nscDNA during replication when replication forks stall and break because of tandemly aligned SDs (and satellite DNA) forming secondary DNA structures, and strand invasion of broken DNA would initiate the formation of circular DNA (*SI Appendix*, Fig. S7). Complex genomic regions can be fragile and subject to spontaneous breaks (54).

The fate of SD-bearing nscDNA (formation, reintegration, and degradation) in germ cells could be associated with genetic variations since SDs are subject to meiosis-specific DNA double-strand breaks (55), which would facilitate the integration of nscDNA into chromosomes. A trace of this mechanism can be found within SD clusters of the human genome (56–58). In yeast (59) and bovine (60), circular intermediates have been implicated in chromosomal amplifications or genic translocation events. It will be profound if future research can touch on the possibility of the involvement of circular DNA in genome evolution.

Limitations. Although eccDNA without in vitro amplification would be an improvement in terms of the quantitative assessment of native populations of eccDNA, there could still be biases for particular populations introduced by our procedures. Based on the enrichment of mtDNA from cell lines, we expected that our preparations enriched nscDNA of greater than 10 kb. We aimed to precisely define the size range of nscDNA species using PFGE and Southern blot. However, circular DNA migrates differently than linear DNA during electrophoresis, and the sonicated nscDNA samples included a mixture of circular and linear DNA. Therefore, the nscDNA sizes were difficult to determine and remain elusive. Future

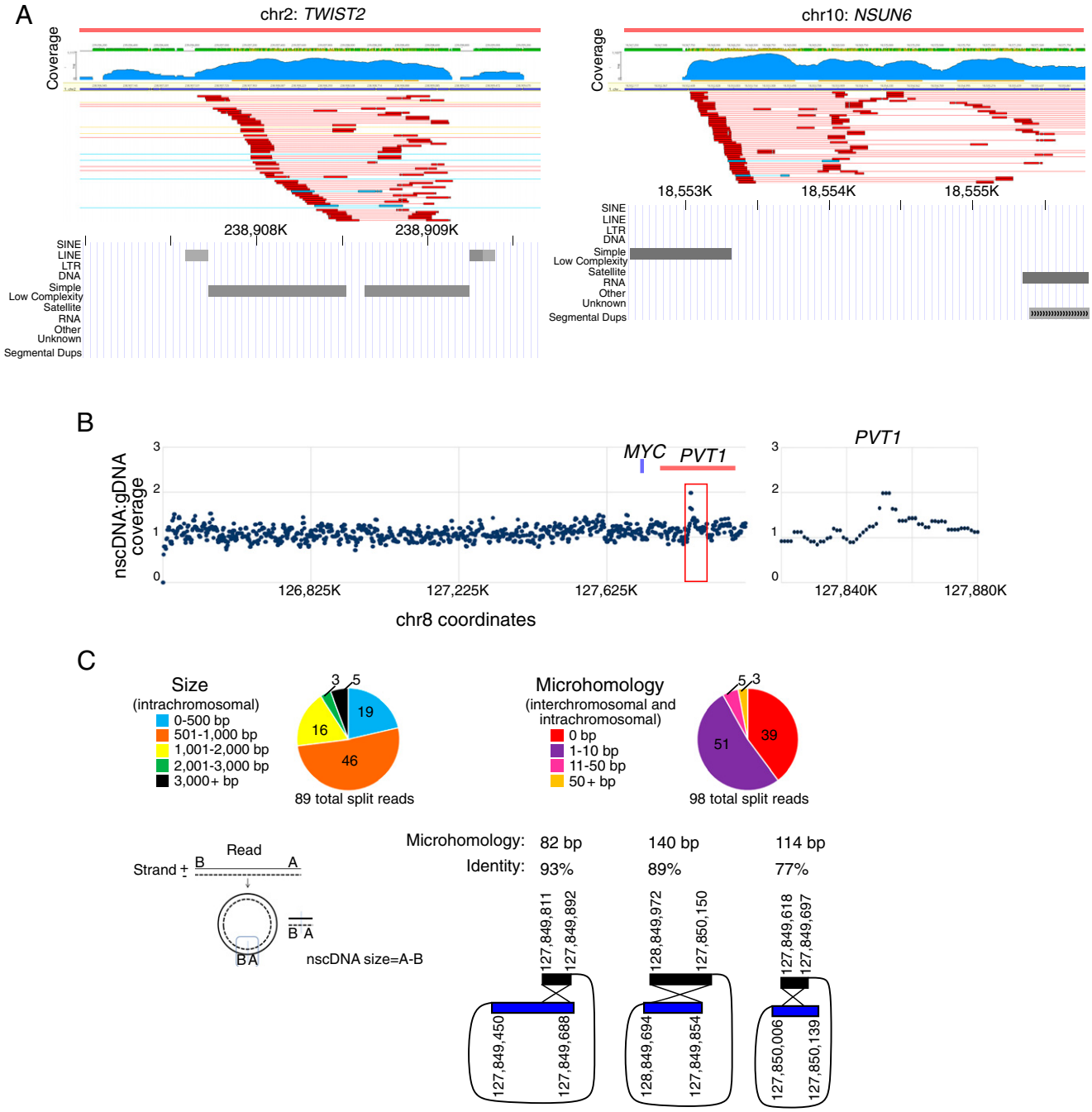

Fig. 4. Fusion junctions in nscDNA. (*A*) Outward-facing paired reads are an indication of circular DNA and cover the fusion point of circles. In the circular orientation, paired reads (red arrows) are facing inward and straddle the fusion point. After fragmentation, library construction, and sequencing, the two reads in the pair map to distant locations in the opposite, outward orientation linked by a red dashed line on hg38. Representative examples of coverage peaks (blue) with linked outward-facing paired reads (red boxes linked by a solid red line) in human sperm nscDNA (hg38 alignment, MAPQ40 reads) are shown for the NSUN6 gene region on chr10 and the TWIST2 gene region on chr2 using Geneious software. The repetitive element and SD contents in these regions are displayed using the UCSC Genome Browser. Light to dark gray SD tracks indicate 90 to 98% sequence similarity between SDs. (*B*) The coverage (excluding reads mapped to chrM) per 1 kb bin (with a five-bin median filter) of the amplified DM (126,425,000 to 128,000,000 kb) locus in Colo320DM nscDNA after normalization to gDNA coverage is shown. A coverage peak consisting of six consecutive bins (127,849,000 to 127,855,000 kb) at the *PVT1* locus is shown. (*C*) Individual reads from Colo320DM nscDNA that partially mapped to the *PVT1* locus in chr8 (hg38) were identified and analyzed for fusion junctions. The junction coordinates were used to calculate the potential sizes of nscDNA. The microhomologies were calculated by nucleotide overlaps at the fusion points with the longest homologies depicted (82, 114, and 140 bp). The junctions between different chromosomes were excluded from the size analysis.

studies employing single-molecule sequencing would provide a better understanding of the size range. Single-molecule sequencing would also provide the detailed structures of each circular DNA: whether each circular DNA consists of either a DNA fragment from a single locus with one circularizing junction (an outwardly oriented read pair in short read sequencing) or a mixture of segments from more than one genomic location.

Materials and Methods

Purification and Enrichment of nscDNA. The Plasmid Midi Kit (Qiagen, 12143) was used to purify circular DNA from cell lines (80×10^6-100×10^6 cells), sperm, and tissue samples. Human sperm nscDNA was extracted from 1 mL pooled semen (Innovative Research, IR100076), which was split into two technical replicates (450 μL/nscDNA replicate, 50 μL for gDNA). Mature mouse sperm was isolated from the epididymis of four C57BL/6 mice (two mice/biological replicate) that were sacrificed at 35 and 33 wk old, respectively. A total of 21×10^6 and 9.5×10^6 sperm were collected from each replicate. gDNA was extracted from 1×10^6 mouse sperm. Liver tissue (370 mg total) from the second pair of mice was extracted and combined before homogenization. A fresh, normal lung tissue sample (frozen) was obtained from the Cedars-Sinai Cancer Biobank and Translational Research Core under protocol Pro00052428 (Circular small DNA as a new cancer biomarker). nscDNA from 50 mg human lung tissue was extracted twice (technical replicates). nscDNA extraction from sperm and tissues required the addition of Proteinase K during cell lysis (500 μg/mL) followed by incubation at 55 °C for 1.5 h (overnight for lung tissue). Sample eluates were phenol-chloroform extracted before DNA precipitation.

The removal of residual, linear double-stranded chromosomal DNA was facilitated by Plasmid-Safe ATP-Dependent DNase (Epicentre, E3101K) digestion reactions. Up to 750 ng purified nscDNA (pre-exonuclease) was digested in 300 μL reactions containing 60 U enzyme and 1 mM ATP and 1× reaction buffer. The reactions were incubated at 37 °C for at least 16 h followed by enzyme inactivation by incubation at 70 °C for 30 min. nscDNA was further purified using DNA Fast Flow PCR Grade centrifugal filters (Microcon, MRCF0R100ET) to exchange buffers and concentrate nscDNA for downstream applications. nscDNA concentration was measured before and after digestion with a Qubit 3.0 fluorometer using the dsDNA HS Assay Kit (Life Technologies, Q32851).

nscDNA and gDNA Library Construction and Sequencing. DNA was fragmented by either Bioruptor Standard (Diagenode) or Covaris M220 sonicators to ~350 bp in length. Libraries were created with the NEBNext Ultra II DNA Library Prep Kit for Illumina (NEB, E7645), starting with 10 ng DNA input from cell line samples. A total of 2 to 5 ng input DNA was used for sperm and tissue samples. The protocol was followed per manufacturer's instructions, bypassing the DNA size selection step. A total of seven PCR cycles were used for the PCR amplification step. The number of libraries sequenced, along with the sequencing platforms, are provided in Dataset S1.

The sequencing reads were trimmed using Trim Galore (v.0.6.1) and Cut Adapt (v2.3) to remove adapters and subsequently aligned to the University of Santa Cruz (UCSC) hg38 human reference genome or mm10 mouse reference genome using Bowtie 2 (v2.3.5). Reads were mapped with unpaired alignment unless otherwise indicated. For paired-read analysis, the orientation of read pairs was determined using the samtools stats function (v1.9). To filter for uniquely mapped reads, only reads with mapping quality scores ≥40 were included. Read depth was normalized using a per-million scaling factor from the total number of mapped reads after filtering. The browser extensible data format files were generated using Bedtools (v2.28.0) and mapped into 1 kb nonoverlapping bins. Coverage maps were visualized using the Interactive Genome Viewer (v2.5.0) (61). The hg38 and mm10 reference genomes were downloaded from the UCSC Genome Browser at http://hgdownload.soe.ucsc.edu/goldenPath/hg38/bigZips/ and http://hgdownload.soe.ucsc.edu/goldenPath/mm10/bigZips/, respectively. The CHM13 reference genome was provided by the Eichler laboratory. The analyzed datasets were uploaded to the Sequence Read Archive (SRA) database with BioProject ID PRJNA641068 (human) and PRJNA655921 (mouse).

Identification of SDs and Repetitive Elements. To gauge the SDs and repetitive DNA contents, repetitive elements in RepeatMasker and SDs (genomicSuperDups) from the UCSC Genome Browser (http://hgdownload.soe.ucsc.edu/goldenPath/hg38/database/ for hg38 and http://hgdownload.soe.ucsc.edu/goldenPath/mm10/database/ for mm10) were intersected with the gDNA and nscDNA sequencing reads. The CHM13 SD tracks were provided by the Eichler laboratory. Reads mapped in 1 kb nonoverlapping bins acted as the control for the repetitive element and SD tracks. For repetitive element analysis, only reads mapping to SINEs, LINEs, Satellites, and simple repeats were considered. Due to the enrichment of circular mtDNA during nscDNA isolation, SDs on nuclear chromosomes with high similarity to those on chrM were excluded from the cell line analysis.

An extended methods section is available in *SI Appendix*.

Data Availability. Sequencing data have been deposited in the SRA at the following accession numbers: PRJNA641068 (62) and PRJNA655921 (63).

ACKNOWLEDGMENTS. We thank Drs. G. V. Boerner and S. Yamada for their critical comments. We also thank the Cedars-Sinai Cancer Applied Genomics, Computation, and Translational Core; the Cedars-Sinai Biobank and Translational Research Core; and the University of California, Los Angeles Electron Imaging Center for NanoMachines for technical support. This work is supported by the National Cancer Institute (2 R01 CA149385 and 1 R03 CA188111-01A1), Department of Defense (W81XWH-18-1-0058), and Cedars-Sinai Medical Center (to H.T.); the Margie and Robert E. Petersen Foundation, the Fashion Footwear Charitable Foundation of New York, Inc., the Avon Foundation, and Associates for Breast and Prostate Cancer Studies (to A.E.G.).

1. Q. Ain, C. Schmeer, D. Wengerodt, O. W. Witte, A. Kretz, Extrachromosomal circular DNA: Current knowledge and implications for CNS aging and neurodegeneration. *Int. J. Mol. Sci.* **21**, 2477 (2020).
2. H. Tanaka, T. Watanabe, Mechanisms underlying recurrent genomic amplification in human cancers. *Trends Cancer* **6**, 462–477 (2020).
3. A. R. Morton *et al.*, Functional enhancers shape extrachromosomal oncogene amplifications. *Cell* **179**, 1330–1341.e13 (2019).
4. S. Wu *et al.*, Circular ecDNA promotes accessible chromatin and high oncogene expression. *Nature* **575**, 699–703 (2019).
5. D. A. Nathanson *et al.*, Targeted therapy resistance mediated by dynamic regulation of extrachromosomal mutant EGFR DNA. *Science* **343**, 72–76 (2014).
6. R. P. Koche *et al.*, Extrachromosomal circular DNA drives oncogenic genome remodeling in neuroblastoma. *Nat. Genet.* **52**, 29–34 (2020).
7. K. M. Turner *et al.*, Extrachromosomal oncogene amplification drives tumour evolution and genetic heterogeneity. *Nature* **543**, 122–125 (2017).
8. Y. Shibata *et al.*, Extrachromosomal microDNAs and chromosomal microdeletions in normal tissues. *Science* **336**, 82–86 (2012).
9. L. W. Dillon *et al.*, Production of extrachromosomal MicroDNAs is linked to mismatch repair pathways and transcriptional activity. *Cell Rep.* **11**, 1749–1759 (2015).
10. J. Zhu *et al.*, Molecular characterization of cell-free eccDNAs in human plasma. *Sci. Rep.* **7**, 10968 (2017).
11. P. Kumar *et al.*, Normal and cancerous tissues release extrachromosomal circular DNA (eccDNA) into the circulation. *Mol. Cancer Res.* **15**, 1197–1205 (2017).
12. H. D. Møller *et al.*, Circular DNA elements of chromosomal origin are common in healthy human somatic tissue. *Nat. Commun.* **9**, 1069 (2018).
13. M. J. Shoura *et al.*, Intricate and cell type-specific populations of endogenous circular DNA (eccDNA) in *Caenorhabditis elegans* and *Homo sapiens*. *G3 (Bethesda)* **7**, 3295–3303 (2017).
14. S. T. K. Sin *et al.*, Identification and characterization of extrachromosomal circular DNA in maternal plasma. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 1658–1665 (2020).
15. P. Mehanna *et al.*, Characterization of the microDNA through the response to chemotherapeutics in lymphoblastoid cell lines. *PLoS One* **12**, e0184365 (2017).
16. J. G. Paez *et al.*, Genome coverage and sequence fidelity of phi29 polymerase-based multiple strand displacement whole genome amplification. *Nucleic Acids Res.* **32**, e71 (2004).
17. A. Fire, S. Q. Xu, Rolling replication of short DNA circles. *Proc. Natl. Acad. Sci. U.S.A.* **92**, 4641–4645 (1995).
18. M. G. Mohsen, E. T. Kool, The discovery of rolling circle amplification and rolling circle transcription. *Acc. Chem. Res.* **49**, 2540–2550 (2016).
19. I. Prada-Luengo, A. Krogh, L. Maretty, B. Regenberg, Sensitive detection of circular DNAs at single-nucleotide resolution using guided realignment of partially aligned reads. *BMC Bioinformatics* **20**, 663 (2019).
20. P. H. Sudmant *et al.*, Global diversity, population stratification, and selection of human copy-number variation. *Science* **349**, aab3761 (2015).
21. M. Y. Dennis, E. E. Eichler, Human adaptation and evolution by segmental duplication. *Curr. Opin. Genet. Dev.* **41**, 44–52 (2016).
22. M. Y. Dennis *et al.*, The evolution and population diversity of human-specific segmental duplications. *Nat. Ecol. Evol.* **1**, 69 (2017).
23. H. D. Møller, L. Parsons, T. S. Jørgensen, D. Botstein, B. Regenberg, Extrachromosomal circular DNA is common in yeast. *Proc. Natl. Acad. Sci. U.S.A.* **112**, E3114–E3122 (2015).
24. S. D. Levene, B. H. Zimm, Separations of open-circular DNA using pulsed-field electrophoresis. *Proc. Natl. Acad. Sci. U.S.A.* **84**, 4054–4057 (1987).
25. S. M. Beverley, Characterization of the 'unusual' mobility of large circular DNAs in pulsed field-gradient electrophoresis. *Nucleic Acids Res.* **16**, 925–939 (1988).
26. D. J. Catanese Jr., J. M. Fogg, D. E. Schrock II, B. E. Gilbert, L. Zechiedrich, Supercoiled minivector DNA resists shear forces associated with gene therapy delivery. *Gene Ther.* **19**, 94–100 (2012).
27. T. Paulsen, P. Kumar, M. M. Koseoglu, A. Dutta, Discoveries of extrachromosomal circles of DNA in normal and tumor cells. *Trends Genet.* **34**, 270–278 (2018).
28. T. Watanabe *et al.*, Impediment of replication forks by long non-coding rna provokes chromosomal rearrangements by error-prone restart. *Cell Rep.* **21**, 2223–2235 (2017).
29. A. L'Abbate *et al.*, Genomic organization and evolution of double minutes/homogeneously staining regions with MYC amplification in human cancer. *Nucleic Acids Res.* **42**, 9131–9145 (2014).
30. R. S. Lasken, T. B. Stockwell, Mechanism of chimera formation during the multiple displacement amplification reaction. *BMC Biotechnol.* **7**, 19 (2007).
31. B. Langmead, S. L. Salzberg, Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).

32. B. Ewing, L. Hillier, M. C. Wendl, P. Green, Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res.* **8**, 175–185 (1998).
33. J. A. Bailey, E. E. Eichler, Primate segmental duplications: Crucibles of evolution, diversity and disease. *Nat. Rev. Genet.* **7**, 552–564 (2006).
34. C. R. Moraes, S. Meyers, The sperm mitochondrion: Organelle of many functions. *Anim. Reprod. Sci.* **194**, 71–80 (2018).
35. J. Vissing, Paternal comeback in mitochondrial DNA inheritance. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 1475–1476 (2019).
36. Y. L. Lyubchenko, L. S. Shlyakhtenko, Visualization of supercoiled DNA with atomic force microscopy in situ. *Proc. Natl. Acad. Sci. U.S.A.* **94**, 496–501 (1997).
37. C. D. Laird, Chromatid structure: Relationship between DNA content and nucleotide sequence diversity. *Chromosoma* **32**, 378–406 (1971).
38. J. A. Bailey, A. M. Yavor, H. F. Massa, B. J. Trask, E. E. Eichler, Segmental duplications: Organization and impact within the current human genome project assembly. *Genome Res.* **11**, 1005–1017 (2001).
39. K. H. Miga *et al.*, Telomere-to-telomere assembly of a complete human X chromosome. *Nature* **585**, 79–84 (2020).
40. E. E. Eichler, R. A. Clark, X. She, An assessment of the sequence gaps: Unfinished business in a finished human genome. *Nat. Rev. Genet.* **5**, 345–354 (2004).
41. S. Louzada *et al.*, Decoding the role of satellite DNA in genome architecture and plasticity-an evolutionary and clinical affair. *Genes (Basel)* **11**, 72 (2020).
42. C. D. Campbell *et al.*, Population-genetic properties of differentiated human copy-number polymorphisms. *Am. J. Hum. Genet.* **88**, 317–332 (2011).
43. S. M. McNulty, B. A. Sullivan, Alpha satellite DNA biology: Finding function in the recesses of the genome. *Chromosome Res.* **26**, 115–138 (2018).
44. S. Cohen, A. Houben, D. Segal, Extrachromosomal circular DNA derived from tandemly repeated genomic sequences in plants. *Plant J.* **53**, 1027–1034 (2008).
45. A. Navrátilová, A. Koblízková, J. Macas, Survey of extrachromosomal circular DNA derived from plant satellite repeats. *BMC Plant Biol.* **8**, 90 (2008).
46. S. Cohen, K. Yacobi, D. Segal, Extrachromosomal circular DNA of tandemly repeated genomic sequences in Drosophila. *Genome Res.* **13**, 1133–1145 (2003).
47. S. Cohen, N. Agmon, K. Yacobi, M. Mislovati, D. Segal, Evidence for rolling circle replication of tandem genes in Drosophila. *Nucleic Acids Res.* **33**, 4519–4526 (2005).
48. Z. Cohen, E. Bacharach, S. Lavi, Mouse major satellite DNA is prone to eccDNA formation via DNA Ligase IV-dependent pathway. *Oncogene* **25**, 4515–4524 (2006).
49. S. Cohen, D. Segal, Extrachromosomal circular DNA in eukaryotes: Possible involvement in the plasticity of tandem repeats. *Cytogenet. Genome Res.* **124**, 327–338 (2009).
50. S. Cohen, N. Agmon, O. Sobol, D. Segal, Extrachromosomal circles of satellite repeats and 5S ribosomal DNA in human cells. *Mob. DNA* **1**, 11 (2010).
51. A. M. Weiner, SINEs and LINEs: The art of biting the hand that feeds you. *Curr. Opin. Cell Biol.* **14**, 343–350 (2002).
52. D. Carnevali, A. Conti, M. Pellegrini, G. Dieci, Whole-genome expression analysis of mammalian-wide interspersed repeat elements in human cell lines. *DNA Res.* **24**, 59–69 (2017).
53. J. A. Bailey, G. Liu, E. E. Eichler, An Alu transposition model for the origin and expansion of human segmental duplications. *Am. J. Hum. Genet.* **73**, 823–834 (2003).
54. R. Suzuki *et al.*, The fragility of a structurally diverse duplication block triggers recurrent genomic amplification. *Nucleic Acids Res.* **49**, 244–256 (2020).
55. J. Lange *et al.*, The landscape of mouse meiotic double-strand break formation, processing, and repair. *Cell* **167**, 695–708.e16 (2016).
56. J. U. Chicote *et al.*, Circular DNA intermediates in the generation of large human segmental duplications. *BMC Genomics* **21**, 593 (2020).
57. K. K. Takahashi, H. Innan, Duplication with structural modification through extrachromosomal circular and lariat DNA in the human genome. *Sci. Rep.* **10**, 7150 (2020).
58. L. Pu, Y. Lin, P. A. Pevzner, Detection and analysis of ancient segmental duplications in mammalian genomes. *Genome Res.* **28**, 901–909 (2018).
59. I. Prada-Luengo *et al.*, Replicative aging is associated with loss of genetic heterogeneity from extrachromosomal circular DNA in Saccharomyces cerevisiae. *Nucleic Acids Res.* **48**, 7883–7898 (2020).
60. K. Durkin *et al.*, Serial translocation by means of circular intermediates underlies colour sidedness in cattle. *Nature* **482**, 81–84 (2012).
61. H. Thorvaldsdóttir, J. T. Robinson, J. P. Mesirov, Integrative genomics viewer (IGV): High-performance genomics data visualization and exploration. *Brief. Bioinform.* **14**, 178–192 (2013).
62. L. Mouakkad-Montoya *et al.*, data at BioProject ID: PRJNA641068. Sequence Read Archive. https://www.ncbi.nlm.nih.gov/bioproject/term=PRJNA641068. Deposited 22 June 2020.
63. L. Mouakkad-Montoya *et al.*, data at: BioProject ID: PRJNA655921. Sequence Read Archive. https://www.ncbi.nlm.nih.gov/bioproject/PRJNA655921. Deposited 7 August 2020.