A classical revival: Human satellite DNAs enter the genomics era

Nicolas Altemose[1*]

[1] Department of Molecular & Cell Biology, University of California, Berkeley, CA, USA
* To whom correspondence should be addressed: altemose@berkeley.edu
345 Weill Hall, UC Berkeley, Berkeley, CA, USA 94720

Abstract

The classical human satellite DNAs, also referred to as human satellites 1, 2 and 3 (HSat1, HSat2, HSat3, or collectively HSat1-3), occur on most human chromosomes as large, pericentromeric tandem repeat arrays, which together constitute roughly 3% of the human genome (100 megabases, on average). Even though HSat1-3 were among the first human DNA sequences to be isolated and characterized at the dawn of molecular biology, they have remained almost entirely missing from the human genome reference assembly for 20 years, hindering studies of their sequence, regulation, and potential structural roles in the nucleus. Recently, the Telomere-to-Telomere Consortium produced the first truly complete assembly of a human genome, paving the way for new studies of HSat1-3 with modern genomic tools. This review provides an account of the history and current understanding of HSat1-3, with a view towards future studies of their evolution and roles in health and disease.

Keywords: satellite DNA, repetitive DNA, tandem repeats, classical human satellites, HSATI, HSATII, HSATIII, HSat1, HSat2, HSat3

1. Introduction

Satellite DNA refers to long arrays of tandemly repeated sequences that make up a major component of many eukaryotic genomes. Within each satellite repeat array, nearly identical DNA sequences are repeated head-to-tail, over and over, often encompassing millions of base pairs. In humans, satellite DNA represents 5-10% of the genome [1–3], found primarily at all centromeres and pericentromeric regions, along the short arms of the acrocentric chromosomes (chromosomes 13, 14, 15, 21, and 22), and on the long arm of the Y chromosome. These satellite-rich, gene-poor regions are typically organized into a chromatin compartment referred to as constitutive heterochromatin [4], which is characteristically more condensed than euchromatic regions in interphase [5]. Satellite DNA arrays can be classified into distinct families based on both their sequence composition, which is often AT-rich overall, and on their characteristic repeat unit lengths, which range from several bases to kilobases. In the human genome, the largest satellite family by total size is alpha satellite DNA (αSat), which encompasses every centromere and plays a key role in centromere function [6,7]. The next largest families by total size, which constitute the largest individual satellite arrays in the genome, are referred to collectively as the classical human satellites, or more specifically as human satellites 1, 2, and 3 (HSat1-3) [8].

Despite the fact that the classical human satellites were among the first human DNA sequences to be biochemically characterized and isolated from the rest of the nuclear genome [9,10] they remain poorly understood, in part because satellite DNA arrays were intractable to early genome sequencing and assembly methods [11]. This intractability stems from the fact that DNA sequencing technologies can only determine the sequence of relatively small fragments of DNA. To assemble a whole genome from these small fragments, assembly algorithms must identify unique overlaps between sequencing reads and stitch them together. This approach works well in regions of the genome containing mostly unique sequences, but it often fails in regions where the same DNA sequence is found repeated over and over, especially when the distance between unique sequences is much larger than the sequencing read length. The sequencing technologies available to the Human Genome Project produced fairly short sequencing reads (~1 kb). This made the problem of assembling satellite-rich regions so insurmountable that they were excluded from the final reference sequence, despite the fact that many satellite-rich regions were known to have important functions. As a result, HSat1-3 and other satellite DNAs have been largely left behind in the genomics era.

Recently, new long-read DNA sequencing and assembly methods enabled the Telomere-to-Telomere (T2T) Consortium to assemble across all satellite DNA arrays in a human genome for the first time [3,12]. This complete reference sequence, along with new long-read approaches for studying epigenetic regulation [13–16], have now opened up HSat1-3 to new discoveries regarding their function, variation, and evolution. To help facilitate these future studies, here, I provide a comprehensive survey of our current understanding of the large and mysterious classical human satellites, HSat1-3.

2. Discovery and initial characterization of the classical human satellites

2.1 Early separation techniques isolate a highly repetitive fraction of human DNA

The term "satellite DNA" originates from early biochemical experiments in which genomic DNA preparations were separated by their base composition using cesium density gradient ultracentrifugation approaches [17–19] (Figure 1a). The concentration of DNA along the cesium gradient could be measured optically, and typically the genetic material would appear as a unimodal, contiguous band of a characteristic density for each species, depending on its overall A/T vs G/C base composition [19]. However, in 1961 Saul Kit discovered that mouse and guinea pig DNA formed a second, smaller, "satellite" DNA band outside the major DNA band [18] (*footnote 1*), and a similar, AT-rich satellite DNA band was later discovered in humans [9]. Further methodological advances achieved finer resolution of DNA fractions by base composition, revealing the presence of additional satellite DNA bands in humans, which were labeled as human satellite fractions I-III [9,10,20] (Figure 1; *footnote 2*). Careful renaturation experiments revealed that DNA isolated from the satellite fraction re-annealed much more quickly after denaturation compared to DNA isolated from the main genomic fraction, consistent with the satellite fraction being composed primarily of repetitive DNA sequences [10,20–23]. Satellite DNAs were found to be enriched in heterochromatic fractions of DNA obtained by centrifugation of chromatin [4,10], and in situ hybridization experiments revealed that satellite DNAs are enriched in pericentromeric heterochromatin [24–27], with the largest blocks in humans found on chromosomes 1, 9, 16, and Y [24,28–33].

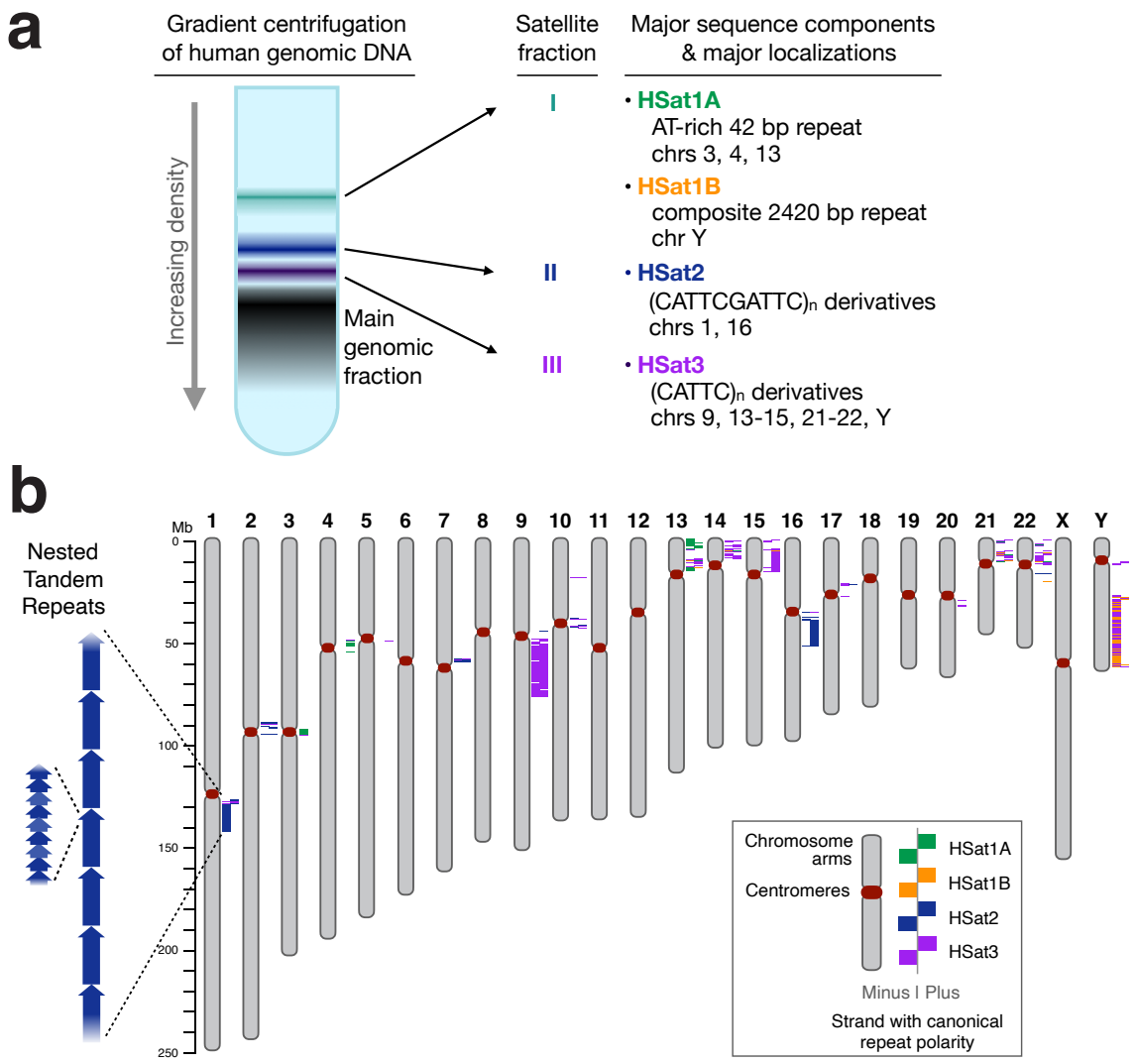

Figure 1. The origins of the classical human satellites. a) A schematic depicts early experiments in which high-molecular-weight human genomic DNA was fractionated by ultracentrifugation in cesium gradients. Three robust "satellite" fractions (I, II, and III) could be separated from the main genomic fraction, each of which contained mixtures of repetitive DNA sequences with different average sequence compositions relative to the bulk of the genome. The predominant components of each satellite fraction, dubbed human satellites 1-3 (HSat1-3), were later mapped by in situ hybridization, and fragments were cloned and sequenced. Below each satellite name is a description of its general sequence characteristics and major chromosomal localizations (chromosomes containing >1 Mb of that satellite). **b**) Ideograms of all human chromosomes in T2T-CHM13v2.0 showing the overall distribution of HSat1-3 across the genome. The schematic on the left illustrates the nested tandem repeat structure typical of HSat1-3 arrays. To the right of each ideogram are two tracks with rectangles representing the span of each HSat1-3 array (arrays smaller than 10 kb are not shown), colored by the family they represent. The left track shows regions where the canonical repeat orientation is found on the minus strand, while the right track shows those found on the plus strand. Arrays with inversions have representation on both strands. Plot was generated with karyoploteR [164].

2.2 Molecular biology techniques identify new classes of tandem repeats

Compared to the non-repetitive components of the genome, the abundance and repetitive nature of satellite DNA made it easier to study with early molecular biology methods like degradation-based DNA sequencing, Southern blotting, and in situ hybridization [25,26,34–36]. Research into satellite DNA benefited immensely from the advent of analytical methods using restriction enzyme digestion followed by electrophoretic separation of DNA fragments by size [7,35,37–39]. In tandem repeats, restriction sites (typically 4-6 bp) often occur only once within each repeat unit, so restriction enzyme digestion will tend to release DNA fragments with lengths equal to the periodicity of the repeat, or its multiples (e.g. when repeat copies carry mutations in the restriction site) [35]. Alternatively, a particular restriction site may occur frequently throughout the genome, but never within a particular satellite array, making it possible to determine total array sizes by pulsed-field gel electrophoresis and Southern blotting [40,41]. In non-repetitive regions, restriction fragment lengths are expected to follow a fairly continuous exponential distribution. By separating and quantifying digested DNA fragments by size, one can observe discrete bands that correspond to tandem repeats and are distinguishable from the continuous smears formed by non-repetitive DNA [39]. With these approaches, the sizes of some of the major repeating units in each satellite fraction could be determined [8,38,42–48], and early work by Howard J. Cooke demonstrated that different repeating units could be found on different chromosomes [38,45,48,49].

Furthermore, restriction digest methods enabled the discovery of new classes of tandemly repeated sequences that could not be separated from the main genomic fraction on cesium density gradients. For example, αSat DNA was not separable from the main genomic fraction but was detected by Laura Manuelidis using restriction enzyme digestion [7]. Although these newly discovered repetitive sequences did not belong to true satellite DNA fractions, they were eventually also referred to as satellite DNA, broadening the definition of the term to include any short tandem repeats occurring in long arrays [50,51]. The sequences in satellite fractions I-III are often referred to as the classical human satellites, to distinguish them from later waves of satellite repeat discovery [51].

2.3 Individual repeat classes are cloned and sequenced

Recombinant DNA and more efficient sequencing technologies [52,53] enabled the discovery of new satellite repeats. Molecular cloning allowed for the isolation and amplification of individual DNA fragments from the complex mixtures of sequences comprising satellite fractions I-III. These fragments were then characterized by analytical digestion, Southern blotting, Sanger sequencing, and in situ hybridization. Foundational work by Jane Prosser and Marianne Frommer, then in the lab of Paul C. Vincent, uncovered the major sequence components of satellite fractions I-III [8,43,44] (Figure 1). They revealed that satellite I, the most AT-rich fraction of the genome, could be subdivided into two unrelated sequence families: a simple 42 bp tandem repeat [8], and a 2.5 kb repeat found predominantly on the Y chromosome [38,43,48]. Satellite fractions II and III were both found to be derived predominantly from a tandem repeat of the pentamer "CATTC," although satellite II sequences appeared to be older and more diverged [8,44]. Because satellite fractions are complex mixtures that can differ by preparation, Prosser et al. suggested naming the specific repetitive DNA families within each satellite fraction using Arabic numerals, while the satellite fractions themselves would retain Roman numerals. For example, human satellite 3 refers to the repeat family that constitutes the majority of human satellite fraction III [8]. I honor this convention here and propose that the disparate naming schemes in the literature be unified moving forward.

Human satellites 1-3 can be abbreviated as HSat1-3, and the two components of HSat1 can be distinguished as HSat1A (the 42-bp simple repeat) and HSat1B (the 2.5 kb repeat predominantly on chrY) [3].

2.4 Satellite DNAs are broadly mapped across the genome

Later fluorescence in situ hybridization (FISH) experiments with DNA oligonucleotides revealed that the 42 bp HSat1A repeat is predominantly found near the centromeres of chromosomes 3 and 4 and on the long arm of chromosome 13 [30,54]. HSat2 probes hybridized primarily to chromosomes 1 and 16, with smaller domains on chromosomes 2, 7, 10, 15, 17, and 22 [31]. HSat3 probes hybridized primarily to chromosome 9, with smaller domains on chromosomes 1, 5, 10, 17, 20, and the acrocentrics (13, 14, 15, 21, 22) [31]. Other studies established that a 3.6 kb HSat3 repeat, along with the 2.5 kb HSat1B repeat, comprised the majority of the Y chromosome, on the q arm, where they are interspersed in large blocks [38,48,49,55–59]. Additional studies isolated clones of HSat2 and HSat3 that hybridized to individual chromosomes or subsets of chromosomes [45,47,60–71]. Classical satellite DNA probes were sometimes used as chromosome-specific markers in FISH studies, given the ease of labeling and visualizing large satellite DNA arrays [72].

Prior to the Human Genome Project, this remained the state of our understanding of the classical human satellites. Their approximate locations in the genome were known at metaphase-chromosome-scale resolution, and a few dozen clones were sequenced as representatives of their respective arrays. Most human satellite DNA research shifted to alpha satellite DNA once it was determined to be associated with centromere function, while the potential functions of the classical satellites remained poorly understood.

3. HSat1-3 in the Genomics Era

3.1 Satellite DNAs are excluded from the Human Genome Project

The repetitive, heterochromatic regions of the genome posed an intractable problem for the Human Genome Project [73] and for Celera Genomics' separate efforts to assemble the human genome [74]. The Human Genome Project decided to use a hierarchical sequencing approach, in which large genomic fragments were cloned and physically mapped along each chromosome. Each fragment was then sheared, Sanger sequenced, and assembled from sequencing reads roughly 1 kb in length. The repetitive nature of satellite DNA caused several problems for this approach: 1) large repetitive regions often could not be cloned efficiently or would undergo structural rearrangements; 2) repetitive DNA clones could not be physically mapped with the same precision as unique DNA clones, because, for example, their FISH probes would hybridize to multiple loci; and 3) it remained challenging to assemble repetitive DNA from short sequencing reads even within a single cloned genomic fragment, as near-identical repeat units often exceeded the read length [11]. Although Celera genomics used a different, Whole Genome Shotgun (WGS) sequencing approach, they still relied on cloning of individual DNA fragments and thus were also potentially susceptible to issues with repetitive DNA amplification. Furthermore, WGS approaches faced the even more difficult task of assembling repetitive regions from sequenced fragments across the entire genome. Thus, neither human genome sequencing effort, nor any that followed for 20 years, succeeded in assembling across large HSat1-3 arrays in the human genome, leaving their approximate locations in the genome assembly as enormous gaps filled with placeholder "N" characters.

3.2 Small amounts of partially assembled satellite DNA are characterized

Occasionally, the human genome assembly would include the very edge of a classical satellite array adjacent to a large gap, and some limited information could be gleaned from these sequences [75,76] (*footnote 3*). Whole-genome shotgun assemblies would often produce partial, unmapped, scaffold assemblies of subregions of satellite arrays [77]. These unmapped scaffolds could sometimes be mapped to specific chromosomes using chromosome-specific sequencing libraries [75], and those that contained unique marker sequences could often be mapped to specific loci using 3D contact information [78] or a clever approach based on genetic information from individuals with recently admixed ancestry [79]. More focused approaches identified additional HSat2 and HSat3 arrays on individual chromosomes [80–83]. However, these approaches still failed to represent the classical human satellites comprehensively.

3.3 HSat1-3 fragments are studied comprehensively using unassembled reads

To address this, in a 2014 study, we developed an alignment-free approach for characterizing all of the HSat2 and HSat3 sequences from a single individual's raw WGS reads [75]. To do so, we converted the sequence of each ~1 kb long HSat2 or HSat3 read in the HuRef genome [77] into a vector of frequencies of every possible 5-mer. Then, by also leveraging paired-read information, we iteratively clustered these read vectors based on their sequence composition and physical proximity, which allowed us to identify broad sequence subfamilies. This approach yielded 11 HSat3 and 3 HSat2 subfamilies, which were then localized to chromosomes using published sequencing data from flow-sorted chromosomes. Subfamilies were named alphanumerically, as HSat2A1-2, HSat2B, HSat3A1-6, and HSat3B1-5. Using these subfamily clusters, we could identify 24-bp sequences that were represented frequently and specifically within each subfamily, yielding a 'pseudoreference' that could be used to identify these satellite DNA sequences even within shorter next-generation sequencing reads [75]. This served as the most comprehensive inventory of HSat2 and HSat3 sequences at the time, although it was also subject to the potential amplification biases used in generating the HuRef clone libraries.

3.4 All HSat1-3 arrays are completely assembled in a human genome

In 2021, the Telomere-to-Telomere (T2T) Consortium released the first complete assembly of a human genome, including all autosomal HSat1-3 arrays [3,12]. This genome originated from CHM13, a diploid cell line derived from a hydatidiform mole, which contains two copies of the paternal haplotype, making it homozygous essentially everywhere and eliminating the challenge of haplotype phasing when assembling the most repetitive regions of the genome. Impressively, this effort fully spanned the largest satellite arrays of any kind in the genome: a 27.6 Mb HSat3 array on chr9, a 13.2 Mb HSat2 array on chr1, a 12.7 Mb HSat2 array on chr16, a 7.5 Mb HSat3 array on chr15, and a 5 Mb HSat1A array on chr13 (Table 1; for reference, the largest αSat array in the genome is 4.8 Mb, on chr18). This represents an enormous improvement over the hg38 reference assembly; the total amount of HSat2 on the chromosomes increased from 0.87 Mb in hg38 to 28.7 Mb in T2T-CHM13, while the total amount of HSat3 increased from 0.14 Mb to 47.7 Mb. The Y chromosome is not present in the CHM13 cell line, although a T2T assembly of chrY from a diploid cell line, HG002, has now been released, which includes 21.7 Mb of HSat3 and 14.2 Mb of HSat1B (Figure 1b, Table 1).

Table 1. Properties of the largest HSat1-3 arrays. For all 16 distinct HSat1-3 arrays that are larger than 500 kb, a list of their coordinates (in T2T-CHM13v2.0), lengths, families, subfamilies from [75], number of inversion breakpoints in the array, repeat unit lengths detected by NTRprism [3], ratio of the array length to the number of unique 24-mers in the array ("24-mer Fold Compression", a measure of repetitiveness), and unique array index (format: chr_family_number). Note: for the large array on chr9, no predominant repeat period larger than the ancestral 5 bp repeat was detected by NTRprism. A complete version of this table including smaller arrays, along with repeat consensus sequences, are available at github.com/Altemose/HSatReview.

Chr	Start (Mb)	End (Mb)	Length (Mb)	Family	Subfam.	Inv. Breaks	NTRprism Repeat Unit Lengths (bp)	24-mer Fold Compression	Array Index
chr1	129.03	129.64	0.62	HSat2	A1	0	1407, 2729, 1322	11.72	1_2_8
chr1	129.64	142.17	12.53	HSat2	A2	0	1776, 1308	93.95	1_2_9
chr2	90.99	91.61	0.62	HSat2	B	1	6571, 4671	6.78	2_2_4
chr3	92.90	95.37	2.46	HSat1A	-	0	378, 751, 42	39.39	3_1A_2
chr4	50.43	52.12	1.68	HSat1A	-	0	42	36.20	4_1A_2
chr9	49.06	76.69	27.64	HSat3	B5	237	(5)	41.67	9_3_3
chr13	0.07	4.97	4.91	HSat1A	-	4	378, 42	26.45	13_1A_1
chr13	13.94	15.55	1.61	HSat1A	-	0	3013, 6330, 42, 3474	44.46	13_1A_3
chr15	6.36	13.91	7.55	HSat3	A5	0	1794	65.42	15_3_12
chr16	39.52	52.22	12.70	HSat2	B	2	6112, 3194	56.96	16_2_10
chr20	32.02	32.89	0.87	HSat3	B3	0	6905	5.99	20_3_2
chr21	1.24	1.95	0.71	HSat1A	-	0	378, 751, 42	14.64	21_1A_1
chr22	2.62	3.50	0.89	HSat1A	-	0	378, 751, 42	15.32	22_1A_1
chrY	11.71	12.66	0.94	HSat3	A3	0	1373, 5540, 3811	4.38	Y_3_4
chrY	27.81	62.00	20.01	HSat3	A6	4	3554, 2758	74.25	Y_3_6
chrY	28.66	62.03	14.21	HSat1B	-	4	2420	56.70	Y_1B_1

We identified the subfamily components of each HSat2 and HSat3 array in this new reference and compared their localization with previous predictions [3] (Table 1). The assembly confirmed the chromosomal assignments predicted previously in [75] and identified several novel arrays beyond the resolution of previous methods. For example, we identified that the B1 subfamily of HSat3 was contained almost entirely in a previously undescribed array on chr17. Furthermore, the assembly confirmed that HSat2 subfamilies A1 and A2 represent distinct subdomains within the HSat2 array on chr1, revealing that 2A1 is the smaller and more centromere-proximal of the two but shares a boundary with 2A2. We could also compare the sequence relationships between different arrays using an alignment-free, k-mer based approach (Figure 2, [3,75]). This can reveal which arrays have the most closely related sequences, consistent with recent duplication and/or homogenization of the arrays, as well as which arrays are the most distinct (e.g. the large HSat3A6 array on chrY, Figure 2).

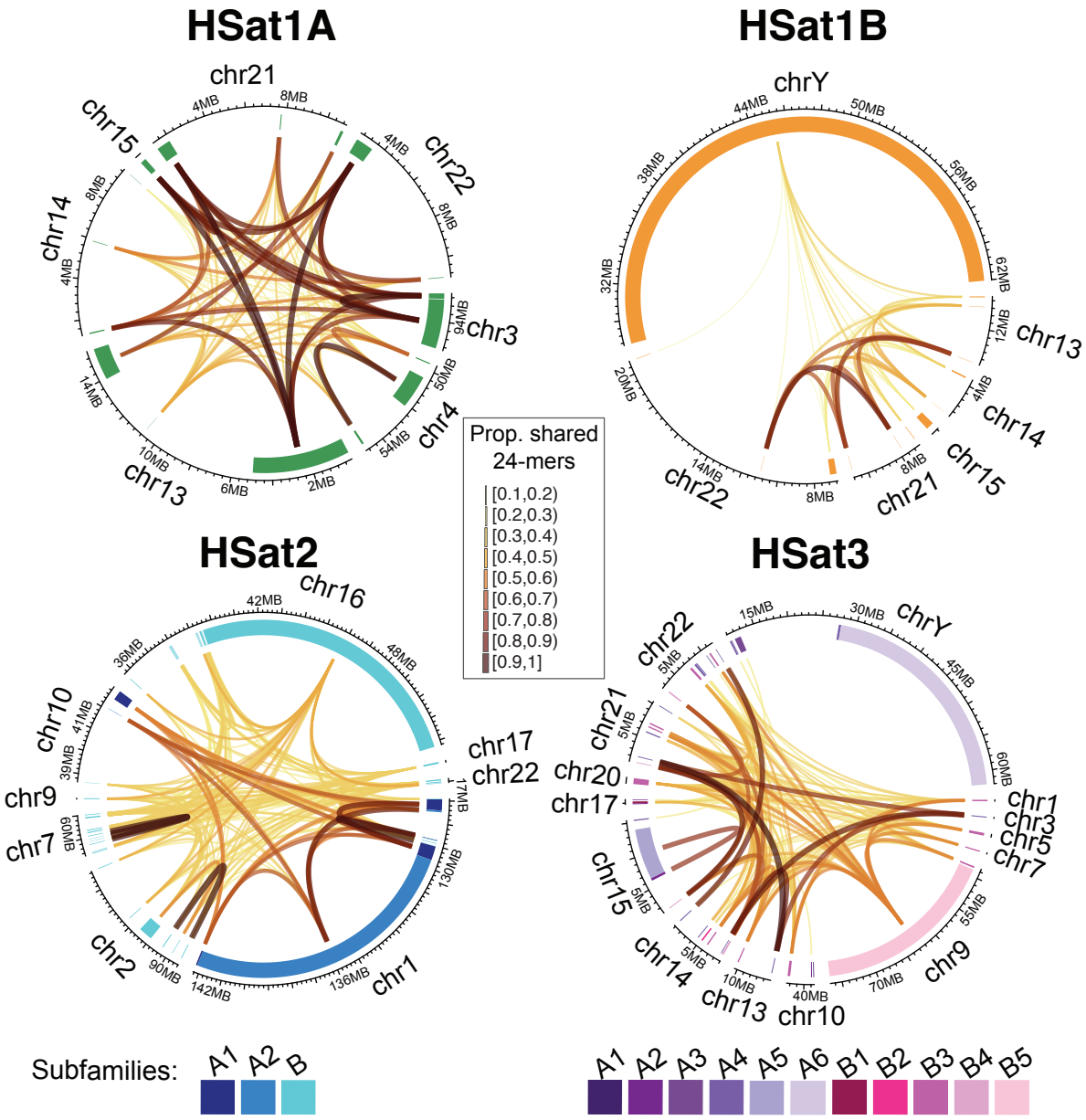

Figure 2. Annotation of HSat1-3 in a complete T2T genome assembly. Circos plots show the proportion of 24-mers shared between HSat1-3 arrays of the same family. The outer track depicts the size and coordinates of each array, and for HSat2 and HSat3, arrays are colored by subfamily assignment. Only arrays larger than 10 kb are shown for HSat1A, HSat1B, and HSat2, and only arrays larger than 100 kb are shown for HSat3. The lines connecting the arrays are scaled and colored according to the proportion of exact 24-mer sequence matches between the arrays, serving as an overall estimate of sequence similarity (no line is drawn below a fixed threshold of 0.1 for HSat1A and HSat1B, and 0.25 for HSat2 and HSat3). Specifically, for two arrays A and B for which A is longer than B, this represents the proportion of all 24 bp substrings in B that can each be matched exactly with a 24 bp substring in A (e.g. if a particular 24-mer is present 5 times in A and 4 times in B, both the numerator and denominator of the proportion would be incremented by 4). Plots were generated with circlize [165]. Source data for this plot are provided at github.com/altemose/HSatReview

The CHM13 assembly also revealed unexpected structural phenomena involving HSat1-3. For example, the active, centromere-containing αSat arrays on chromosomes 3 and 4 are interrupted by large HSat1A arrays (1.7 and 2.7 Mb, respectively [3]). Furthermore, the large HSat3B5 array on chr9 contains 237 inversion breakpoints, a phenomenon never previously described for this array or any other HSat1-3 arrays (Figure 1b, [3]). It remains unknown whether this highly inverted array organization is fixed or polymorphic, whether it arose incrementally or by a mechanism like chromothripsis, or whether these inversions play any sort of functional role, such as in further suppressing recombination in this region. The successful assembly of large HSat1-3 arrays also enabled analyses not possible with unassembled sequencing reads, such as examining long-range differences in sequence homogeneity across entire arrays (discussed further in section 4). Furthermore, we could examine repeat periodicity and its variation across arrays. To do so, we developed an algorithm, Nested Tandem Repeat Prism (NTRprism), which can uncover repeat periodicities in long tandem repeat arrays ([3], *footnote 4*). For example, NTRprism uncovered different repeat periodicities in the adjacent HSat2A1 (1.4 kb) and HSat2A2 (1.8 kb) arrays on chr1, and it revealed different periodicities in different subregions of the HSat3 array on chr20 [3].

Although the CHM13 assembly sequence only represents one human haplotype, it provides an important point of reference against which one can compare reads and assembly scaffolds from other genomes, shining light on the diversity and evolution of these newly added satellite DNA sequences. This reference also provides an atlas to guide future studies of the regulation and function of HSat1-3 arrays using modern tools, allowing researchers to revisit and test old hypotheses about the roles of HSat1-3 in healthy and diseased cells.

4. Evolution and variation of HSat1-3

4.1 Satellite array size polymorphisms are observed by comparing karyotypes

Before the discovery of satellite DNA, it was known from studies of banding patterns on human karyotypes that large heterochromatic blocks could differ enormously between individuals, and even between the homologous chromosomes in one individual [24,84]. These effects could be so dramatic as to visibly change the size of entire chromosomes, such as the Y chromosome [85–90]. Once satellite DNA was associated with heterochromatin, it became clear that something about the evolution of satellite DNA must explain these enormous size polymorphisms [55,84,91]. Because large heteromorphisms in HSat1-3 arrays were frequently present in individuals without congenital diseases or other health issues [55,56,68,89,92–95], and because HSat1-3 are not universally present on every chromosome, it was often believed that these satellites likely do not play an essential function in the cell, such as guiding centromere function [24]. The high degree of polymorphism also suggested a high rate of structural rearrangements in these regions, which could not be explained by the same molecular mechanisms responsible for variation in the rest of the genome.

4.2 Unequal crossover is hypothesized to explain array size polymorphisms

The favored hypothesis for how human satellite DNA arrays evolve, which was suggested by Ann P. Craig-Homes and Margery W. Shaw in 1971, became evolution by unequal crossover [84,92,96] (Figure 3a). That is, when a double-strand break occurs during the S or G2 phases of the cell cycle, the predominant mode of repair is homologous recombination (HR) from the sister chromatid.

This process involves a homology search step in which a short stretch of DNA adjacent to the break must find an exact sequence match along the sister chromatid. Once found, the repair process often results in a crossover, in which the two distal chromosome arms are exchanged near the break site. In repetitive DNA, the cell runs into the same problem as genome mapping or assembly algorithms: during homology search, the short stretch of break-adjacent DNA can encounter many exact sequence matches, and the correct homologous stretch cannot be determined. As a result, sometimes the crossover occurs between the break site and a site at a different position within the sister chromatid's satellite array, producing two recombinant chromatids of different lengths: one long, and one short (Figure 3a). If unequal crossover events occur in the germline, these expanded or contracted arrays can be passed on to offspring. Occasionally ectopic crossovers can occur between classical human satellite arrays on different chromosomes, yielding translocations that can cause congenital diseases or contribute to cancer [70,83,97–99].

Additional support for the unequal crossover hypothesis stemmed from theoretical predictions and experimental observations of tandem ribosomal gene arrays in yeast [91,100–104]. Mitotic, rather than meiotic, crossover is the favored mechanism because crossovers are known to be suppressed in heterochromatin during meiotic recombination [92,105]. As an alternative to crossovers, sites of HR can resolve as gene conversions, in which the sister chromatid is used as a template for synthesis without an exchange of chromosome arms. This may contribute to the homogenization of satellite arrays both within and between chromosomes, a phenomenon referred to as concerted evolution by molecular drive ([106] and reviewed by [107]). Alternative proposed satellite expansion/contraction mechanisms involve RNA-derived intermediates [108], rolling circle amplification and re-integration [109], or break-induced replication at stalled replication forks [110,111].

4.3 High rates of structural rearrangements are observed in satellite arrays

By comparing restriction digest patterns between father/son pairs, one study measured the rate of detectable mutations in the heterochromatic long arm of the Y chromosome, which is composed primarily of interspersed HSat1B and HSat3A6 arrays. They found at least one structural rearrangement per 40 Mb per meiosis, which is one of the highest mutation rates ever reported in the genome [59,112]. A comparison of HSat3A6 array size estimates on chrY from short-read WGS sequencing data across 396 individuals found a large range of size variation, from 7 Mb to 98 Mb [75]. Similarly, estimates of the total amount of HSat2 and HSat3 in the genome varied widely across hundreds of individuals (1-7% of the genome, combined; 2.1% median, compared to 2.5% in CHM13) [1]. Unexpectedly, chr1 in CHM13 lacked the predicted HSat3B2 subfamily, prompting us to investigate this centromere in partial assemblies from 16 diploid individuals (32 haplotypes, 27 of which were sufficiently assembled in this region of the genome), revealing that a ~400 kb HSat3B2 array was variably present in the pericentromeric region of chr1 (deleted in 29% of ascertainable haplotypes) [3]. This is similar to an HSat3 array previously shown to have variable presence on chr14 in different individuals [68].

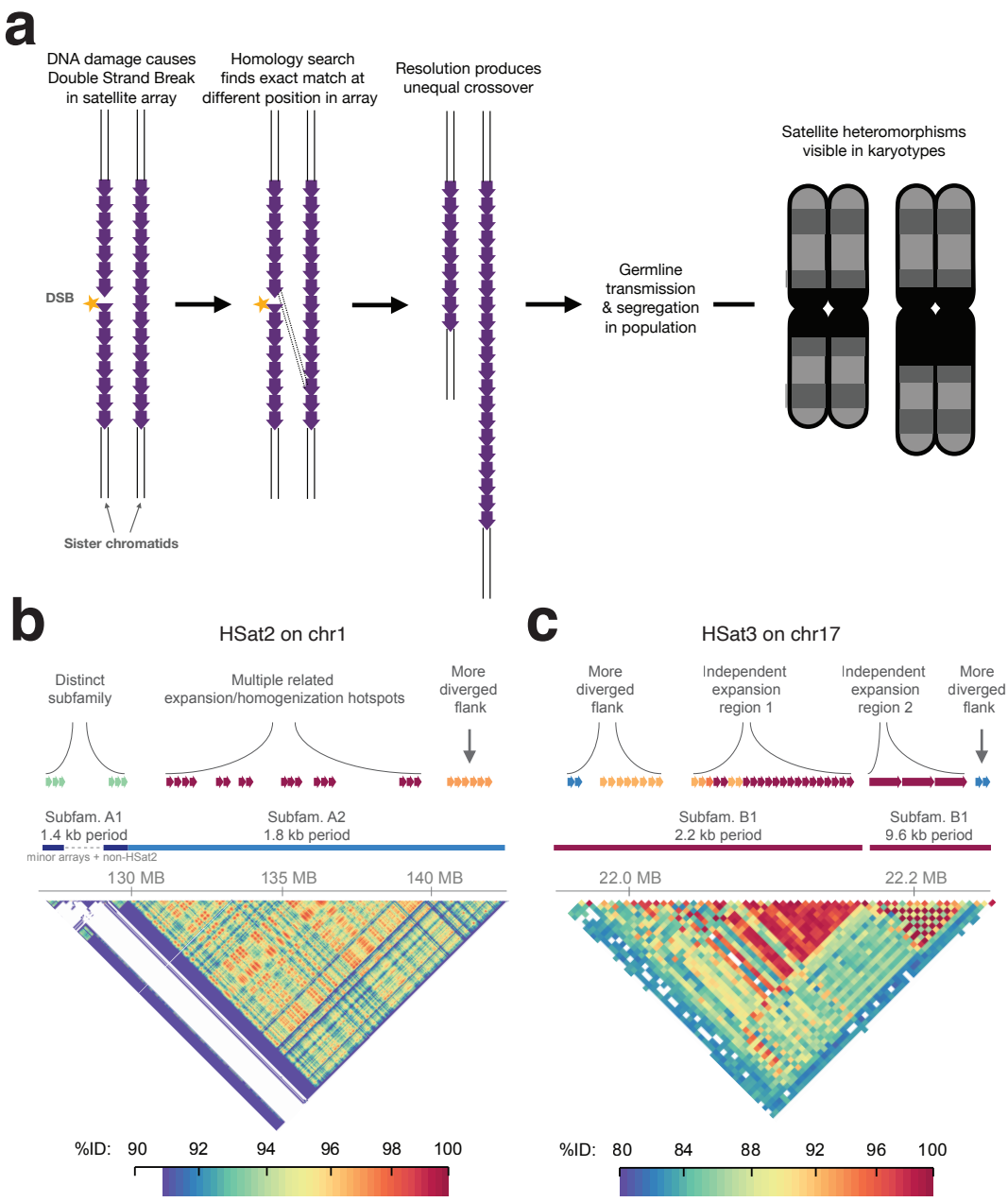

Figure 3. Mechanisms driving the evolution of HSat1-3. a) A schematic illustrating the expansion and contraction of tandem repeat arrays by an unequal non-meiotic crossover mechanism. **b**) An illustration of the repeat structure within the large HSat2 arrays near the centromere of chr1. Below is a StainedGlass plot [113] illustrating the percent identity of pairwise sequence alignments across the array, for which each pixel represents 40 kb. Tandemly arranged arrows illustrate the repeat unit length (arrow length) and homogeneity (arrow color). Generally, the peripheral sequences are more diverged, and labels highlight multiple internal regions of recent expansion and/or homogenization of repeats with the same periodicity. **c**) As in b, but for a ~250 kb HSat3B1 array on chr17 (each StainedGlass pixel represents 5 kb). In this array, different subregions are homogenizing independently and taking on different periodicities.

4.4 Substructure within complete array assemblies reveals local expansion hotspots

To better understand how individual arrays evolve, we examined multiple different HSat2 and HSat3 arrays across the genome using tools like StainedGlass [113] and NTRprism [3] (Figure 3b-c). Most arrays do show higher divergence levels at the fringes, as predicted by theory for unequal crossover mechanisms [100]. Interestingly, one can often observe multiple pockets of high homogeneity within each array, indicative of recent sequence expansions or conversions. In the large HSat2A2 array on chr1, nearly the entire array shares the same 1.8 kb periodicity (Figure 3b). However, in a 900 kb HSat3B3 array on chr20, we observed evidence for recent sequence expansion/homogenization in at least four distinct subregions, each with a different repeat periodicity, suggesting hyper-local and independent evolution of different regions within the array [3]. A similar pattern is observed for the HSat3B1 array on chr17 (Figure 3c). These patterns differ from active αSat arrays, which generally tend to have a single region of recent expansion that is frequently coincident with the centromere [3].

4.5 Inter-species comparisons of HSat1-3 reveal older and newer subfamilies

Comparisons between the genomes of humans and other primates have shed some light on the longer-term evolution of the classical human satellites. Southern blots and in situ hybridization of probes made from satellite fractions I-III found evidence for their presence in the genomes of chimpanzees, orangutans, and gorillas [114–116]. Satellite fraction III also hybridized with New World monkey DNA [116], but given the impurity of early satellite fractions it is difficult to know if this would be true for purer HSat3 repeats. Southern blots identified HSat1B tandem repeats in gorilla DNA but not in chimp or gibbons [48], consistent with a loss of most chrY q-arm heterochromatin in chimpanzees [117]. One study found that probes that are specific to the large HSat3B5 array on chr9 hybridize to multiple pericentromeric regions in gorilla metaphase spreads [118]. Another study showed using PCR and in situ hybridization that individual clones from HSat3A4 arrays on the human acrocentric chromosomes are present in chimps, gorillas, bonobos, orangutangs, and gibbons, which indicates that they appeared in a common ancestor 16-23 million years ago [119]. Some of these HSat3A4 clones were shared only between humans and chimpanzees, consistent with them having appeared less than 5 million years ago [119]. Other clones (corresponding to subfamilies HSat3 subfamilies A1, A2, and 3B2) appeared to be specific to humans [119]. Finally, a comparison of sequencing reads from multiple technologies found evidence for tandem CATTC repeats, which are characteristic of HSat2 and HSat3, in chimpanzees, bonobos, gorillas, and orangutans, with very different abundances in each species [120]. The prospect of telomere-to-telomere assemblies across many primate lineages will open up opportunities to more comprehensively study the evolution of the classical human satellites, resolving the incomplete picture left by prior methods.

5. Epigenomic and functional studies of HSat1-3

5.1 Pericentromeric satellites are hypothesized to play structural roles

Although some researchers hypothesized that human satellites 2 and 3 may play a role in centromere function [121], the fact that these sequences are not present on every chromosome disfavored them as the likely centromeric sequences [24], and later analyses more definitively established αSat as the centromeric satellite [122]. Researchers initially hypothesized that pericentromeric satellite DNAs might play a structural role in the nucleus by providing a platform for the formation of constitutive heterochromatin, which can locally alter gene expression, meiotic

recombination, and nuclear architecture [123]. This hypothesis is still favored, bolstered by recent work in model organisms proposing that pericentromeric satellites function to recruit satellite-binding proteins that establish inter-chromosomal links, which organize chromocenters in the nucleus and prevent chromosomes from drifting away [124,125]. The classical human satellites may play a similar role in the human genome.

5.2 DNA methylation influences satellite DNA organization and regulation

Blocks of constitutive heterochromatin formed by the classical human satellites were shown to contain high levels of 5-methylcytosine by electron microscopy [126] (Figure 4). This was recently confirmed in a human lymphocyte cell line using long-read sequencing, which also found that methylation patterns were periodic in some of the HSat arrays, generally following the satellite sequence periodicity [15]. In contrast, the CHM13 cell line, which resembles early embryonic cells, shows greatly reduced, though still periodic, methylation in these regions [15]. Adding drugs that inhibit DNA methylation results in decondensation of pericentromeric heterochromatin in human cells [127], along with chromosomal rearrangements [128,129] and chromosomal segregation errors that result in the formation of micronuclei [130] (Figure 4). Senescent cells and other cells with natural hypomethylation similarly show satellite DNA decondensation [130–132] along with segregation errors [130,133] and chromosomal rearrangements [134], suggesting a role for satellite DNA misregulation in aging (Figure 4). The new T2T-CHM13 assembly, along with long-read methods for interrogating DNA methylation in repetitive regions, will help to shed even more light on these phenomena [15].

5.3 Satellite DNA transcription can occur in cancerous, senescent, and stressed cells

Other clues about the potential functions or biological effects of the classical human satellites stem from studies of the circumstances under which they are transcribed (Figure 4). HSat2 and/or HSat3 transcripts have been detected in early embryonic cells [135], senescent cells [131], cancer cells [108,131,136–140], cells with DNA damage [141], virus-infected cells [141,142], and stressed cells [143–146]. For example, the large HSat3B5 array on chr9 has been shown to play a role in heat shock responses in human cells in tissue culture [147]. When cells are heat shocked, a protein called heat shock transcription factor 1 (HSF1) localizes predominantly to the chr9q12 heterochromatin domain, where nuclear stress bodies form, in a manner dependent on its DNA-binding and trimerization domains [147]. These stress bodies recruit polymerase II and act like transcription factories, yielding long, non-coding RNAs (lncRNAs) complementary to the G-rich strand of HSat3B5, which stay localized near the satellite DNA [143,144]. Similar behavior was subsequently observed with other stress stimuli, including exposure to heavy metals, UV-C radiation, oxidative stress, and hyperosmotic stress [145,146].

In cancers, nuclear stress bodies also accumulate at other HSat2 and/or HSat3 loci [136,148], which are also sometimes demethylated and bound by polycomb bodies [136,149]. HSat2/3 lncRNAs play a role in recruiting splicing factors and other proteins such as MeCP2 to nuclear stress bodies [136,150–152], and they are associated with mitotic segregation defects [153]. A recent study examined these phenomena further by stably expressing ectopic HSat2 lncRNAs from a transgene randomly integrated into a primary human fibroblast line [154]. They found that lncRNAs from HSat2, but not from αSat, accumulated in visible foci and recruited MeCP2 in cis with the transgene. Furthermore, they found that ectopic satellite expression resulted in segregation defects [154]. These results suggest that the formation of nuclear stress bodies and chromosome

instability phenotypes in cancer cells result directly from aberrant satellite transcription rather than some other property of cancer. Moving forward, the T2T-CHM13 assembly and long-read lncRNA sequencing methods will help to interrogate the exact origins of satellite transcripts at much finer resolution [155].

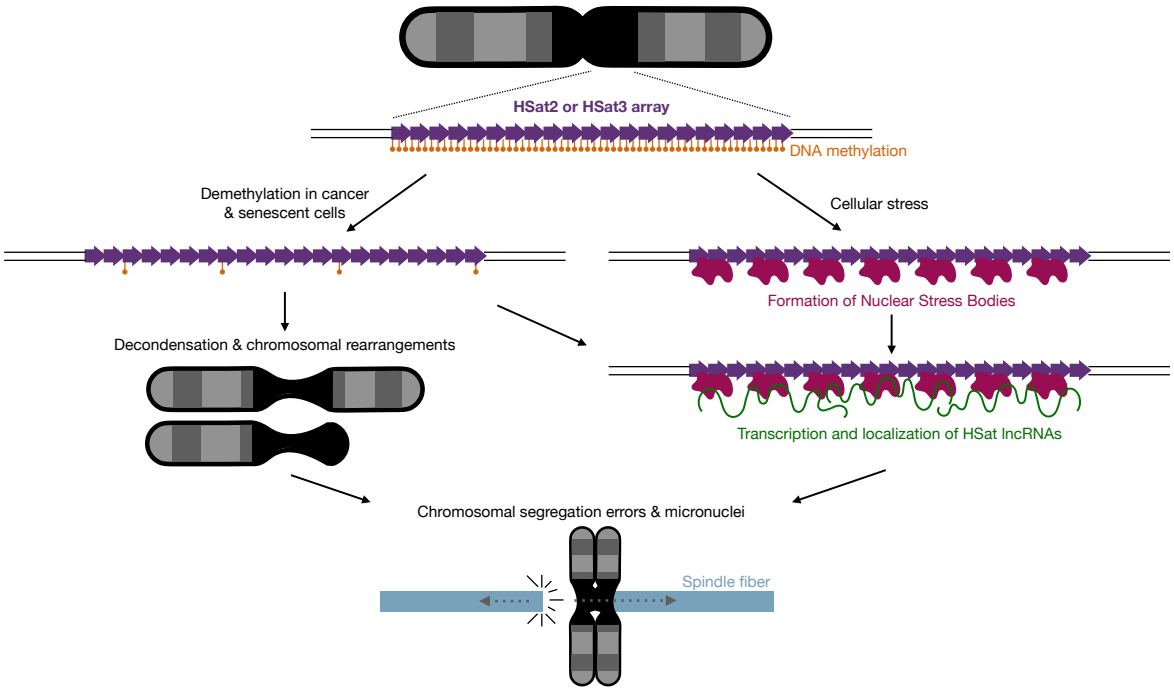

Figure 4. Roles for HSat2,3 arrays in stress responses, cancer, and senescence. At the top is a schematic of a typical HSat2 or HSat3 array in a healthy cell. The DNA is heavily methylated and the region becomes highly condensed in mitosis. In senescent cells and certain cancer cells, genome misregulation can cause the array to become demethylated, resulting in decondensation, chromosomal rearrangements, satellite transcription, and chromosomal segregation errors. When cells are exposed to stress stimuli, a subset of HSat3 arrays become hubs for nuclear stress bodies, which recruit polymerases that transcribe the satellite DNA into lncRNAs that remain localized in cis. Presence of these transcripts can also lead to chromosomal segregation errors.

5.4 Satellite DNAs may play a role in speciation

The high mutation rate of satellite DNA was also proposed as a driver of speciation by reproductive isolation [123], which can result from genetic conflict or Dobzhansky-Muller incompatibilities between rapidly evolving elements. Although this hypothesis is difficult to test in mammals, careful experiments in *Drosophila* species have shed some light on this phenomenon. Satellite DNA content can differ dramatically between reproductively incompatible *Drosophila* species [156,157], but it can also differ dramatically between reproductively compatible *Drosophila* strains [156]. A specific non-centromeric satellite repeat was shown to cause hybrid incompatibility in crosses between *melanogaster* and *simulans* [158,159], and later it was shown that satellite DNAs fail to cluster properly into chromocenters in incompatible hybrids [160]. Whether satellite DNAs play a broader role in speciation remains to be seen.

6. Conclusion

Fifty-five years have passed since the first human satellite fraction was described in 1967. Although our understanding of the classical human satellites has lagged behind the rest of the genome, the ability to fully assemble across these formerly intractable repetitive arrays will enable careful studies of their regulation and function like never before. Technological advances in DNA editing will allow researchers to delete entire HSat1-3 arrays and study the phenotypic consequences of these knockouts in human cell lines and organoids. The long-read DNA sequencing technologies and assembly methods used to create the first human T2T assembly will also enable the assembly of analogous pericentromeric satellite arrays in model organisms, in which perturbation experiments may reveal broader principles that govern the behavior and evolution of pericentromeric heterochromatin. Comparing T2T assemblies across primate and other mammalian lineages will enable the study of the deeper evolutionary origins of HSat1-3, and comparisons across human T2T assemblies will enhance our understanding the variability and recent evolution of these regions. Future studies of HSat1-3 will also benefit from new long-read technologies for mapping protein-DNA interactions, DNA methylation, and DNA accessibility in repetitive regions [13–16]. These future studies will also benefit from new computational methods for comparing complex satellite array sequences and modeling their evolutionary histories. In the coming years, our understanding of the classical human satellites may finally catch up to the rest of the genome as researchers are newly equipped to investigate their roles in human evolution, health, and disease.

Footnotes

1) Saul Kit's choice in 1961 to describe the minor DNA band as a "satellite" may have been inspired by the news at the time [112], since the Space Age had just begun and the first crop of artificial satellites were in orbit. Though apocryphal, this theory on the origination of the name was believed to be true by several of the leading figures in early satellite DNA research (Chris Tyler-Smith, personal correspondence).

2) Cesium density gradient fractions, although later shown to be enriched for certain repeat sequence families, were known to be impure composites of sequences that could differ depending on ultracentrifugation conditions [42,44,161]. A satellite fraction IV was also isolated using a column fractionation method [162], but later studies found that its components were indistinguishable from those found in satellite fraction III [29,42,44,161]. Repeat classes labeled as "Human Satellites 4-6" in RepBase bear no relation to the classical human satellite fractions and are not discussed here.

3) Typically, automated annotation of HSat1-3 in human assemblies relies on RepeatMasker. RepeatMasker annotates HSat1A sequences as "SAR" and HSat1B sequences as "HSATI". Unfortunately, the current version of RepeatMasker, which depends on a Repbase library for comparison, does a poor job of distinguishing HSat2 and HSat3, which both tend to be annotated as either an "HSATII" satellite repeat or a "(CATTC)n" simple repeat. Until this issue is resolved by RepeatMasker/Repbase, an alternative automated method for annotating HSat2 and HSat3 has been provided at https://github.com/altemose/chm13_hsat [3].

4) NTRprism is akin to the classical analytical restriction digest experiments used to characterize satellite arrays in the past, and it is similar in approach to previous computational tools for detecting repeat periodicity [163]. This algorithm takes advantage of the complete information in a full array assembly by essentially simulating digestion with restriction enzymes that could cut all possible recognition sites. Then, it combines information across these simulations to reveal the predominant periodicities within the array. Running NTRprism in windows across an array can reveal variation in periodicity within that array.

Acknowledgements

I am grateful to Karen Miga, Hunt Willard, Gary Karpen, Lucy Brennan, Varsha Rajshekar, Sasha Langley, Chuck Langley, Serafin Colmenares, Collin Hickmann, and Chris Tyler-Smith for their helpful discussions and feedback.

Funding: NA is supported by a Howard Hughes Medical Institute Hanna H. Gray Fellowship.

References

[1] K.H. Miga, Centromeric Satellite DNAs: Hidden Sequence Variation in the Human Population, Genes-Basel. 10 (2019) 352. https://doi.org/10.3390/genes10050352.

[2] C.W. Schmid, P.L. Deininger, Sequence organization of the human genome, Cell. 6 (1975) 345–358. https://doi.org/10.1016/0092-8674(75)90184-1.

[3] N. Altemose, G.A. Logsdon, A.V. Bzikadze, P. Sidhwani, S.A. Langley, G.V. Caldas, S.J. Hoyt, L. Uralsky, F.D. Ryabov, C.J. Shew, M.E.G. Sauria, M. Borchers, A. Gershman, A. Mikheenko, V.A. Shepelev, T. Dvorkina, O. Kunyavskaya, M.R. Vollger, A. Rhie, A.M. McCartney, M. Asri, R. Lorig-Roach, K. Shafin, J.K. Lucas, S. Aganezov, D. Olson, L.G. de Lima, T. Potapova, G.A. Hartley, M. Haukness, P. Kerpedjiev, F. Gusev, K. Tigyi, S. Brooks, A. Young, S. Nurk, S. Koren, S.R. Salama, B. Paten, E.I. Rogaev, A. Streets, G.H. Karpen, A.F. Dernburg, B.A. Sullivan, A.F. Straight, T.J. Wheeler, J.L. Gerton, E.E. Eichler, A.M. Phillippy, W. Timp, M.Y. Dennis, R.J. O'Neill, J.M. Zook, M.C. Schatz, P.A. Pevzner, M. Diekhans, C.H. Langley, I.A. Alexandrov, K.H. Miga, Complete genomic and epigenetic maps of human centromeres., Sci New York N Y. 376 (2022) eabl4178. https://doi.org/10.1126/science.abl4178.

[4] W.G. Yasmineh, J.J. Yunis, Satellite DNA in mouse autosomal heterochromatin., Biochem Bioph Res Co. 35 (1969) 779–82. https://doi.org/10.1016/0006-291x(69)90690-1.

[5] E. Heitz, Das Heterochromatin der Moose, Jahrb. Wiss. Bot. (1928) 762–818.

[6] M.G. Schueler, A.W. Higgins, M.K. Rudd, K. Gustashaw, H.F. Willard, Genomic and Genetic Definition of a Functional Human Centromere, Science. 294 (2001) 109–115. https://doi.org/10.1126/science.1065042.

[7] L. Manuelidis, Repeating restriction fragments of human DNA, Nucleic Acids Res. 3 (1976) 3063–3076. https://doi.org/10.1093/nar/3.11.3063.

[8] J. Prosser, M. Frommer, C. Paul, P.C. Vincent, Sequence relationships of three human satellite DNAs, J Mol Biol. 187 (1986) 145–155. https://doi.org/10.1016/0022-2836(86)90224-x.

[9] G. Corneo, E. Ginelli, E. Polli, A satellite DNA isolated from human tissues, J Mol Biol. 23 (1967) 619–622. https://doi.org/10.1016/s0022-2836(67)80130-x.

[10] G. Corneo, E. Ginelli, E. Polli, Renaturation properties and localization in heterochromatin of human satellite DNA's, Biochimica Et Biophysica Acta Bba - Nucleic Acids Protein Synthesis. 247 (1971) 528–534. https://doi.org/10.1016/0005-2787(71)90689-7.

[11] 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 (2004) 345–354. https://doi.org/10.1038/nrg1322.

[12] S. Nurk, S. Koren, A. Rhie, M. Rautiainen, A.V. Bzikadze, A. Mikheenko, M.R. Vollger, N. Altemose, L. Uralsky, A. Gershman, S. Aganezov, S.J. Hoyt, M. Diekhans, G.A. Logsdon, M. Alonge, S.E. Antonarakis, M. Borchers, G.G. Bouffard, S.Y. Brooks, G.V. Caldas, N.-C. Chen, H. Cheng, C.-S. Chin, W. Chow, L.G. de Lima, P.C. Dishuck, R. Durbin, T. Dvorkina, I.T. Fiddes, G. Formenti, R.S. Fulton, A. Fungtammasan, E. Garrison, P.G.S. Grady, T.A. Graves-Lindsay, I.M. Hall, N.F. Hansen, G.A. Hartley, M. Haukness, K. Howe, M.W. Hunkapiller, C. Jain, M. Jain, E.D. Jarvis, P. Kerpedjiev, M. Kirsche, M. Kolmogorov, J. Korlach, M. Kremitzki, H. Li, V.V. Maduro, T. Marschall, A.M. McCartney, J. McDaniel, D.E. Miller, J.C. Mullikin, E.W. Myers, N.D. Olson, B. Paten, P. Peluso, P.A. Pevzner, D. Porubsky, T. Potapova, E.I. Rogaev, J.A. Rosenfeld, S.L. Salzberg, V.A. Schneider, F.J. Sedlazeck, K. Shafin, C.J. Shew, A. Shumate, Y. Sims, A.F.A. Smit, D.C. Soto, I. Sović, J.M. Storer, A. Streets, B.A. Sullivan, F. Thibaud-Nissen, J. Torrance, J. Wagner, B.P. Walenz, A. Wenger, J.M.D. Wood, C. Xiao, S.M. Yan, A.C. Young, S. Zarate, U. Surti, R.C. McCoy, M.Y. Dennis, I.A. Alexandrov, J.L. Gerton, R.J. O'Neill, W. Timp, J.M. Zook, M.C. Schatz, E.E. Eichler, K.H. Miga, A.M. Phillippy, The complete sequence of a human genome., Sci New York N Y. 376 (2022) 44–53. https://doi.org/10.1126/science.abj6987.

[13] N. Altemose, A. Maslan, O.K. Smith, K. Sundararajan, R.R. Brown, R. Meeshra, A.M. Detweiler, N. Neff, K.H. Miga, A.F. Straight, A. Streets, DiMeLo-seq: a long-read, single-molecule method for mapping protein-DNA interactions genome-wide, Nat. Meth. (2022). https://doi.org/10.1038/s41592-022-01475-6.

[14] A.B. Stergachis, B.M. Debo, E. Haugen, L.S. Churchman, J.A. Stamatoyannopoulos, Single-molecule regulatory architectures captured by chromatin fiber sequencing, Science. 368 (2020) 1449–1454. https://doi.org/10.1126/science.aaz1646.

[15] A. Gershman, M.E.G. Sauria, X. Guitart, M.R. Vollger, P.W. Hook, S.J. Hoyt, M. Jain, A. Shumate, R. Razaghi, S. Koren, N. Altemose, G.V. Caldas, G.A. Logsdon, A. Rhie, E.E. Eichler, M.C. Schatz, R.J. O'Neill, A.M. Phillippy, K.H. Miga, W. Timp, Epigenetic patterns in a complete human genome, Science. 376 (2022) eabj5089. https://doi.org/10.1126/science.abj5089.

[16] I. Lee, R. Razaghi, T. Gilpatrick, M. Molnar, A. Gershman, N. Sadowski, F.J. Sedlazeck, K.D. Hansen, J.T. Simpson, W. Timp, Simultaneous profiling of chromatin accessibility and methylation on human cell lines with nanopore sequencing, Nat Methods. 17 (2020) 1191–1199. https://doi.org/10.1038/s41592-020-01000-7.

[17] M. Meselson, F.W. Stahl, J. Vinograd, EQUILIBRIUM SEDIMENTATION OF MACROMOLECULES IN DENSITY GRADIENTS, Proc National Acad Sci. 43 (1957) 581–588. https://doi.org/10.1073/pnas.43.7.581.

[18] S. Kit, Equilibrium sedimentation in density gradients of DNA preparations from animal tissues, Journal of Molecular Biology. 3 (1961) 711-IN2. https://doi.org/10.1016/s0022-2836(61)80075-2.

[19] N. Sueoka, Variation and heterogeneity of base composition of deoxyribonucleic acids: A compilation of old and new data, J Mol Biol. 3 (1961) 31-IN15. https://doi.org/10.1016/s0022-2836(61)80005-3.

[20] G. Corneo, E. Ginelli, E. Polli, Repeated sequences in human DNA, J Mol Biol. 48 (1970) 319–327. https://doi.org/10.1016/0022-2836(70)90163-4.

[21] M. Waring, R.J. Britten, Nucleotide Sequence Repetition: A Rapidly Reassociating Fraction of Mouse DNA, Science. 154 (1966) 791–794. https://doi.org/10.1126/science.154.3750.791.

[22] R.J. Britten, D.E. Kohne, Repeated Sequences in DNA, Science. 161 (1968) 529–540. https://doi.org/10.1126/science.161.3841.529.

[23] L.M. Kunkel, K.D. Smith, S.H. Boyer, Human Y-chromosome-specific reiterated DNA, Science (New York, NY). 191 (1976) 1189–1190. https://doi.org/10.1126/science.1257744.

[24] K.W. Jones, G. Corneo, Location of Satellite and Homogeneous DNA Sequences on Human Chromosomes, Nat New Biology. 233 (1971) 268–271. https://doi.org/10.1038/newbio233268a0.

[25] K.W. Jones, Chromosomal and Nuclear Location of Mouse Satellite DNA in Individual Cells, Nature. 225 (1970) 912–915. https://doi.org/10.1038/225912a0.

[26] M.L. Pardue, J.G. Gall, Chromosomal Localization of Mouse Satellite DNA, Science. 168 (1970) 1356–1358. https://doi.org/10.1126/science.168.3937.1356.

[27] F.E. Arrighi, P.P. Saunders, G.F. Saunders, T.C. Hsu, Distribution of repetitious DNA in human chromosomes, Experientia. 27 (1971) 964–966. https://doi.org/10.1007/bf02135776.

[28] K.W. Jones, I.F. Purdom, J. Prosser, G. Corneo, The chromosomal localisation of human satellite DNA I, Chromosoma. 49 (1974) 161–171. https://doi.org/10.1007/bf00348888.

[29] J.R. Gosden, A.R. Mitchell, R.A. Buckland, R.P. Clayton, H.J. Evans, The location of four human satellite DNAs on human chromosomes, Exp Cell Res. 92 (1975) 148–158. https://doi.org/10.1016/0014-4827(75)90648-5.

[30] I. Tagarro, J. Wiegant, A.K. Raap, J.J. González-Aguilera, A.M. Fernández-Peralta, Assignment of human satellite 1 DNA as revealed by fluorescent in situ hybridization with oligonucleotides, Hum Genet. 93 (1994) 125–128. https://doi.org/10.1007/bf00210595.

[31] I. Tagarro, A.M. Fernández-Peralta, J.J. González-Aguilera, Chromosomal localization of human satellites 2 and 3 by a FISH method using oligonucleotides as probes, Hum Genet. 93 (1994) 383–388. https://doi.org/10.1007/bf00201662.

[32] G.F. Saunders, T.C. Hsu, M.J. Getz, E.L. Simes, F.E. Arrighi, Locations of a Human Satellite DNA in Human Chromosomes, Nat New Biology. 236 (1972) 244–246. https://doi.org/10.1038/newbio236244a0.

[33] K.W. Jones, J. Prosser, G. Corneo, E. Ginelli, The chromosomal location of human satellite DNA III, Chromosoma. 42 (1973) 445–451. https://doi.org/10.1007/bf00399411.

[34] E.M. Southern, Base Sequence and Evolution of Guinea-pig α-Satellite DNA, Nature. 227 (1970) 794–798. https://doi.org/10.1038/227794a0.

[35] E.M. Southern, Long range periodicities in mouse satellite DNA, J Mol Biol. 94 (1975) 51–69. https://doi.org/10.1016/0022-2836(75)90404-0.

[36] E.M. Southern, Detection of specific sequences among DNA fragments separated by gel electrophoresis, J Mol Biol. 98 (1975) 503–517. https://doi.org/10.1016/s0022-2836(75)80083-0.

[37] C.J. Shen, G. Wiesehahn, J.E. Hearst, Cleavage patterns of Drosophila melanogaster satellite DNA by restriction enzymes, Nucleic Acids Res. 3 (1976) 931–952. https://doi.org/10.1093/nar/3.4.931.

[38] H.J. Cooke, Repeated sequence specific to human males, Nature. 262 (1976) 182–186. https://doi.org/10.1038/262182a0.

[39] P. Philippsen, R.E. Streeck, H.G. Zachau, Defined Fragments of Calf, Human, and Rat DNA Produced by Restriction Nucleases, Eur J Biochem. 45 (1974) 479–488. https://doi.org/10.1111/j.1432-1033.1974.tb03573.x.

[40] J.S. Waye, S.J. Durfy, D. Pinkel, S. Kenwrick, M. Patterson, K.E. Davies, H.F. Willard, Chromosome-specific alpha satellite DNA from human chromosome 1: hierarchical structure and genomic organization of a polymorphic domain spanning several hundred kilobase pairs of centromeric DNA., Genomics. 1 (1987) 43–51. https://doi.org/10.1016/0888-7543(87)90103-0.

[41] C. Tyler-Smith, W.R.A. Brown, Structure of the major block of alphoid satellite DNA on the human Y chromosome, J Mol Biol. 195 (1987) 457–470. https://doi.org/10.1016/0022-2836(87)90175-6.

[42] A.R. Mitchell, R.S. Beauchamp, C.J. Bostock, A study of sequence homologies in four satellite DNAs of man, J Mol Biol. 135 (1979) 127–149. https://doi.org/10.1016/0022-2836(79)90344-9.

[43] M. Frommer, J. Prosser, P.C. Vincent, Human satellite I sequences include a male specific 2.47 kb tandemly repeated unit containing one Alu family member per repeat, Nucleic Acids Res. 12 (1984) 2887–2900. https://doi.org/10.1093/nar/12.6.2887.

[44] M. Frommer, J. Prosser, D. Tkachuk, A.H. Reisner, P.C. Vincent, Simple repeated sequences in human satellite DNA, Nucleic Acids Res. 10 (1982) 547–563. https://doi.org/10.1093/nar/10.2.547.

[45] H.J. Cooke, J. Hindley, Cloning of human satellite III DNA: different components are on different chromosomes, Nucleic Acids Res. 6 (1979) 3177–3198. https://doi.org/10.1093/nar/6.10.3177.

[46] R.S. Beauchamp, A.R. Mitchell, R.A. Buckland, C.J. Bostock, Specific arrangements of human satellite III DNA sequences in human chromosomes, Chromosoma. 71 (1979) 153–166. https://doi.org/10.1007/bf00292820.

[47] M.J. Higgins, H. Wang, I. Shtromas, T. Haliotis, J.C. Roder, J.J.A. Holden, B.N. White, Organization of a repetitive human 1.8 kb KpnI sequence localized in the heterochromatin of chromosome 15, Chromosoma. 93 (1985) 77–86. https://doi.org/10.1007/bf01259449.

[48] H.J. Cooke, J. Schmidtke, J.R. Gosden, Characterisation of a human Y chromosome repeated sequence and related sequences in higher primates, Chromosoma. 87 (1982) 491–502. https://doi.org/10.1007/bf00333470.

[49] H.J. Cooke, R.D.G. McKay, Evolution of a human Y chromosome-specific repeated sequence, Cell. 13 (1978) 453–460. https://doi.org/10.1016/0092-8674(78)90319-7.

[50] M. Pech, T. Igo-Kemenes, H.G. Zachau, Nucleotide sequence of a highly repetitive component of rat DNA, Nucleic Acids Res. 7 (1979) 417–432. https://doi.org/10.1093/nar/7.2.417.

[51] M.F. Singer, Highly Repeated Sequences in Mammalian Genomes, Int Rev Cytol. 76 (1982) 67–112. https://doi.org/10.1016/s0074-7696(08)61789-1.

[52] F. Sanger, S. Nicklen, A.R. Coulson, DNA sequencing with chain-terminating inhibitors., P Natl Acad Sci Usa. 74 (1977) 5463–7. https://doi.org/10.1073/pnas.74.12.5463.

[53] A.M. Maxam, W. Gilbert, Sequencing end-labeled DNA with base-specific chemical cleavages., Methods Enzymol. 65 (1980) 499–560. https://doi.org/10.1016/s0076-6879(80)65059-9.

[54] J. Meyne, E.H. Goodwin, R.K. Moyzis, Chromosome localization and orientation of the simple sequence repeat of human satellite I DNA, Chromosoma. 103 (1994) 99–103. https://doi.org/10.1007/bf00352318.

[55] R.D.G. McKay, M. Bobrow, H.J. Cooke, The identification of a repeated DNA sequence involved in the karyotype polymorphism of the human Y chromosome, Cytogenet Genome Res. 21 (1978) 19–32. https://doi.org/10.1159/000130875.

[56] E. Manz, M. Alkan, E. Bühler, J. Schmidtke, Arrangement of DYZ1 and DYZ2 repeats on the human Y-chromosome: a case with presence of DYZ1 and absence of DYZ2, Mol Cell Probe. 6 (1992) 257–259. https://doi.org/10.1016/0890-8508(92)90025-s.

[57] M. Schmid, M. Guttenbach, I. Nanda, R. Studer, J.T. Epplen, Organization of DYZ2 repetitive DNA on the human Y chromosome, Genomics. 6 (1990) 212–218. https://doi.org/10.1016/0888-7543(90)90559-d.

[58] Y. Nakahori, K. Mitani, M. Yamada, Y. Nakagome, A human Y-chromosome specific repeated DNA family (DYZ1) consists of a tandem array of pentanucleotides, Nucleic Acids Res. 14 (1986) 7569–7580. https://doi.org/10.1093/nar/14.19.7569.

[59] N. Mathias, M. Bayes, C. Tyler-Smith, Highly informative compound haplotypes for the human Y chromosome, Hum Mol Genet. 3 (1994) 115–123. https://doi.org/10.1093/hmg/3.1.115.

[60] M.S. Jackson, S.E. Mole, B.A. Ponder, Characterisation of a boundary between satellite III and alphoid sequences on human chromosome 10., Nucleic Acids Res. 20 (1992) 4781–7.

[61] M.S. Jackson, P. Slijepcevic, B.A.J. Ponder, The organisation of repetitive sequences in the pericentromeric region of human chromosome 10, Nucleic Acids Res. 21 (1993) 5865–5874. https://doi.org/10.1093/nar/21.25.5865.

[62] M. Jeanpierre, D. Weil, P. Gallano, N. Creau-Goldberg, C. Junien, The organization of two related subfamilies of a human tandemly repeated DNA is chromosome specific, Hum Genet. 70 (1985) 302–310. https://doi.org/10.1007/bf00295365.

[63] M. Jeanpierre, Human satellites 2 and 3., Ann De Génétique. 37 (1994) 163–71.

[64] R.K. Moyzis, K.L. Albright, M.F. Bartholdi, L.S. Cram, L.L. Deaven, C.E. Hildebrand, N.E. Joste, J.L. Longmire, J. Meyne, T. Schwarzacher-Robinson, Human chromosome-specific repetitive DNA sequences: novel markers for genetic analysis, Chromosoma. 95 (1987) 375–386. https://doi.org/10.1007/bf00333988.

[65] P.L. Deininger, D.J. Jolly, C.M. Rubin, T. Friedmann, C.W. Schmid, Base sequence studies of 300 nucleotide renatured repeated human DNA clones, J Mol Biol. 151 (1981) 17–33. https://doi.org/10.1016/0022-2836(81)90219-9.

[66] R. Bandyopadhyay, C. McQuillan, S.L. Page, K.H.A. Choo, L.G. Shaffer, Identification and characterization of satellite III subfamilies to the acrocentric chromosomes, Chromosome Res. 9 (2001) 223–233. https://doi.org/10.1023/a:1016648404388.

[67] K.H. Choo, E. Earle, C. McQuillan, A homologous subfamily of satellite III DNA on human chromosomes 14 and 22, Nucleic Acids Res. 18 (1990) 5641–5648. https://doi.org/10.1093/nar/18.19.5641.

[68] K.H. Choo, E. Earle, B. Vissel, P. Kalitsis, A chromosome 14-specific human satellite III DNA subfamily that shows variable presence on different chromosomes 14., Am J Hum Genet. 50 (1992) 706–16.

[69] B. Vissel, A. Nagy, K.H. Choo, A satellite III sequence shared by human chromosomes 13, 14, and 21 that is contiguous with alpha satellite DNA., Cytogenet Cell Genet. 61 (1992) 81–6.

[70] P. Kalitsis, E. Earle, B. Vissel, L.G. Shaffer, K.H.A. Choo, A Chromosome 13-Specific Human Satellite I DNA Subfamily with Minor Presence on Chromosome 21: Further Studies on Robertsonian Translocations, Genomics. 16 (1993) 104–112. https://doi.org/10.1006/geno.1993.1147.

[71] A. Legrand, D.B. Zimonjic, N.C. Popescu, M. Nachtigal, A Chromosome 4 Satellite I DNA Isolated from SV40-Transformed Human Cells, Dna Res. 5 (1998) 271–276. https://doi.org/10.1093/dnares/5.5.271.

[72] G. Sauter, H. Moch, P. Carroll, R. Kerschmann, M.J. Mihatsch, F.M. Waldman, Chromosome-9 loss detected by fluorescencein situ hybridization in bladder cancer, Int J Cancer. 64 (1995) 99–103. https://doi.org/10.1002/ijc.2910640205.

[73] I.H.G.S. Consortium, Initial sequencing and analysis of the human genome, Nature. 409 (2001) 860–921. https://doi.org/10.1038/35057062.

[74] J.C. Venter, M.D. Adams, E.W. Myers, P.W. Li, R.J. Mural, G.G. Sutton, H.O. Smith, M. Yandell, C.A. Evans, R.A. Holt, J.D. Gocayne, P. Amanatides, R.M. Ballew, D.H. Huson, J.R. Wortman, Q. Zhang, C.D. Kodira, X.H. Zheng, L. Chen, M. Skupski, G. Subramanian, P.D. Thomas, J. Zhang, G.L.G. Miklos, C. Nelson, S. Broder, A.G. Clark, J. Nadeau, V.A. McKusick, N. Zinder, A.J. Levine, R.J. Roberts, M. Simon, C. Slayman, M. Hunkapiller, R. Bolanos, A. Delcher, I. Dew, D. Fasulo, M. Flanigan, L. Florea, A. Halpern, S. Hannenhalli, S. Kravitz, S. Levy, C. Mobarry, K. Reinert, K. Remington, J. Abu-Threideh, E. Beasley, K. Biddick, V. Bonazzi, R. Brandon, M. Cargill, I. Chandramouliswaran, R. Charlab, K. Chaturvedi, Z. Deng, V.D. Francesco, P. Dunn, K. Eilbeck, C. Evangelista, A.E. Gabrielian, W. Gan, W. Ge, F. Gong, Z. Gu, P. Guan, T.J. Heiman, M.E. Higgins, R.R. Ji, Z. Ke, K.A. Ketchum, Z. Lai, Y. Lei, Z. Li, J. Li, Y. Liang, X. Lin, F. Lu, G.V. Merkulov, N. Milshina, H.M. Moore, A.K. Naik, V.A. Narayan, B. Neelam, D. Nusskern, D.B. Rusch, S. Salzberg, W. Shao, B. Shue, J. Sun, Z. Wang, A. Wang, X. Wang, J. Wang, M. Wei, R. Wides, C. Xiao, C. Yan, A. Yao, J. Ye, M. Zhan, W. Zhang, H. Zhang, Q. Zhao, L. Zheng, F. Zhong, W. Zhong, S. Zhu, S. Zhao, D. Gilbert, S. Baumhueter, G. Spier, C. Carter, A. Cravchik, T. Woodage, F. Ali, H. An, A. Awe, D. Baldwin, H. Baden, M. Barnstead, I. Barrow, K. Beeson, D. Busam, A. Carver, A. Center, M.L. Cheng, L. Curry, S. Danaher, L. Davenport, R. Desilets, S. Dietz, K. Dodson, L. Doup, S. Ferriera, N. Garg, A. Gluecksmann, B. Hart, J. Haynes, C. Haynes, C. Heiner, S. Hladun, D. Hostin, J. Houck, T. Howland, C. Ibegwam, J. Johnson, F. Kalush, L. Kline, S. Koduru, A. Love, F. Mann, D. May, S. McCawley, T. McIntosh, I. McMullen, M. Moy, L. Moy, B. Murphy, K. Nelson, C. Pfannkoch, E. Pratts, V. Puri, H. Qureshi, M. Reardon, R. Rodriguez, Y.H. Rogers, D. Romblad, B. Ruhfel, R. Scott, C. Sitter, M. Smallwood, E. Stewart, R. Strong, E. Suh, R. Thomas, N.N. Tint, S. Tse, C. Vech, G. Wang, J. Wetter, S. Williams, M. Williams, S. Windsor, E. Winn-Deen,

K. Wolfe, J. Zaveri, K. Zaveri, J.F. Abril, R. Guigó, M.J. Campbell, K.V. Sjolander, B. Karlak, A. Kejariwal, H. Mi, B. Lazareva, T. Hatton, A. Narechania, K. Diemer, A. Muruganujan, N. Guo, S. Sato, V. Bafna, S. Istrail, R. Lippert, R. Schwartz, B. Walenz, S. Yooseph, D. Allen, A. Basu, J. Baxendale, L. Blick, M. Caminha, J. Carnes-Stine, P. Caulk, Y.H. Chiang, M. Coyne, C. Dahlke, A. Mays, M. Dombroski, M. Donnelly, D. Ely, S. Esparham, C. Fosler, H. Gire, S. Glanowski, K. Glasser, A. Glodek, M. Gorokhov, K. Graham, B. Gropman, M. Harris, J. Heil, S. Henderson, J. Hoover, D. Jennings, C. Jordan, J. Jordan, J. Kasha, L. Kagan, C. Kraft, A. Levitsky, M. Lewis, X. Liu, J. Lopez, D. Ma, W. Majoros, J. McDaniel, S. Murphy, M. Newman, T. Nguyen, N. Nguyen, M. Nodell, S. Pan, J. Peck, M. Peterson, W. Rowe, R. Sanders, J. Scott, M. Simpson, T. Smith, A. Sprague, T. Stockwell, R. Turner, E. Venter, M. Wang, M. Wen, D. Wu, M. Wu, A. Xia, A. Zandieh, X. Zhu, The sequence of the human genome., Science (New York, NY). 291 (2001) 1304–1351. https://doi.org/10.1126/science.1058040.

[75] N. Altemose, K.H. Miga, M. Maggioni, H.F. Willard, Genomic characterization of large heterochromatic gaps in the human genome assembly, Plos Comput Biol. 10 (2014) e1003628. https://doi.org/10.1371/journal.pcbi.1003628.

[76] P.E. Warburton, D. Hasson, F. Guillem, C. Lescale, X. Jin, G. Abrusan, Analysis of the largest tandemly repeated DNA families in the human genome, Bmc Genomics. 9 (2008) 533. https://doi.org/10.1186/1471-2164-9-533.

[77] S. Levy, G. Sutton, P.C. Ng, L. Feuk, A.L. Halpern, B.P. Walenz, N. Axelrod, J. Huang, E.F. Kirkness, G. Denisov, Y. Lin, J.R. MacDonald, A.W.C. Pang, M. Shago, T.B. Stockwell, A. Tsiamouri, V. Bafna, V. Bansal, S.A. Kravitz, D.A. Busam, K.Y. Beeson, T.C. McIntosh, K.A. Remington, J.F. Abril, J. Gill, J. Borman, Y.-H. Rogers, M.E. Frazier, S.W. Scherer, R.L. Strausberg, J.C. Venter, The Diploid Genome Sequence of an Individual Human, Plos Biol. 5 (2007) e254. https://doi.org/10.1371/journal.pbio.0050254.

[78] N. Kaplan, J. Dekker, High-throughput genome scaffolding from in vivo DNA interaction frequency, Nat Biotechnol. 31 (2013) 1143–1147. https://doi.org/10.1038/nbt.2768.

[79] G. Genovese, R.E. Handsaker, H. Li, N. Altemose, A.M. Lindgren, K. Chambert, B. Pasaniuc, A.L. Price, D. Reich, C.C. Morton, M.R. Pollak, J.G. Wilson, S.A. McCarroll, Using population admixture to help complete maps of the human genome, Nature Publishing Group. 45 (2013) 406–414. https://doi.org/10.1038/ng.2565.

[80] H. Skaletsky, T. Kuroda-Kawaguchi, P.J. Minx, H.S. Cordum, L. Hillier, L.G. Brown, S. Repping, T. Pyntikova, J. Ali, T. Bieri, A. Chinwalla, A. Delehaunty, K. Delehaunty, H. Du, G. Fewell, L. Fulton, R. Fulton, T. Graves, S.-F. Hou, P. Latrielle, S. Leonard, E. Mardis, R. Maupin, J. McPherson, T. Miner, W. Nash, C. Nguyen, P. Ozersky, K. Pepin, S. Rock, T. Rohlfing, K. Scott, B. Schultz, C. Strong, A. Tin-Wollam, S.-P. Yang, R.H. Waterston, R.K. Wilson, S. Rozen, D.C. Page, The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes, Nature. 423 (2003) 825–837. https://doi.org/10.1038/nature01722.

[81] J. Guy, T. Hearn, M. Crosier, J. Mudge, L. Viggiano, D. Koczan, H.-J. Thiesen, J.A. Bailey, J.E. Horvath, E.E. Eichler, M.E. Earthrowl, P. Deloukas, L. French, J. Rogers, D. Bentley, M.S.

Jackson, Genomic sequence and transcriptional profile of the boundary between pericentromeric satellites and genes on human chromosome arm 10p., Genome Res. 13 (2003) 159–72. https://doi.org/10.1101/gr.644503.

[82] X. She, J.E. Horvath, Z. Jiang, G. Liu, T.S. Furey, L. Christ, R. Clark, T. Graves, C.L. Gulden, C. Alkan, J.A. Bailey, C. Sahinalp, M. Rocchi, D. Haussler, R.K. Wilson, W. Miller, S. Schwartz, E.E. Eichler, The structure and evolution of centromeric transition regions within the human genome, Nature. 430 (2004) 857–864. https://doi.org/10.1038/nature02806.

[83] M. Babcock, S. Yatsenko, P. Stankiewicz, J.R. Lupski, B.E. Morrow, AT-rich repeats associated with chromosome 22q11.2 rearrangement disorders shape human genome architecture on Yq12, Genome Res. 17 (2007) 451–460. https://doi.org/10.1101/gr.5651507.

[84] A.P. Craig-Holmes, M.W. Shaw, Polymorphism of Human Constitutive Heterochromatin, Science. 174 (1971) 702–704. https://doi.org/10.1126/science.174.4010.702.

[85] M.A. Bender, P.C. Gooch, AN UNUSUALLY LONG HUMAN Y CHROMOSOME, Lancet. 278 (1961) 463–464. https://doi.org/10.1016/s0140-6736(61)92433-3.

[86] U. Gripenberg, Size variation and orientation of the human Y chromosome, Chromosoma. 15 (1964) 618–629. https://doi.org/10.1007/bf00319995.

[87] V. Unnérus, J. Fellman, A. de la Chapelle, The Length of the Human Y Chromosome, Cytogenetic and Genome Research. 6 (1967) 213–227. https://doi.org/10.1159/000129943.

[88] M. Bobrow, P.L. Pearson, M.C. Pike, O.S. el-Alfi, Length variation in the quinacrine-binding segment of human Y chromosomes of different sizes., Cytogenetics. 10 (1971) 190–8. https://doi.org/10.1159/000130138.

[89] W.H. McKenzie, T.L. Hostetter, H.A. Lubs, Y family study: Heritable variation in the length of the human Y chromosome, American Journal of Human Genetics. 24 (1972) 686.

[90] C. Laberge, R. Gagne, Quinacrine mustard staining solves the length variations of the human Y chromosome., Johns Hopkins Medical J. 128 (1971) 79–83.

[91] D.M. Kurnit, Satellite DNA and heterochromatin variants: The case for unequal mitotic crossing over, Hum Genet. 47 (1979) 169–186. https://doi.org/10.1007/bf00273199.

[92] A.P. Craig-Holmes, F.B. Moore, M.W. Shaw, Polymporphism of human C-band heterochromatin. II. Family studies with suggestive evidence for somatic crossing over., Am J Hum Genet. 27 (1975) 178–89.

[93] O.A. Podugolnikova, A.P. Korostelev, The quantitative analysis of polymorphism on human chromosomes 1,9,16, and Y. IV. Heterogeneity of a normal population., Hum Genet. 54 (1980) 163–9.

[94] L.Y.F. Hsu, P.A. Benn, H.L. Tannenbaum, T.E. Perlis, A.D. Carlson, J.M. Opitz, J.F. Reynolds, Chromosomal polymorphisms of 1, 9, 16, and Y in 4 major ethnic groups: A large prenatal study, Am J Med Genet. 26 (1987) 95–101. https://doi.org/10.1002/ajmg.1320260116.

[95] C. Fowler, R. Drinkwater, L. Burgoyne, J. Skinner, Hypervariable lengths of human DNA associated with a human satellite III sequence found in the 3, 4kb Y-specific fragment, Nucleic Acids Res. 15 (1987) 3929–3929. https://doi.org/10.1093/nar/15.9.3929.

[96] A.H. Sturtevant, THE EFFECTS OF UNEQUAL CROSSING OVER AT THE BAR LOCUS IN DROSOPHILA, Genetics. 10 (1925) 117–147. https://doi.org/10.1093/genetics/10.2.117.

[97] C.H. Gravholt, U. Friedrich, M. Caprani, A.L. Jørgensen, Breakpoints in Robertsonian translocations are localized to satellite III DNA by fluorescence in situ hybridization, Genomics. 14 (1992) 924–930. https://doi.org/10.1016/s0888-7543(05)80113-2.

[98] E. Earle, L.G. Shaffer, P. Kalitsis, C. McQuillan, S. Dale, K.H. Choo, Identification of DNA sequences flanking the breakpoint of human t(14q21q) Robertsonian translocations., Am J Hum Genet. 50 (1992) 717–24.

[99] A. Fournier, A. McLeer-Florin, C. Lefebvre, S. Duley, L. Barki, J. Ribeyron, K. Alboukadel, S. Hamaidia, A. Granjon, R. Gressin, A. Lajmanovich, T. Bonnefoix, S. Chauvelier, A. Debernardi, S. Rousseaux, F. de Fraipont, M. Figeac, J. Kerckaert, J.D. Vos, Y. Usson, K. Delaval, A. Grichine, C. Vourc'h, S. Khochbin, R. Feil, D. Leroux, M.B. Callanan, 1q12 chromosome translocations form aberrant heterochromatic foci associated with changes in nuclear architecture and gene expression in B cell lymphoma, Embo Mol Med. 2 (2010) 159–171. https://doi.org/10.1002/emmm.201000067.

[100] G.P. Smith, Evolution of Repeated DNA Sequences by Unequal Crossover, Science. 191 (1976) 528–535. https://doi.org/10.1126/science.1251186.

[101] J. Krüger, F. Vogel, Population genetics of unequal crossing over, J Mol Evol. 4 (1975) 201–247. https://doi.org/10.1007/bf01732983.

[102] K.D. Tartof, Unequal mitotic sister chromatid exchange and disproportionate replication as mechanisms regulating ribosomal RNA gene redundancy., Cold Spring Harb Sym. 38 (1974) 491–500. https://doi.org/10.1101/sqb.1974.038.01.053.

[103] T.D. Petes, Unequal meiotic recombination within tandem arrays of yeast ribosomal DNA genes., Cell. 19 (1980) 765–74. https://doi.org/10.1016/s0092-8674(80)80052-3.

[104] J.W. Szostak, R. Wu, Unequal crossing over in the ribosomal DNA of Saccharomyces cerevisiae, Nature. 284 (1980) 426–430. https://doi.org/10.1038/284426a0.

[105] W.K. Baker, Crossing Over in Heterochromatin, Am Nat. 92 (1958) 59–60. https://doi.org/10.1086/282010.

[106] G. Dover, Molecular drive: a cohesive mode of species evolution, Nature. 299 (1982) 111–117. https://doi.org/10.1038/299111a0.

[107] J.F. Elder, B.J. Turner, Concerted evolution of repetitive DNA sequences in eukaryotes., Q Rev Biology. 70 (1995) 297–320. https://doi.org/10.1086/419073.

[108] F. Bersani, E. Lee, P.V. Kharchenko, A.W. Xu, M. Liu, K. Xega, O.C. MacKenzie, B.W. Brannigan, B.S. Wittner, H. Jung, S. Ramaswamy, P.J. Park, S. Maheswaran, D.T. Ting, D.A. Haber, Pericentromeric satellite repeat expansions through RNA-derived DNA intermediates in cancer, Proc National Acad Sci. 112 (2015) 15148–15153. https://doi.org/10.1073/pnas.1518008112.

[109] S. Cohen, N. Agmon, O. Sobol, D. Segal, Extrachromosomal circles of satellite repeats and 5S ribosomal DNA in human cells, Mobile Dna-Uk. 1 (2010) 11–11. https://doi.org/10.1186/1759-8753-1-11.

[110] W.R. Rice, A Game of Thrones at Human Centromeres II. A new molecular/evolutionary model, Biorxiv. (2019) 731471. https://doi.org/10.1101/731471.

[111] P.B. Talbert, S. Henikoff, The genetics and epigenetics of satellite centromeres., Genome Res. (2022). https://doi.org/10.1101/gr.275351.121.

[112] C. Tyler-Smith, The Neglected Parts of The Genome: Half a Century of Satellite DNA, in: R.K. Mishra (Ed.), Chromosomes to Genome, International Pub. House, New Delhi, 2009.

[113] M.R. Vollger, P. Kerpedjiev, A.M. Phillippy, E.E. Eichler, StainedGlass: interactive visualization of massive tandem repeat structures with identity heatmaps, Bioinformatics. (2022). https://doi.org/10.1093/bioinformatics/btac018.

[114] A.R. Mitchell, H.N. Seuanez, S.S. Lawrie, D.E. Martin, J.R. Gosden, The location of DNA homologous to human satellite III DNA in the chromosomes of chimpanzee (Pan troglodytes), gorilla (Gorilla gorilla) and orang utan (Pongo pygmaeus)., Chromosoma. 61 (1977) 345–58. https://doi.org/10.1007/bf00288618.

[115] J.R. Gosden, A.R. Mitchell, H.N. Seuanez, C.M. Gosden, The distribution of sequences complementary to human satellite DNAs I, II and IV in the chromosomes of chimpanzee (Pan troglodytes), gorilla (Gorilla gorilla) and orang utan (Pongo pygmaeus), Chromosoma. 63 (1977) 253–271. https://doi.org/10.1007/bf00327453.

[116] A.R. Mitchell, J.R. Gosden, O.A. Ryder, Satellite DNA relationships in man and the primates, Nucleic Acids Res. 9 (1981) 3235–3249. https://doi.org/10.1093/nar/9.14.3235.

[117] B. Gläser, F. Grützner, U. Willmann, R. Stanyon, N. Arnold, K. Taylor, W. Rietschel, S. Zeitler, R. Toder, W. Schempp, Simian Y Chromosomes: species-specific rearrangements of DAZ, RBM, and TSPY versus contiguity of PAR and SRY, Mamm Genome. 9 (1998) 226–231. https://doi.org/10.1007/s003359900730.

[118] M. Pita, J. Gosálvez, A. Gosálvez, M. Nieddu, C. López-Fernández, R. Mezzanotte, A Highly Conserved Pericentromeric Domain in Human and Gorilla Chromosomes, Cytogenet Genome Res. 126 (2010) 253–258. https://doi.org/10.1159/000251962.

[119] M. Jarmuż, C.D. Glotzbach, K.A. Bailey, R. Bandyopadhyay, L.G. Shaffer, The Evolution of Satellite III DNA Subfamilies among Primates, Am J Hum Genetics. 80 (2007) 495–501. https://doi.org/10.1086/512132.

[120] M. Cechova, R.S. Harris, M. Tomaszkiewicz, B. Arbeithuber, F. Chiaromonte, K.D. Makova, High Satellite Repeat Turnover in Great Apes Studied with Short- and Long-Read Technologies, Mol Biol Evol. 36 (2019) 2415–2431. https://doi.org/10.1093/molbev/msz156.

[121] D.L. Grady, R.L. Ratliff, D.L. Robinson, E.C. McCanlies, J. Meyne, R.K. Moyzis, Highly conserved repetitive DNA sequences are present at human centromeres., Proc National Acad Sci. 89 (1992) 1695–1699. https://doi.org/10.1073/pnas.89.5.1695.

[122] O. Vafa, K.F. Sullivan, Chromatin containing CENP-A and α-satellite DNA is a major component of the inner kinetochore plate, Curr Biol. 7 (1997) 897–900. https://doi.org/10.1016/s0960-9822(06)00381-2.

[123] J.J. Yunis, W.G. Yasmineh, Heterochromatin, Satellite DNA, and Cell Function: Structural DNA of eucaryotes may support and protect genes and aid in speciation., Science. 174 (1971) 1200–1209. https://doi.org/10.1126/science.174.4015.1200.

[124] M. Jagannathan, R. Cummings, Y.M. Yamashita, A conserved function for pericentromeric satellite DNA, Elife. 7 (2018) e34122. https://doi.org/10.7554/elife.34122.

[125] M. Jagannathan, R. Cummings, Y.M. Yamashita, The modular mechanism of chromocenter formation in Drosophila., Elife. 8 (2019) e43938. https://doi.org/10.7554/elife.43938.

[126] B.W. Lubit, P.T. Duc, O.J. Miller, B.F. Erlanger, Localization of 5-methylcytosine in human metaphase chromosomes by immunoelectron microscopy, Cell. 9 (1976) 503–509. https://doi.org/10.1016/0092-8674(76)90032-5.

[127] E. Viegas-Pequignot, B. Dutrillaux, Segmentation of human chromosomes induced by 5-ACR (5-azacytidine), Hum Genet. 34 (1976) 247–254. https://doi.org/10.1007/bf00295287.

[128] R. Hernandez, A. Frady, X.Y. Zhang, M. Varela, M. Ehrlich, Preferential induction of chromosome 1 multibranched figures and whole-arm deletions in a human pro-B cell line treated with 5-azacytidine or 5-azadeoxycytidine., Cytogenet Cell Genet. 76 (1997) 196–201. https://doi.org/10.1159/000134548.

[129] N. Kokalj-Vokac, A. Almeida, E. Viegas-Péquignot, M. Jeanpierre, B. Malfoy, B. Dutrillaux, Specific induction of uncoiling and recombination by azacytidine in classical

satellite-containing constitutive heterochromatin., Cytogenet Cell Genet. 63 (1993) 11–5. https://doi.org/10.1159/000133492.

[130] T. Haaf, M. Schmid, Experimental condensation inhibition in constitutive and facultative heterochromatin of mammalian chromosomes, Cytogenet Genome Res. 91 (2000) 113–123. https://doi.org/10.1159/000056830.

[131] N.I. Enukashvily, R. Donev, I.S.-R. Waisertreiger, O.I. Podgornaya, Human chromosome 1 satellite 3 DNA is decondensed, demethylated and transcribed in senescent cells and in A431 epithelial carcinoma cells, Cytogenet Genome Res. 118 (2007) 42–54. https://doi.org/10.1159/000106440.

[132] E.C. Swanson, B. Manning, H. Zhang, J.B. Lawrence, Higher-order unfolding of satellite heterochromatin is a consistent and early event in cell senescence, J Cell Biol. 203 (2013) 929–942. https://doi.org/10.1083/jcb.201306073.

[133] R.B. Slee, C.M. Steiner, B.-S. Herbert, G.H. Vance, R.J. Hickey, T. Schwarz, S. Christan, M. Radovich, B.P. Schneider, D. Schindelhauer, B.R. Grimes, Cancer-associated alteration of pericentromeric heterochromatin may contribute to chromosome instability, Oncogene. 31 (2012) 3244–3253. https://doi.org/10.1038/onc.2011.502.

[134] A. Almeida, N. Kokalj-Vokac, D. Lefrancois, E. Viegas-Pequignot, M. Jeanpierre, B. Dutrillaux, B. Malfoy, Hypomethylation of classical satellite DNA and chromosome instability in lymphoblastoid cell lines, Hum Genet. 91 (1993) 538–46. https://doi.org/10.1007/bf00205077.

[135] C. Yandım, G. Karakülah, Expression dynamics of repetitive DNA in early human embryonic development, Bmc Genomics. 20 (2019) 439. https://doi.org/10.1186/s12864-019-5803-1.

[136] L.L. Hall, M. Byron, D.M. Carone, T.W. Whitfield, G.P. Pouliot, A. Fischer, P. Jones, J.B. Lawrence, Demethylated HSATII DNA and HSATII RNA Foci Sequester PRC1 and MeCP2 into Cancer-Specific Nuclear Bodies., Cell Reports. 18 (2017) 2943–2956. https://doi.org/10.1016/j.celrep.2017.02.072.

[137] Q. Zhu, N. Hoong, A. Aslanian, T. Hara, C. Benner, S. Heinz, K.H. Miga, E. Ke, S. Verma, J. Soroczynski, J.R. Yates, T. Hunter, I.M. Verma, Heterochromatin-Encoded Satellite RNAs Induce Breast Cancer, Mol Cell. 70 (2018) 842-853.e7. https://doi.org/10.1016/j.molcel.2018.04.023.

[138] D.T. Ting, D. Lipson, S. Paul, B.W. Brannigan, S. Akhavanfard, E.J. Coffman, G. Contino, V. Deshpande, A.J. Iafrate, S. Letovsky, M.N. Rivera, N. Bardeesy, S. Maheswaran, D.A. Haber, Aberrant Overexpression of Satellite Repeats in Pancreatic and Other Epithelial Cancers, Science. 331 (2011) 593–596. https://doi.org/10.1126/science.1200801.

[139] Q. Zhu, G.M. Pao, A.M. Huynh, H. Suh, N. Tonnu, P.M. Nederlof, F.H. Gage, I.M. Verma, BRCA1 tumour suppression occurs via heterochromatin-mediated silencing, Nature. 477 (2011) 179–184. https://doi.org/10.1038/nature10371.

[140] C. Yandım, G. Karakülah, Dysregulated expression of repetitive DNA in ER+/HER2- breast cancer, Cancer Genet-Ny. 239 (2019) 36–45. https://doi.org/10.1016/j.cancergen.2019.09.002.

[141] M.T. Nogalski, T. Shenk, HSATII RNA is induced via a noncanonical ATM-regulated DNA damage response pathway and promotes tumor cell proliferation and movement, Proc National Acad Sci. 117 (2020) 31891–31901. https://doi.org/10.1073/pnas.2017734117.

[142] M.T. Nogalski, A. Solovyov, A.S. Kulkarni, N. Desai, A. Oberstein, A.J. Levine, D.T. Ting, T. Shenk, B.D. Greenbaum, A tumor-specific endogenous repetitive element is induced by herpesviruses, Nat Commun. 10 (2019) 90. https://doi.org/10.1038/s41467-018-07944-x.

[143] C. Jolly, A. Metz, J. Govin, M. Vigneron, B.M. Turner, S. Khochbin, C. Vourc'h, Stress-induced transcription of satellite III repeats, J Cell Biol. 164 (2003) 25–33. https://doi.org/10.1083/jcb.200306104.

[144] N. Rizzi, M. Denegri, I. Chiodi, M. Corioni, R. Valgardsdottir, F. Cobianchi, S. Riva, G. Biamonti, Transcriptional activation of a constitutive heterochromatic domain of the human genome in response to heat shock., Mol Biol Cell. 15 (2003) 543–51. https://doi.org/10.1091/mbc.e03-07-0487.

[145] S. Sengupta, R. Parihar, S. Ganesh, Satellite III non-coding RNAs show distinct and stress-specific patterns of induction, Biochem Bioph Res Co. 382 (2009) 102–107. https://doi.org/10.1016/j.bbrc.2009.02.137.

[146] R. Valgardsdottir, I. Chiodi, M. Giordano, A. Rossi, S. Bazzini, C. Ghigna, S. Riva, G. Biamonti, Transcription of Satellite III non-coding RNAs is a general stress response in human cells, Nucleic Acids Res. 36 (2008) 423–434. https://doi.org/10.1093/nar/gkm1056.

[147] C. Jolly, L. Konecny, D.L. Grady, Y.A. Kutskova, J.J. Cotto, R.I. Morimoto, C. Vourc'h, In vivo binding of active heat shock transcription factor 1 to human chromosome 9 heterochromatin during stress, J Cell Biology. 156 (2002) 775–781. https://doi.org/10.1083/jcb.200109018.

[148] A. Eymery, C. Souchier, C. Vourc'h, C. Jolly, Heat shock factor 1 binds to and transcribes satellite II and III sequences at several pericentromeric regions in heat-shocked cells, Exp Cell Res. 316 (2010) 1845–1855. https://doi.org/10.1016/j.yexcr.2010.02.002.

[149] N.H. Brückmann, C.B. Pedersen, H.J. Ditzel, M.F. Gjerstorff, Epigenetic Reprogramming of Pericentromeric Satellite DNA in Premalignant and Malignant Lesions, Mol Cancer Res. 16 (2018) 417–427. https://doi.org/10.1158/1541-7786.mcr-17-0477.

[150] M.K. Aly, K. Ninomiya, S. Adachi, T. Natsume, T. Hirose, Two distinct nuclear stress bodies containing different sets of RNA-binding proteins are formed with HSATIII architectural noncoding RNAs upon thermal stress exposure, Biochem Bioph Res Co. 516 (2019) 419–423. https://doi.org/10.1016/j.bbrc.2019.06.061.

[151] K. Ninomiya, S. Adachi, T. Natsume, J. Iwakiri, G. Terai, K. Asai, T. Hirose, LncRNA-dependent nuclear stress bodies promote intron retention through SR protein phosphorylation, Embo J. 39 (2020) e102729. https://doi.org/10.15252/embj.2019102729.

[152] K. Ninomiya, J. Iwakiri, M.K. Aly, Y. Sakaguchi, S. Adachi, T. Natsume, G. Terai, K. Asai, T. Suzuki, T. Hirose, m6A modification of HSATIII lncRNAs regulates temperature-dependent splicing, Embo J. 40 (2021) e107976. https://doi.org/10.15252/embj.2021107976.

[153] M. Giordano, L. Infantino, M. Biggiogera, A. Montecucco, G. Biamonti, Heat Shock Affects Mitotic Segregation of Human Chromosomes Bound to Stress-Induced Satellite III RNAs, Int J Mol Sci. 21 (2020) 2812. https://doi.org/10.3390/ijms21082812.

[154] C.C. Landers, C.A. Rabeler, E.K. Ferrari, L.R. D'Alessandro, D.D. Kang, J. Malisa, S.M. Bashir, D.M. Carone, Ectopic expression of pericentric HSATII RNA results in nuclear RNA accumulation, MeCP2 recruitment, and cell division defects, Chromosoma. 130 (2021) 75–90. https://doi.org/10.1007/s00412-021-00753-0.

[155] S.J. Hoyt, J.M. Storer, G.A. Hartley, P.G.S. Grady, A. Gershman, L.G. de Lima, C. Limouse, R. Halabian, L. Wojenski, M. Rodriguez, N. Altemose, A. Rhie, L.J. Core, J.L. Gerton, W. Makalowski, D. Olson, J. Rosen, A.F.A. Smit, A.F. Straight, M.R. Vollger, T.J. Wheeler, M.C. Schatz, E.E. Eichler, A.M. Phillippy, W. Timp, K.H. Miga, R.J. O'Neill, From telomere to telomere: The transcriptional and epigenetic state of human repeat elements., Sci New York N Y. 376 (2022) eabk3112. https://doi.org/10.1126/science.abk3112.

[156] G. Bosco, P. Campbell, J.T. Leiva-Neto, T.A. Markow, Analysis of Drosophila Species Genome Size and Satellite DNA Content Reveals Significant Differences Among Strains as Well as Between Species, Genetics. 177 (2007) 1277–1290. https://doi.org/10.1534/genetics.107.075069.

[157] M. Jagannathan, N. Warsinger-Pepe, G.J. Watase, Y.M. Yamashita, Comparative Analysis of Satellite DNA in the Drosophila melanogaster Species Complex, G3 Genes Genomes Genetics. 7 (2017) 693–704. https://doi.org/10.1534/g3.116.035352.

[158] K. Sawamura, M.T. Yamamoto, T.K. Watanabe, Hybrid lethal systems in the Drosophila melanogaster species complex. II. The Zygotic hybrid rescue (Zhr) gene of D. melanogaster., Genetics. 133 (1993) 307–13. https://doi.org/10.1093/genetics/133.2.307.

[159] P.M. Ferree, D.A. Barbash, Species-Specific Heterochromatin Prevents Mitotic Chromosome Segregation to Cause Hybrid Lethality in Drosophila, Plos Biol. 7 (2009) e1000234. https://doi.org/10.1371/journal.pbio.1000234.

[160] M. Jagannathan, Y.M. Yamashita, Defective Satellite DNA Clustering into Chromocenters Underlies Hybrid Incompatibility in Drosophila, Mol Biol Evol. 38 (2021) msab221. https://doi.org/10.1093/molbev/msab221.

[161] J. Prosser, A.H. Reisner, M.L. Bradley, K. Ho, P.C. Vincent, Buoyant density and hybridization analysis of human DNA sequences, including three satellite DNAs, Biochimica Et Biophysica Acta Bba - Nucleic Acids Protein Synthesis. 656 (1981) 93–102. https://doi.org/10.1016/0005-2787(81)90031-9.

[162] G. Corneo, L. Zardi, E. Polli, Elution of human satellite DNAs on a methylated albumin kieselguhr chromatographic column:Isolation of satellite DNA IV, Biochimica Et Biophysica Acta Bba - Nucleic Acids Protein Synthesis. 269 (1972) 201–204. https://doi.org/10.1016/0005-2787(72)90427-3.

[163] V. Paar, I. Basar, M. Rosandic, M. Gluncic, Consensus Higher Order Repeats and Frequency of String Distributions in Human Genome, Curr Genomics. 8 (2007) 93–111. https://doi.org/10.2174/138920207780368169.

[164] B. Gel, E. Serra, karyoploteR: an R/Bioconductor package to plot customizable genomes displaying arbitrary data, Bioinformatics. 33 (2017) 3088–3090. https://doi.org/10.1093/bioinformatics/btx346.

[165] Z. Gu, L. Gu, R. Eils, M. Schlesner, B. Brors, circlize implements and enhances circular visualization in R, Bioinformatics. 30 (2014) 2811–2812. https://doi.org/10.1093/bioinformatics/btu393.