Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 February 2022 doi:10.20944/preprints202202.0009.v1

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, collectively HSat1-3) constitute the largest individual arrays of tandemly repeated DNA sequences in the genome. Even though they were among the first human DNA sequences to be isolated and characterized at the dawn of molecular biology, HSat1-3 have been left behind in the genomics era and remain among the most enigmatic sequences in the human genome. Although HSat1-3 total roughly 3% of the genome on average, they were almost entirely missing from the human genome reference assembly for 20 years. Recently, the Telomere-to-Telomere Consortium produced the first truly complete assembly of a human genome, including the enormous HSat1-3 arrays, opening them up for a new wave of discovery. Towards this end, here, I provide an account of the history and current understanding of HSat1-3 genomics, evolution, and roles in 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, sometimes encompassing millions of base pairs. In humans, satellite DNA represents 5-10% of the genome (Altemose, Logsdon, et al. 2021, Miga 2019, C. W. Schmid & Deininger 1975), found primarily around all centromeres, 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 (Yasmineh & Yunis 1969), which is characteristically more condensed than euchromatic regions (Heitz 1928). Satellite DNA arrays can be classified into distinct families based on their sequence composition. In the human genome, the largest satellite family by total size genome-wide is alpha satellite DNA (αSat), which encompasses every centromere and plays a key role in the establishment and maintenance of centromere function (Manuelidis 1976, Schueler et al. 2001).

The next largest families by total size, which also comprise the largest individual satellite arrays in the genome, are referred to collectively as the classical human satellites, or individually as Human Satellites 1-3 (HSat1-3) (Prosser et al. 1986). However, the potential functions of HSat1-3 remain poorly understood, in part because satellite DNA arrays were intractable to early DNA amplification and sequence assembly methods (Eichler et al. 2004). As a result, although the classical human satellites were among the first human DNA sequences to be biochemically

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 February 2022 doi:10.20944/preprints202202.0009.v1

characterized and isolated from the rest of the nuclear genome (Corneo et al. 1967, 1971), they have remained almost entirely missing from the human reference sequence for the last 20 years. However, thanks to the recent efforts of the Telomere-to-Telomere (T2T) Consortium, all satellite DNA arrays have now been fully assembled in a human genome for the first time (Altemose, Logsdon, et al. 2021, Nurk et al. 2021), opening them up for 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

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 (Kit 1961, Meselson et al. 1957, Sueoka 1961) (Figure 1). 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 (Sueoka 1961). However, in 1961 Saul Kit discovered that mouse and guinea pig DNA formed a second, smaller, "satellite" DNA band outside the major DNA band (Kit 1961) (*footnote 1*), and a similar, AT-rich satellite DNA band was later discovered in humans (Corneo et al. 1967). 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 (Corneo et al. 1967, 1970, 1971) (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 (Britten & Kohne 1968, Corneo et al. 1970, 1971, Kunkel et al. 1976, Waring & Britten 1966). Satellite DNAs were found to be enriched in heterochromatic fractions of DNA obtained by centrifugation of chromatin (Corneo et al. 1971, Yasmineh & Yunis 1969), and in situ hybridization experiments revealed that satellite DNAs are enriched in pericentromeric heterochromatin (Arrighi et al. 1971, Jones 1970, Jones & Corneo 1971, Pardue & Gall 1970), with the largest blocks in humans found on chromosomes 1, 9, 16, and Y (Gosden et al. 1975, Jones et al. 1973, 1974, Jones & Corneo 1971, Saunders et al. 1972, Tagarro, Fernández-Peralta, et al. 1994, Tagarro, Wiegant, et al. 1994).

Satellite DNA served as an original testing ground for methods like degradation-based DNA sequencing, Southern blotting, and in situ hybridization (Jones 1970, Pardue & Gall 1970, Southern 1970, 1975b, 1975a). 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 (Cooke 1976, Manuelidis 1976, Philippsen et al. 1974, Shen et al. 1976, Southern 1975b). Restriction enzymes bind and cut specific short DNA sequences (enzymes recognizing 4-6 bp sites were typically used) (Kelly & Smith 1970, H. O. Smith & Welcox 1970). In tandem repeats, these restriction sites 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) (Southern 1975b). In non-repetitive regions, the resulting fragment lengths are expected to follow a fairly continuous exponential distribution. By separating and quantifying digested DNA

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 February 2022 doi:10.20944/preprints202202.0009.v1

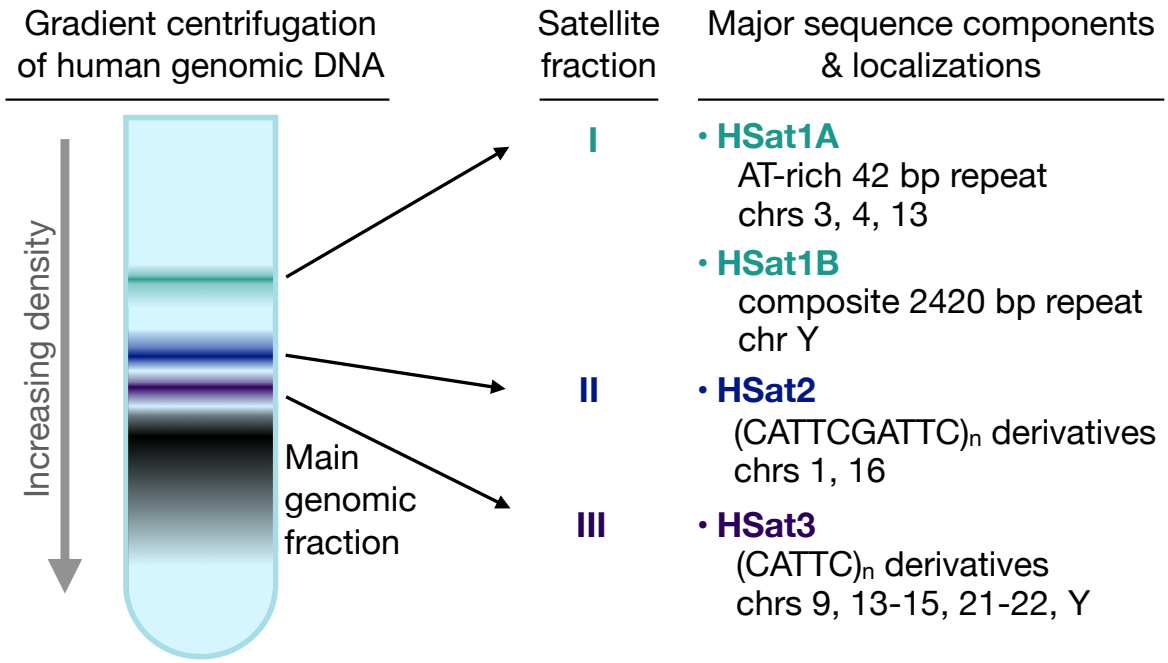

Figure 1. The origins of the classical human satellites. A schematic depicts early experiments in which high-molecular-weight human genomic DNA was fractionated by ultracentrifugation in cesium gradients. Three robust "satellite" fractions (I, II, and III) could be separated from the main genomic fraction, 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).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 February 2022 doi:10.20944/preprints202202.0009.v1

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 (Philippsen et al. 1974). With these approaches, the sizes of some of the major repeating units in each satellite fraction could be determined (Beauchamp et al. 1979, Cooke 1976, Cooke et al. 1982, Cooke & Hindley 1979, Frommer et al. 1982, 1984, Higgins et al. 1985, Mitchell et al. 1979, Prosser et al. 1986), and early work by Howard J. Cooke demonstrated that different repeating units could be found on different chromosomes (Cooke 1976, Cooke et al. 1982, Cooke & Hindley 1979, Cooke & McKay 1978).

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 (Manuelidis 1976). 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 (Pech et al. 1979, Singer 1982). 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 (Singer 1982).

Recombinant DNA and more efficient sequencing technologies (Maxam & Gilbert 1980, Sanger et al. 1977) enabled the next wave of human satellite repeat discovery. 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 (Frommer et al. 1982, 1984, Prosser et al. 1986) (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 (Prosser et al. 1986), and a 2.5 kb repeat found predominantly on the Y chromosome (Cooke 1976, Cooke et al. 1982, Frommer et al. 1984). 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 (Frommer et al. 1982, Prosser et al. 1986). 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 comprising the majority of human satellite fraction III (Prosser et al. 1986). 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) (Altemose, Logsdon, et al. 2021).

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 (Meyne et al. 1994, Tagarro, Wiegant, et al. 1994). HSat2 probes hybridized primarily to chromosomes 1 and 16, with smaller domains on chromosomes 2, 7, 10, 15, 17, and 22 (Tagarro, Fernández-Peralta, et al. 1994). HSat3 probes hybridized primarily

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 February 2022 doi:10.20944/preprints202202.0009.v1

to chromosome 9, with smaller domains on chromosomes 1, 5, 10, 17, 20, and the acrocentrics (13, 14, 15, 21, 22) (Tagarro, Fernández-Peralta, et al. 1994). 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 (Cooke 1976, Cooke et al. 1982, Cooke & McKay 1978, Manz et al. 1992, Mathias et al. 1994, McKay et al. 1978, Nakahori et al. 1986, M. Schmid et al. 1990). Additional studies isolated clones of HSat2 and HSat3 that hybridized to individual chromosomes or subsets of chromosomes (Bandyopadhyay et al. 2001, Choo et al. 1990, 1992, Cooke & Hindley 1979, Deininger et al. 1981, Higgins et al. 1985, M S Jackson et al. 1992, Michael S. Jackson et al. 1993, Jeanpierre 1994, Jeanpierre et al. 1985, Kalitsis et al. 1993, Legrand et al. 1998, Moyzis et al. 1987, Vissel et al. 1992). 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 (Sauter et al. 1995).

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 relevant for centromere function, while the potential functions of the classical satellites remained mysterious.

3. HSat1-3 in the Genomics Era

The repetitive, heterochromatic regions of the genome posed an intractable problem for the Human Genome Project (Consortium 2001) and for Celera Genomics' separate efforts to assemble the human genome (Venter et al. 2001). The Human Genome Project decided to use a hierarchical sequencing approach, in which large genomic fragments were cloned in bacteria 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 in bacteria 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; 3) assembling repetitive DNA from short sequencing reads even within a single genomic fragment remained challenging, as near-identical repeat units often exceeded the read length (Eichler et al. 2004). 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 had an even more difficult task of assembling repetitive regions from sequenced fragments across the entire genome. Thus, neither initial human genome sequencing effort, nor any that followed for 20 years, succeeded in assembling across HSat1-3 arrays in the human genome, leaving their approximate locations in the genome assembly as enormous gaps filled with placeholder "N" characters.

Occasionally, the 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 (Altemose et al. 2014, Warburton et al. 2008) (*footnote 3*). Whole-genome shotgun assemblies would often produce partial, unmapped, scaffold assemblies of subregions of satellite arrays (Levy et al. 2007). These

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 February 2022 doi:10.20944/preprints202202.0009.v1

unmapped scaffolds could sometimes be mapped to specific chromosomes using chromosome-specific sequencing libraries (Altemose et al. 2014), and those that contained unique marker sequences could often be mapped to specific loci using 3D contact information (Kaplan & Dekker 2013) or a clever approach based on genetic information from individuals with recently admixed ancestry (Genovese et al. 2013). More focused approaches identified additional HSat2 and HSat3 arrays on individual chromosomes (Babcock et al. 2007, Guy et al. 2003, She et al. 2004, Skaletsky et al. 2003). However, these approaches still failed to represent the classical human satellites comprehensively.

To address this, in a 2014 study, colleagues and I developed an alignment-free approach for characterizing all of the HSat2 and HSat3 sequences from a single individual's raw WGS reads (Altemose et al. 2014). To do so, we converted the sequence of each HSat2 or HSat3 read in the HuRef genome (Levy et al. 2007) 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 among even shorter next-generation sequencing reads (Altemose et al. 2014). 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.

In 2021, the Telomere-to-Telomere (T2T) Consortium released the first complete assembly of a human genome, including all autosomal HSat1-3 arrays (Altemose, Logsdon, et al. 2021, Nurk et al. 2021). 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 (compared to the largest αSat array in the genome: 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. Colleagues and I worked to identify the subfamily components of each HSat2 and HSat3 array in this new reference and compared their localization with previous predictions (Altemose, Logsdon, et al. 2021) (Figure 2). The assembly confirmed the chromosomal assignments predicted previously 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.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 February 2022 doi:10.20944/preprints202202.0009.v1

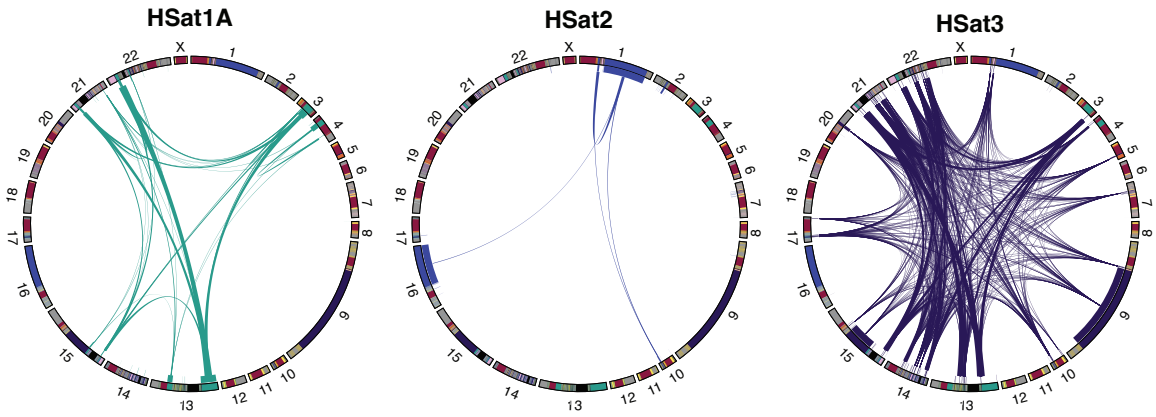

Figure 2. Annotation of HSat1-3 in a complete T2T genome assembly. Circos plots show color-coded annotations of the pericentromeric regions of all chromosomes in the T2T-CHM13 assembly. The locations of HSat1A, HSat2, or HSat3 arrays are highlighted in an inner concentric ring. In the middle of each circle, the width of lines connecting the arrays is scaled according to the proportion of exact 75-mer sequence matches between the arrays, serving as an overall estimate of sequence similarity (no line is drawn below a fixed threshold). Overall, HSat3 appears to be the least diverged satellite, while HSat2 appears to be the most diverged, consistent with published findings (Prosser et al. 1986). [note: chrY and HSat1B will be added before final proof stage. I also plan to add a table with more detailed information about each HSat array in the genome].

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 February 2022 doi:10.20944/preprints202202.0009.v1

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). 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. 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) akin to the analytical restriction digest experiments used to characterize satellite arrays in the past, and similar in approach to previous computational tools for detecting repeat periodicity (Paar et al. 2007). This algorithm takes advantage of the complete information in a full array assembly by essentially simulating restriction digests with enzymes that could cut all possible recognition sites. Then, it combines information across these simulations to reveal the predominant periodicities within the array. By running NTRprism in windows across an array, variation in periodicity can be uncovered. 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. The Y chromosome remains missing from this assembly, as it is not present in the CHM13 cell line, although a T2T assembly of chrY from a different cell line is promised in the near future. This will enable a detailed examination of the enormous HSat3 and HSat1B arrays present on chrY.

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.

4. Evolution and variation of HSat1-3

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 (Ann P Craig-Holmes & Shaw 1971, Jones & Corneo 1971). These effects could be so dramatic as to visibly change the size of entire chromosomes, such as the Y chromosome (Bender & Gooch 1961, Bobrow et al. 1971, Gripenberg 1964, Laberge & Gagne 1971, McKenzie et al. 1972, Unnérus et al. 1967). Once satellite DNA was associated with heterochromatin, it became clear that something about the evolution of satellite DNA must explain these enormous size polymorphisms (Ann P Craig-Holmes & Shaw 1971, Kurnit 1979, McKay et al. 1978). Because large heteromorphisms in HSat1-3 arrays were frequently present in individuals without congenital diseases or other health issues (Choo et al. 1992, A P Craig-Holmes et al. 1975, Fowler et al. 1987, Hsu et al. 1987, Manz et al. 1992, McKay et al. 1978, McKenzie et al. 1972, Podugolnikova & Korostelev 1980), 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 (Jones & Corneo 1971). The high degree of polymorphism also suggested a high rate of structural rearrangements in these regions, which could not be explained by the molecular mechanisms predominating mutations in the rest of the genome.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 February 2022 doi:10.20944/preprints202202.0009.v1

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 (A P Craig-Holmes et al. 1975, Ann P Craig-Holmes & Shaw 1971) (Figure 3a). That is, when a double-strand break occurs during the S or G2 phases of the cell cycle, the favored 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 satellite 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 throughout the entire array, and the correct homologous stretch cannot be determined. As a result, sometimes the crossover occurs between the break site and a non-homologous site at a different position within the sister chromatid's array, producing two recombinant chromatids of different lengths: one long, and one short. If unequal crossover events occur in the germline, these expanded or contracted arrays can be passed on to offspring. Occasionally unequal crossover results in non-allelic recombination events between classical human satellite arrays on different chromosomes, yielding translocations that can cause congenital diseases or contribute to cancer (Babcock et al. 2007, Earle et al. 1992, Fournier et al. 2010, Gravholt et al. 1992, Kalitsis et al. 1993).

Additional support for the unequal crossover hypothesis stemmed from theoretical predictions and experimental observations of tandem ribosomal gene arrays in yeast (Krüger & Vogel 1975, Kurnit 1979, Petes 1980, G. P. Smith 1976, Szostak & Wu 1980, Tartof 1974). Mitotic, rather than meiotic, crossover is the favored mechanism because crossovers are known to be suppressed in heterochromatin during meiotic recombination (Baker 1958, A P Craig-Holmes et al. 1975). 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 (reviewed by (Elder & Turner 1995)). Alternative proposed satellite expansion mechanisms include expansion via RNA-derived intermediates (Bersani et al. 2015) and expansion via rolling circle amplification (Cohen et al. 2010).

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, composed primarily of interspersed HSat1B and HSat3A6 arrays, finding at least one mutation per 40 Mb per meiosis, which is one of the highest mutation rates ever reported in the genome (Mathias et al. 1994, Tyler-Smith 2009). 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 (Altemose et al. 2014). 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) (Miga 2019). 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) (Altemose, Logsdon, et al. 2021). This

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 February 2022 doi:10.20944/preprints202202.0009.v1

is similar to an HSat3 array previously shown to have variable presence on chr14 in different individuals (Choo et al. 1992).

To better understand how individual arrays evolve, we examined multiple different HSat2 and HSat3 arrays across the genome using tools like StainedGlass (Vollger et al. 2022) and NTRprism (Altemose, Logsdon, et al. 2021) (Figure 3b-c). Most arrays do show higher divergence levels at the fringes, as predicted by theory for unequal crossover mechanisms (G. P. Smith 1976). 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 (Altemose, Logsdon, et al. 2021). 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 (Altemose, Logsdon, et al. 2021).

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, orangutangs, and gorillas (Gosden et al. 1977, Mitchell et al. 1977, 1981). Satellite fraction III also hybridized with New World monkey DNA (Mitchell et al. 1981), 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 (Cooke et al. 1982), consistent with a loss of most chrY q-arm heterochromatin in chimpanzees (Gläser et al. 1998). One study found that probes that are specific to the large HSat3B5 array on chr9 hybridize to multiple pericentromeric regions in gorilla metaphase spreads (Pita et al. 2010). 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 (Jarmuż et al. 2007). Some of these HSat3A4 clones were shared only between humans and chimpanzees, consistent with them having appeared less than 5 million years ago (Jarmuż et al. 2007). Other clones (corresponding to subfamilies HSat3 subfamilies A1, A2, and 3B2) appeared to be specific to humans (Jarmuż et al. 2007). 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 (Cechova et al. 2019). The prospect of telomere-to-telomere assemblies across many primate lineages will open up more comprehensive opportunities to study the evolution of the classical human satellites, resolving the incomplete picture left by prior methods.

5. Epigenomic and functional studies of HSat1-3

Although some researchers hypothesized that human satellites 2 and 3 may play a role in centromere function (Grady et al. 1992), the fact that these sequences are not present on every chromosome disfavored them as the likely centromeric sequences (Jones & Corneo 1971), and

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 February 2022 doi:10.20944/preprints202202.0009.v1

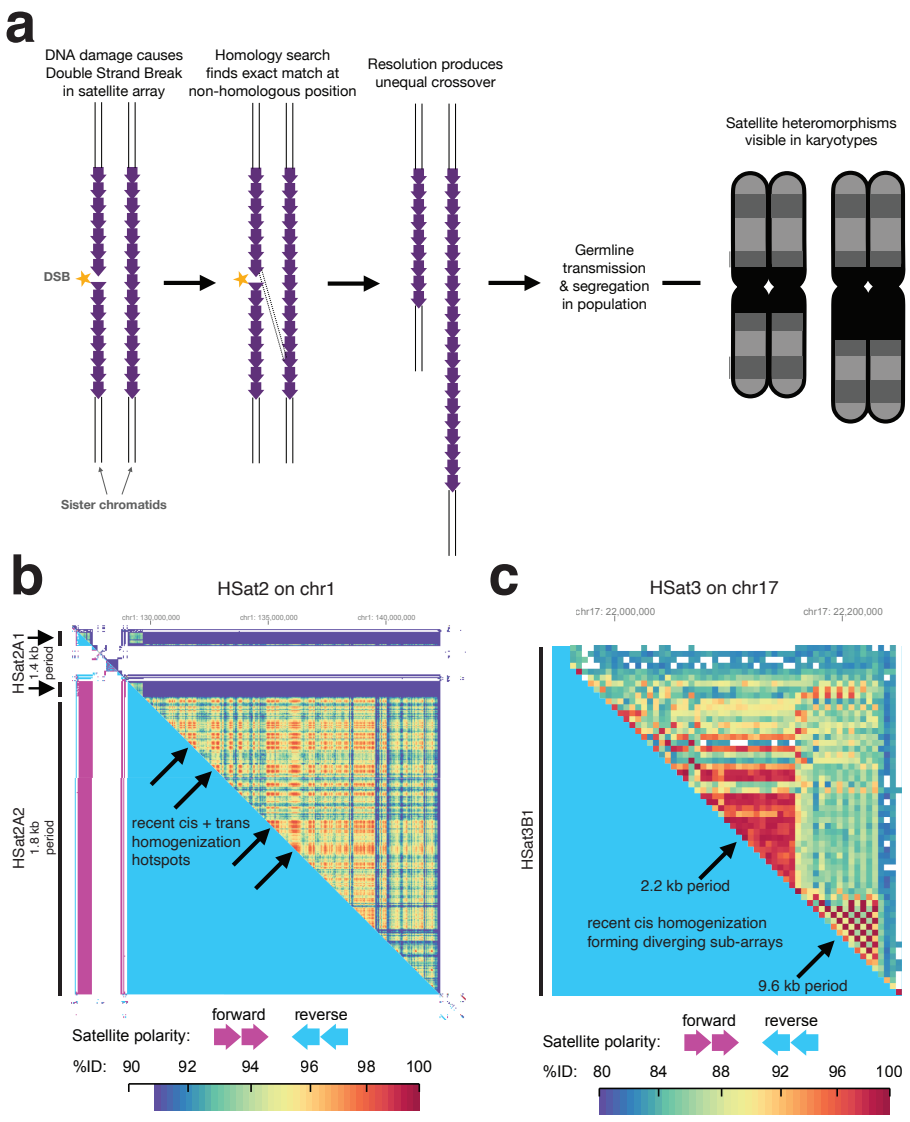

Figure 3. Mechanisms driving the evolution of HSat1-3. a) A schematic illustrating the expansion and contraction of tandem repeat arrays by an unequal non-meiotic crossover mechanism. **b**) A StainedGlass plot (Vollger et al. 2022) illustrates pairwise sequence relationships across the large HSat2 array near the centromere of chr1. Each pixel represents 40 kb. Above the diagonal, pixels are colored by the maximum percent identity of an alignment between the two corresponding sequences. Below the diagonal they are colored by their array polarity (forward means the C-rich strand is on the forward strand of the assembly). Generally, the peripheral sequences are more diverged, and multiple internal regions of recent homogenization are indicated with arrows. Regions containing HSat2A1 and HSat2A2, which have different periodicities, are indicated on the left. Note that HSat2A1 also occurs in a distinct centromere-proximal array with reversed polarity, perhaps resulting from an inversion event. **c**) As in b, but for a ~250 kb HSat3B1 array on chr17 (each pixel represents 5 kb). In this array, different subregions are homogenizing independently and taking on different periodicities.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 February 2022 doi:10.20944/preprints202202.0009.v1

later analyses more definitively established αSat as the centromeric satellite (Vafa & Sullivan 1997). 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 (Yunis & Yasmineh 1971). This hypothesis is still favored, bolstered by recent work in model organisms proposing that a key function for pericentromeric satellites is to organize chromocenters, regions where satellite-binding proteins establish inter-chromosomal links that prevent chromosomes from drifting out of the nucleus (Jagannathan et al. 2018, 2019). The classical human satellites may play a similar role in the human genome.

The high mutation rate of satellite DNA was also proposed as a driver of speciation by reproductive isolation (Yunis & Yasmineh 1971). 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 (Bosco et al. 2007, Jagannathan et al. 2017), but it can also differ dramatically between reproductively compatible *Drosophila* strains (Bosco et al. 2007). A specific non-centromeric satellite repeat was shown to cause hybrid incompatibility in crosses between *melanogaster* and *simulans* (Ferree & Barbash 2009, Sawamura et al. 1993), and later it was shown that satellite DNAs fail to cluster properly into chromocenters in incompatible hybrids (Jagannathan & Yamashita 2021). Whether satellite DNAs play a broader role in speciation remains to be seen.

Blocks of constitutive heterochromatin formed by the classical human satellites were shown to contain high levels of 5-methylcytosine by electron microscopy (Lubit et al. 1976) (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 (Gershman et al. 2021). In contrast, the CHM13 cell line, which resembles early embryonic cells, shows greatly reduced, though still periodic, methylation in these regions (Gershman et al. 2021). Adding drugs that inhibit DNA methylation results in decondensation of pericentromeric heterochromatin in human cells (Viegas-Pequignot & Dutrillaux 1976), which can lead to chromosomal rearrangements (Hernandez et al. 1997, Kokalj-Vokac et al. 1993) and chromosomal segregation errors that result in the formation of micronuclei (Haaf & Schmid 2000) (Figure 4). Senescent cells and other cells with natural hypomethylation similarly show satellite DNA decondensation (Enukashvily et al. 2007, Haaf & Schmid 2000, Swanson et al. 2013) along with segregation errors (Haaf & Schmid 2000, Slee et al. 2012) and chromosomal rearrangements (Almeida et al. 1993), suggesting a role for satellite DNA misregulation in aging (Figure 4).

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 (Yandım & Karakülah 2019b), senescent cells (Enukashvily et al. 2007), cancer cells (Bersani et al. 2015, Enukashvily et al. 2007, Hall et al. 2017, Ting et al. 2011, Yandım & Karakülah 2019a, Zhu et al. 2011, 2018), cells with DNA damage (Nogalski & Shenk 2020), virus-infected cells (Nogalski et al. 2019, Nogalski & Shenk 2020), and stressed cells (Jolly et al. 2003, Rizzi et al. 2003, Sengupta et al. 2009, Valgardsdottir et al. 2008). 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 (Jolly et al. 2002). When cells are heat shocked,

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 February 2022 doi:10.20944/preprints202202.0009.v1

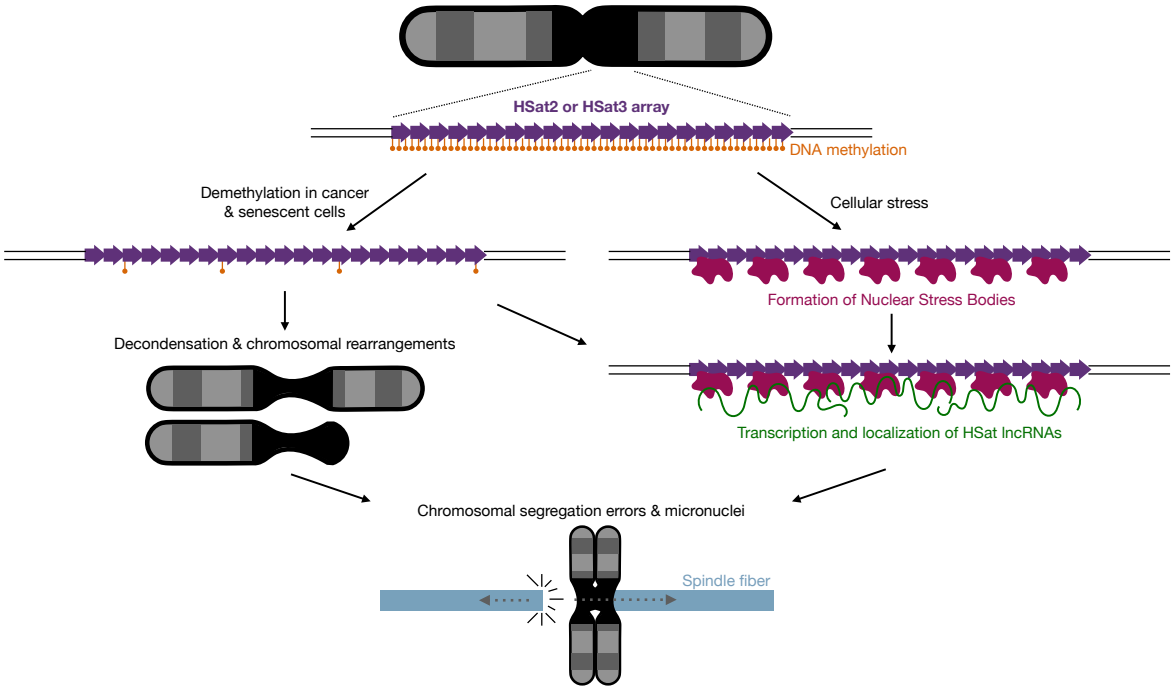

Figure 4. Roles for HSat arrays in stress responses, cancer, and senescence. At the top is a schematic of a typical HSat2 or HSat3 array in a healthy cell. The DNA is heavily methylated and the region becomes highly condensed in mitosis. In senescent cells and certain cancer cells, genome misregulation 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.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 February 2022 doi:10.20944/preprints202202.0009.v1

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 (Jolly et al. 2002). 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 (Jolly et al. 2003, Rizzi et al. 2003). Similar behavior was subsequently observed with other stress stimuli, including exposure to heavy metals, UV-C radiation, oxidative stress, and hyperosmotic stress (Sengupta et al. 2009, Valgardsdottir et al. 2008).

In cancers, nuclear stress bodies also accumulate at other HSat2 and/or HSat3 loci (Eymery et al. 2010, Hall et al. 2017), which are also sometimes demethylated and bound by polycomb bodies (Brückmann et al. 2018, Hall et al. 2017). HSat2/3 lncRNAs play a role in recruiting splicing factors and other proteins such as MeCP2 to nuclear stress bodies (Aly et al. 2019, Hall et al. 2017, Ninomiya et al. 2020, 2021), and they are associated with mitotic segregation defects (Giordano et al. 2020). A recent study examined these phenomena further by stably expressing ectopic HSat2 lncRNAs from a transgene randomly integrated into a primary human fibroblast line (Landers et al. 2021). 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 (Landers et al. 2021). 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.

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 and DNA accessibility in repetitive regions (Altemose, Maslan, et al. 2021, Stergachis et al. 2020). 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.

 doi:10.20944/preprints202202.0009.v1

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 (Tyler-Smith 2009), 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 (Frommer et al. 1982, Mitchell et al. 1979, Prosser et al. 1981). A satellite fraction IV was also isolated using a column fractionation method (Corneo et al. 1972), but later studies found that its components were indistinguishable from those found in satellite fraction III (Frommer et al. 1982, Gosden et al. 1975, Mitchell et al. 1979, Prosser et al. 1981). 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 (Altemose, Logsdon, et al. 2021).

Acknowledgements

I am grateful to Karen Miga, Hunt Willard, Gary Karpen, and Chris Tyler-Smith for their helpful discussions on these topics.

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

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 February 2022 doi:10.20944/preprints202202.0009.v1

Figure captions

Figure 1. The origins of the classical human satellites. 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).

Figure 2. Annotation of HSat1-3 in a complete T2T genome assembly. Circos plots show color-coded annotations of the pericentromeric regions of all chromosomes in the T2T-CHM13 assembly. The locations of HSat1A, HSat2, or HSat3 arrays are highlighted in an inner concentric ring. In the middle of each circle, the width of lines connecting the arrays is scaled according to the proportion of exact 75-mer sequence matches between the arrays, serving as an overall estimate of sequence similarity (no line is drawn below a fixed threshold). Overall, HSat3 appears to be the least diverged satellite, while HSat2 appears to be the most diverged, consistent with published findings (Prosser et al. 1986). [note: chrY and HSat1B will be added before final proof stage. I also plan to add a table with more detailed information about each HSat array in the genome].

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**) A StainedGlass plot (Vollger et al. 2022) illustrates pairwise sequence relationships across the large HSat2 array near the centromere of chr1. Each pixel represents 40 kb. Above the diagonal, pixels are colored by the maximum percent identity of an alignment between the two corresponding sequences. Below the diagonal they are colored by their array polarity (forward means the C-rich strand is on the forward strand of the assembly). Generally, the peripheral sequences are more diverged, and multiple internal regions of recent homogenization are indicated with arrows. Regions containing HSat2A1 and HSat2A2, which have different periodicities, are indicated on the left. Note that HSat2A1 also occurs in a distinct centromere-proximal array with reversed polarity, perhaps resulting from an inversion event. **c**) As in b, but for a ~250 kb HSat3B1 array on chr17 (each pixel represents 5 kb). In this array, different subregions are homogenizing independently and taking on different periodicities.

Figure 4. Roles for HSat 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.

References

Almeida A, Kokalj-Vokac N, Lefrancois D, Viegas-Pequignot E, Jeanpierre M, Dutrillaux B, & Malfoy B. (1993). Hypomethylation of classical satellite DNA and chromosome instability in lymphoblastoid cell lines. *Human Genetics*, 91(6), 538–546. https://doi.org/10.1007/bf00205077

Altemose N, Logsdon GA, Bzikadze AV, Sidhwani P, Langley SA, Caldas GV, Hoyt SJ, Uralsky L, Ryabov FD, Shew CJ, Sauria MEG, Borchers M, Gershman A, Mikheenko A, Shepelev VA, Dvorkina T, Kunyavskaya O, Vollger MR, Rhie A, … Miga KH. (2021). Complete genomic and epigenetic maps of human centromeres. *BioRxiv*, 2021.07.12.452052. https://doi.org/10.1101/2021.07.12.452052

Altemose N, Maslan A, Smith OK, Sundararajan K, Brown RR, Detweiler AM, Neff N, Miga KH, Straight AF, & Streets A. (2021). DiMeLo-seq: a long-read, single-molecule method for mapping protein-DNA interactions genome-wide. *BioRxiv*, 2021.07.06.451383. https://doi.org/10.1101/2021.07.06.451383

Altemose N, Miga KH, Maggioni M, & Willard HF. (2014). Genomic characterization of large heterochromatic gaps in the human genome assembly. *PLoS Computational Biology*, 10(5), e1003628. https://doi.org/10.1371/journal.pcbi.1003628

Aly MK, Ninomiya K, Adachi S, Natsume T, & Hirose T. (2019). Two distinct nuclear stress bodies containing different sets of RNA-binding proteins are formed with HSATIII architectural noncoding RNAs upon thermal stress exposure. *Biochemical and Biophysical Research Communications*, 516(2), 419–423. https://doi.org/10.1016/j.bbrc.2019.06.061

Arrighi FE, Saunders PP, Saunders GF, & Hsu TC. (1971). Distribution of repetitious DNA in human chromosomes. *Experientia*, 27(8), 964–966. https://doi.org/10.1007/bf02135776

Babcock M, Yatsenko S, Stankiewicz P, Lupski JR, & Morrow BE. (2007). AT-rich repeats associated with chromosome 22q11.2 rearrangement disorders shape human genome architecture on Yq12. *Genome Research*, 17(4), 451–460. https://doi.org/10.1101/gr.5651507

Baker WK. (1958). Crossing Over in Heterochromatin. *The American Naturalist*, 92(862), 59–60. https://doi.org/10.1086/282010

Bandyopadhyay R, McQuillan C, Page SL, Choo KHA, & Shaffer LG. (2001). Identification and characterization of satellite III subfamilies to the acrocentric chromosomes. *Chromosome Research*, 9(3), 223–233. https://doi.org/10.1023/a:1016648404388

589 **Beauchamp RS, Mitchell AR, Buckland RA, & Bostock CJ. (1979). Specific arrangements**
590 **of human satellite III DNA sequences in human chromosomes. *Chromosoma*, 71(2), 153–**
591 **166. https://doi.org/10.1007/bf00292820**

592 **Bender MA, & Gooch PC. (1961). AN UNUSUALLY LONG HUMAN Y**
593 **CHROMOSOME. *The Lancet*, 278(7200), 463–464. https://doi.org/10.1016/s0140-**
594 **6736(61)92433-3**

595 **Bersani F, Lee E, Kharchenko PV, Xu AW, Liu M, Xega K, MacKenzie OC, Brannigan**
596 **BW, Wittner BS, Jung H, Ramaswamy S, Park PJ, Maheswaran S, Ting DT, & Haber**
597 **DA. (2015). Pericentromeric satellite repeat expansions through RNA-derived DNA**
598 **intermediates in cancer. *Proceedings of the National Academy of Sciences*, 112(49),**
599 **15148–15153. https://doi.org/10.1073/pnas.1518008112**

600 **Bobrow M, Pearson PL, Pike MC, & el-Alfi OS. (1971). Length variation in the quinacrine-**
601 **binding segment of human Y chromosomes of different sizes. *Cytogenetics*, 10(3), 190–**
602 **198. https://doi.org/10.1159/000130138**

603 **Bosco G, Campbell P, Leiva-Neto JT, & Markow TA. (2007). Analysis of Drosophila**
604 **Species Genome Size and Satellite DNA Content Reveals Significant Differences Among**
605 **Strains as Well as Between Species. *Genetics*, 177(3), 1277–1290.**
606 **https://doi.org/10.1534/genetics.107.075069**

607 **Britten RJ, & Kohne DE. (1968). Repeated Sequences in DNA. *Science*, 161(3841), 529–540.**
608 **https://doi.org/10.1126/science.161.3841.529**

609 **Brückmann NH, Pedersen CB, Ditzel HJ, & Gjerstorff MF. (2018). Epigenetic**
610 **Reprogramming of Pericentromeric Satellite DNA in Premalignant and Malignant**
611 **Lesions. *Molecular Cancer Research*, 16(3), 417–427. https://doi.org/10.1158/1541-**
612 **7786.mcr-17-0477**

613 **Cechova M, Harris RS, Tomaszkiewicz M, Arbeithuber B, Chiaromonte F, & Makova KD.**
614 **(2019). High Satellite Repeat Turnover in Great Apes Studied with Short- and Long-**
615 **Read Technologies. *Molecular Biology and Evolution*, 36(11), 2415–2431.**
616 **https://doi.org/10.1093/molbev/msz156**

617 **Choo KH, Earle E, & McQuillan C. (1990). A homologous subfamily of satellite III DNA on**
618 **human chromosomes 14 and 22. *Nucleic Acids Research*, 18(19), 5641–5648.**
619 **https://doi.org/10.1093/nar/18.19.5641**

620 **Choo KH, Earle E, Vissel B, & Kalitsis P. (1992). A chromosome 14-specific human**
621 **satellite III DNA subfamily that shows variable presence on different chromosomes 14.**
622 ***American Journal of Human Genetics*, 50(4), 706–716.**

623 **Cohen S, Agmon N, Sobol O, & Segal D. (2010). Extrachromosomal circles of satellite**
624 **repeats and 5S ribosomal DNA in human cells. *Mobile DNA*, 1(1), 11–11.**
625 **https://doi.org/10.1186/1759-8753-1-11**

626 **Consortium IHGS. (2001). Initial sequencing and analysis of the human genome. *Nature*,**
627 **409(6822), 860–921. https://doi.org/10.1038/35057062**

628 **Cooke HJ. (1976). Repeated sequence specific to human males. *Nature*, 262(5565), 182–186.**
629 **https://doi.org/10.1038/262182a0**

630 **Cooke HJ, & Hindley J. (1979). Cloning of human satellite III DNA: different components**
631 **are on different chromosomes. *Nucleic Acids Research*, 6(10), 3177–3198.**
632 **https://doi.org/10.1093/nar/6.10.3177**

633 **Cooke HJ, & McKay RDG. (1978). Evolution of a human Y chromosome-specific repeated**
634 **sequence. *Cell*, 13(3), 453–460. https://doi.org/10.1016/0092-8674(78)90319-7**

635 **Cooke HJ, Schmidtke J, & Gosden JR. (1982). Characterisation of a human Y chromosome**
636 **repeated sequence and related sequences in higher primates. *Chromosoma*, 87(5), 491–**
637 **502. https://doi.org/10.1007/bf00333470**

638 **Corneo G, Ginelli E, & Polli E. (1967). A satellite DNA isolated from human tissues.**
639 ***Journal of Molecular Biology*, 23(3), 619–622. https://doi.org/10.1016/s0022-**
640 **2836(67)80130-x**

641 **Corneo G, Ginelli E, & Polli E. (1970). Repeated sequences in human DNA. *Journal of***
642 ***Molecular Biology*, 48(2), 319–327. https://doi.org/10.1016/0022-2836(70)90163-4**

643 **Corneo G, Ginelli E, & Polli E. (1971). Renaturation properties and localization in**
644 **heterochromatin of human satellite DNA's. *Biochimica et Biophysica Acta (BBA) -***
645 ***Nucleic Acids and Protein Synthesis*, 247(4), 528–534. https://doi.org/10.1016/0005-**
646 **2787(71)90689-7**

647 **Corneo G, Zardi L, & Polli E. (1972). Elution of human satellite DNAs on a methylated**
648 **albumin kieselguhr chromatographic column:Isolation of satellite DNA IV. *Biochimica***
649 ***et Biophysica Acta (BBA) - Nucleic Acids and Protein Synthesis*, 269(2), 201–204.**
650 **https://doi.org/10.1016/0005-2787(72)90427-3**

651 **Craig-Holmes A P, Moore FB, & Shaw MW. (1975). Polymporphism of human C-band**
652 **heterochromatin. II. Family studies with suggestive evidence for somatic crossing over.**
653 ***American Journal of Human Genetics*, 27(2), 178–189.**

654 **Craig-Holmes Ann P, & Shaw MW. (1971). Polymorphism of Human Constitutive**
655 **Heterochromatin. *Science*, 174(4010), 702–704.**
656 **https://doi.org/10.1126/science.174.4010.702**

Deininger PL, Jolly DJ, Rubin CM, Friedmann T, & Schmid CW. (1981). Base sequence studies of 300 nucleotide renatured repeated human DNA clones. *Journal of Molecular Biology*, 151(1), 17–33. https://doi.org/10.1016/0022-2836(81)90219-9

Earle E, Shaffer LG, Kalitsis P, McQuillan C, Dale S, & Choo KH. (1992). Identification of DNA sequences flanking the breakpoint of human t(14q21q) Robertsonian translocations. *American Journal of Human Genetics*, 50(4), 717–724.

Eichler EE, Clark RA, & She X. (2004). An assessment of the sequence gaps: Unfinished business in a finished human genome. *Nature Reviews Genetics*, 5(5), 345–354. https://doi.org/10.1038/nrg1322

Elder JF, & Turner BJ. (1995). Concerted evolution of repetitive DNA sequences in eukaryotes. *The Quarterly Review of Biology*, 70(3), 297–320. https://doi.org/10.1086/419073

Enukashvily NI, Donev R, Waisertreiger IS-R, & Podgornaya OI. (2007). Human chromosome 1 satellite 3 DNA is decondensed, demethylated and transcribed in senescent cells and in A431 epithelial carcinoma cells. *Cytogenetic and Genome Research*, 118(1), 42–54. https://doi.org/10.1159/000106440

Eymery A, Souchier C, Vourc'h C, & Jolly C. (2010). Heat shock factor 1 binds to and transcribes satellite II and III sequences at several pericentromeric regions in heat-shocked cells. *Experimental Cell Research*, 316(11), 1845–1855. https://doi.org/10.1016/j.yexcr.2010.02.002

Ferree PM, & Barbash DA. (2009). Species-Specific Heterochromatin Prevents Mitotic Chromosome Segregation to Cause Hybrid Lethality in Drosophila. *PLoS Biology*, 7(10), e1000234. https://doi.org/10.1371/journal.pbio.1000234

Fournier A, McLeer-Florin A, Lefebvre C, Duley S, Barki L, Ribeyron J, Alboukadel K, Hamaidia S, Granjon A, Gressin R, Lajmanovich A, Bonnefoix T, Chauvelier S, Debernardi A, Rousseaux S, Fraipont F de, Figeac M, Kerckaert J, Vos JD, … Callanan MB. (2010). 1q12 chromosome translocations form aberrant heterochromatic foci associated with changes in nuclear architecture and gene expression in B cell lymphoma. *EMBO Molecular Medicine*, 2(5), 159–171. https://doi.org/10.1002/emmm.201000067

Fowler C, Drinkwater R, Burgoyne L, & Skinner J. (1987). Hypervariable lengths of human DNA associated with a human satellite III sequence found in the 3, 4kb Y-specific fragment. *Nucleic Acids Research*, 15(9), 3929–3929. https://doi.org/10.1093/nar/15.9.3929

Frommer M, Prosser J, Tkachuk D, Reisner AH, & Vincent PC. (1982). Simple repeated sequences in human satellite DNA. *Nucleic Acids Research*, 10(2), 547–563. https://doi.org/10.1093/nar/10.2.547

693 **Frommer M, Prosser J, & Vincent PC. (1984). Human satellite I sequences include a male**
694 **specific 2.47 kb tandemly repeated unit containing one Alu family member per repeat.**
695 ***Nucleic Acids Research*, 12(6), 2887–2900. https://doi.org/10.1093/nar/12.6.2887**

696 **Genovese G, Handsaker RE, Li H, Altemose N, Lindgren AM, Chambert K, Pasaniuc B,**
697 **Price AL, Reich D, Morton CC, Pollak MR, Wilson JG, & McCarroll SA. (2013). Using**
698 **population admixture to help complete maps of the human genome. *Nature Publishing***
699 ***Group*, 45(4), 406–414. https://doi.org/10.1038/ng.2565**

700 **Gershman A, Sauria MEG, Hook PW, Hoyt SJ, Razaghi R, Koren S, Altemose N, Caldas**
701 **GV, Vollger MR, Logsdon GA, Rhie A, Eichler EE, Schatz MC, O'Neill RJ, Phillippy**
702 **AM, Miga KH, & Timp W. (2021). Epigenetic Patterns in a Complete Human Genome.**
703 ***BioRxiv*, 2021.05.26.443420. https://doi.org/10.1101/2021.05.26.443420**

704 **Giordano M, Infantino L, Biggiogera M, Montecucco A, & Biamonti G. (2020). Heat Shock**
705 **Affects Mitotic Segregation of Human Chromosomes Bound to Stress-Induced Satellite**
706 **III RNAs. *International Journal of Molecular Sciences*, 21(8), 2812.**
707 **https://doi.org/10.3390/ijms21082812**

708 **Gläser B, Grützner F, Willmann U, Stanyon R, Arnold N, Taylor K, Rietschel W, Zeitler S,**
709 **Toder R, & Schempp W. (1998). Simian Y Chromosomes: species-specific**
710 **rearrangements of DAZ, RBM, and TSPY versus contiguity of PAR and SRY.**
711 ***Mammalian Genome*, 9(3), 226–231. https://doi.org/10.1007/s003359900730**

712 **Gosden JR, Mitchell AR, Buckland RA, Clayton RP, & Evans HJ. (1975). The location of**
713 **four human satellite DNAs on human chromosomes. *Experimental Cell Research*, 92(1),**
714 **148–158. https://doi.org/10.1016/0014-4827(75)90648-5**

715 **Gosden JR, Mitchell AR, Seuanez HN, & Gosden CM. (1977). The distribution of**
716 **sequences complementary to human satellite DNAs I, II and IV in the chromosomes of**
717 **chimpanzee (Pan troglodytes), gorilla (Gorilla gorilla) and orang utan (Pongo**
718 **pygmaeus). *Chromosoma*, 63(3), 253–271. https://doi.org/10.1007/bf00327453**

719 **Grady DL, Ratliff RL, Robinson DL, McCanlies EC, Meyne J, & Moyzis RK. (1992).**
720 **Highly conserved repetitive DNA sequences are present at human centromeres.**
721 ***Proceedings of the National Academy of Sciences*, 89(5), 1695–1699.**
722 **https://doi.org/10.1073/pnas.89.5.1695**

723 **Gravholt CH, Friedrich U, Caprani M, & Jørgensen AL. (1992). Breakpoints in**
724 **Robertsonian translocations are localized to satellite III DNA by fluorescence in situ**
725 **hybridization. *Genomics*, 14(4), 924–930. https://doi.org/10.1016/s0888-7543(05)80113-2**

726 **Gripenberg U. (1964). Size variation and orientation of the human Y chromosome.**
727 ***Chromosoma*, 15(5), 618–629. https://doi.org/10.1007/bf00319995**

728 **Guy J, Hearn T, Crosier M, Mudge J, Viggiano L, Koczan D, Thiesen H-J, Bailey JA,**
729 **Horvath JE, Eichler EE, Earthrowl ME, Deloukas P, French L, Rogers J, Bentley D, &**
730 **Jackson MS. (2003). Genomic sequence and transcriptional profile of the boundary**
731 **between pericentromeric satellites and genes on human chromosome arm 10p. *Genome***
732 ***Research*, 13(2), 159–172. https://doi.org/10.1101/gr.644503**

733 **Haaf T, & Schmid M. (2000). Experimental condensation inhibition in constitutive and**
734 **facultative heterochromatin of mammalian chromosomes. *Cytogenetic and Genome***
735 ***Research*, 91(1–4), 113–123. https://doi.org/10.1159/000056830**

736 **Hall LL, Byron M, Carone DM, Whitfield TW, Pouliot GP, Fischer A, Jones P, &**
737 **Lawrence JB. (2017). Demethylated HSATII DNA and HSATII RNA Foci Sequester**
738 **PRC1 and MeCP2 into Cancer-Specific Nuclear Bodies. *Cell Reports*, 18(12), 2943–2956.**
739 **https://doi.org/10.1016/j.celrep.2017.02.072**

740 **Heitz E. (1928). Das Heterochromatin der Moose. *Jahrbücher Für Wissenschaftliche***
741 ***Botanik*, *69*, 762–818.**

742 **Hernandez R, Frady A, Zhang XY, Varela M, & Ehrlich M. (1997). Preferential induction**
743 **of chromosome 1 multibranched figures and whole-arm deletions in a human pro-B cell**
744 **line treated with 5-azacytidine or 5-azadeoxycytidine. *Cytogenetics and Cell Genetics*,**
745 **76(3–4), 196–201. https://doi.org/10.1159/000134548**

746 **Higgins MJ, Wang H, Shtromas I, Haliotis T, Roder JC, Holden JJA, & White BN. (1985).**
747 **Organization of a repetitive human 1.8 kb KpnI sequence localized in the**
748 **heterochromatin of chromosome 15. *Chromosoma*, 93(1), 77–86.**
749 **https://doi.org/10.1007/bf01259449**

750 **Hsu LYF, Benn PA, Tannenbaum HL, Perlis TE, Carlson AD, Opitz JM, & Reynolds JF.**
751 **(1987). Chromosomal polymorphisms of 1, 9, 16, and Y in 4 major ethnic groups: A**
752 **large prenatal study. *American Journal of Medical Genetics*, 26(1), 95–101.**
753 **https://doi.org/10.1002/ajmg.1320260116**

754 **Jackson M S, Mole SE, & Ponder BA. (1992). Characterisation of a boundary between**
755 **satellite III and alphoid sequences on human chromosome 10. *Nucleic Acids Research*,**
756 **20(18), 4781–4787.**

757 **Jackson Michael S., Slijepcevic P, & Ponder BAJ. (1993). The organisation of repetitive**
758 **sequences in the pericentromeric region of human chromosome 10. *Nucleic Acids***
759 ***Research*, 21(25), 5865–5874. https://doi.org/10.1093/nar/21.25.5865**

760 **Jagannathan M, Cummings R, & Yamashita YM. (2018). A conserved function for**
761 **pericentromeric satellite DNA. *ELife*, 7, e34122. https://doi.org/10.7554/elife.34122**

762 **Jagannathan M, Cummings R, & Yamashita YM. (2019). The modular mechanism of**
763 **chromocenter formation in Drosophila. *ELife*, 8, e43938.**
764 **https://doi.org/10.7554/elife.43938**

765 **Jagannathan M, Warsinger-Pepe N, Watase GJ, & Yamashita YM. (2017). Comparative**
766 **Analysis of Satellite DNA in the Drosophila melanogaster Species Complex. *G3***
767 ***Genes|Genomes|Genetics*, 7(2), 693–704. https://doi.org/10.1534/g3.116.035352**

768 **Jagannathan M, & Yamashita YM. (2021). Defective Satellite DNA Clustering into**
769 **Chromocenters Underlies Hybrid Incompatibility in Drosophila. *Molecular Biology and***
770 ***Evolution*, 38(11), msab221. https://doi.org/10.1093/molbev/msab221**

771 **Jarmuż M, Glotzbach CD, Bailey KA, Bandyopadhyay R, & Shaffer LG. (2007). The**
772 **Evolution of Satellite III DNA Subfamilies among Primates. *The American Journal of***
773 ***Human Genetics*, 80(3), 495–501. https://doi.org/10.1086/512132**

774 **Jeanpierre M. (1994). Human satellites 2 and 3. *Annales de Génétique*, 37(4), 163–171.**

775 **Jeanpierre M, Weil D, Gallano P, Creau-Goldberg N, & Junien C. (1985). The organization**
776 **of two related subfamilies of a human tandemly repeated DNA is chromosome specific.**
777 ***Human Genetics*, 70(4), 302–310. https://doi.org/10.1007/bf00295365**

778 **Jolly C, Konecny L, Grady DL, Kutskova YA, Cotto JJ, Morimoto RI, & Vourc'h C.**
779 **(2002). In vivo binding of active heat shock transcription factor 1 to human chromosome**
780 **9 heterochromatin during stress. *The Journal of Cell Biology*, 156(5), 775–781.**
781 **https://doi.org/10.1083/jcb.200109018**

782 **Jolly C, Metz A, Govin J, Vigneron M, Turner BM, Khochbin S, & Vourc'h C. (2003).**
783 **Stress-induced transcription of satellite III repeats. *Journal of Cell Biology*, 164(1), 25–**
784 **33. https://doi.org/10.1083/jcb.200306104**

785 **Jones KW. (1970). Chromosomal and Nuclear Location of Mouse Satellite DNA in**
786 **Individual Cells. *Nature*, 225(5236), 912–915. https://doi.org/10.1038/225912a0**

787 **Jones KW, & Corneo G. (1971). Location of Satellite and Homogeneous DNA Sequences on**
788 **Human Chromosomes. *Nature New Biology*, 233(43), 268–271.**
789 **https://doi.org/10.1038/newbio233268a0**

790 **Jones KW, Prosser J, Corneo G, & Ginelli E. (1973). The chromosomal location of human**
791 **satellite DNA III. *Chromosoma*, 42(4), 445–451. https://doi.org/10.1007/bf00399411**

792 **Jones KW, Purdom IF, Prosser J, & Corneo G. (1974). The chromosomal localisation of**
793 **human satellite DNA I. *Chromosoma*, 49(2), 161–171. https://doi.org/10.1007/bf00348888**

794 **Kalitsis P, Earle E, Vissel B, Shaffer LG, & Choo KHA. (1993). A Chromosome 13-Specific**
795 **Human Satellite I DNA Subfamily with Minor Presence on Chromosome 21: Further**

796 Studies on Robertsonian Translocations. *Genomics*, 16(1), 104–112.
797 https://doi.org/10.1006/geno.1993.1147

798 Kaplan N, & Dekker J. (2013). High-throughput genome scaffolding from in vivo DNA
799 interaction frequency. *Nature Biotechnology*, 31(12), 1143–1147.
800 https://doi.org/10.1038/nbt.2768

801 Kelly TJ, & Smith HO. (1970). A restriction enzyme from Hemophilus influenzae. *Journal*
802 *of Molecular Biology*, 51(2), 393–409. https://doi.org/10.1016/0022-2836(70)90150-6

803 Kit S. (1961). Equilibrium sedimentation in density gradients of DNA preparations from
804 animal tissues. *Journal of Molecular Biology*, 3(6), 711-IN2.
805 https://doi.org/10.1016/s0022-2836(61)80075-2

806 Kokalj-Vokac N, Almeida A, Viegas-Péquignot E, Jeanpierre M, Malfoy B, & Dutrillaux B.
807 (1993). Specific induction of uncoiling and recombination by azacytidine in classical
808 satellite-containing constitutive heterochromatin. *Cytogenetics and Cell Genetics*, 63(1),
809 11–15. https://doi.org/10.1159/000133492

810 Krüger J, & Vogel F. (1975). Population genetics of unequal crossing over. *Journal of*
811 *Molecular Evolution*, 4(3), 201–247. https://doi.org/10.1007/bf01732983

812 Kunkel LM, Smith KD, & Boyer SH. (1976). Human Y-chromosome-specific reiterated
813 DNA. *Science (New York, NY)*, 191(4232), 1189–1190.
814 https://doi.org/10.1126/science.1257744

815 Kurnit DM. (1979). Satellite DNA and heterochromatin variants: The case for unequal
816 mitotic crossing over. *Human Genetics*, 47(2), 169–186.
817 https://doi.org/10.1007/bf00273199

818 Laberge C, & Gagne R. (1971). Quinacrine mustard staining solves the length variations of
819 the human Y chromosome. *The Johns Hopkins Medical Journal*, 128(2), 79–83.

820 Landers CC, Rabeler CA, Ferrari EK, D'Alessandro LR, Kang DD, Malisa J, Bashir SM,
821 & Carone DM. (2021). Ectopic expression of pericentric HSATII RNA results in nuclear
822 RNA accumulation, MeCP2 recruitment, and cell division defects. *Chromosoma*, 130(1),
823 75–90. https://doi.org/10.1007/s00412-021-00753-0

824 Legrand A, Zimonjic DB, Popescu NC, & Nachtigal M. (1998). A Chromosome 4 Satellite I
825 DNA Isolated from SV40-Transformed Human Cells. *DNA Research*, 5(5), 271–276.
826 https://doi.org/10.1093/dnares/5.5.271

827 Levy S, Sutton G, Ng PC, Feuk L, Halpern AL, Walenz BP, Axelrod N, Huang J, Kirkness
828 EF, Denisov G, Lin Y, MacDonald JR, Pang AWC, Shago M, Stockwell TB, Tsiamouri
829 A, Bafna V, Bansal V, Kravitz SA, … Venter JC. (2007). The Diploid Genome Sequence

of an Individual Human. *PLoS Biology*, 5(10), e254. https://doi.org/10.1371/journal.pbio.0050254

Lubit BW, Duc PT, Miller OJ, & Erlanger BF. (1976). Localization of 5-methylcytosine in human metaphase chromosomes by immunoelectron microscopy. *Cell*, 9(4), 503–509. https://doi.org/10.1016/0092-8674(76)90032-5

Manuelidis L. (1976). Repeating restriction fragments of human DNA. *Nucleic Acids Research*, 3(11), 3063–3076. https://doi.org/10.1093/nar/3.11.3063

Manz E, Alkan M, Bühler E, & Schmidtke J. (1992). Arrangement of DYZ1 and DYZ2 repeats on the human Y-chromosome: a case with presence of DYZ1 and absence of DYZ2. *Molecular and Cellular Probes*, 6(3), 257–259. https://doi.org/10.1016/0890-8508(92)90025-s

Mathias N, Bayes M, & Tyler-Smith C. (1994). Highly informative compound haplotypes for the human Y chromosome. *Human Molecular Genetics*, 3(1), 115–123. https://doi.org/10.1093/hmg/3.1.115

Maxam AM, & Gilbert W. (1980). Sequencing end-labeled DNA with base-specific chemical cleavages. *Methods in Enzymology*, 65(1), 499–560. https://doi.org/10.1016/s0076-6879(80)65059-9

McKay RDG, Bobrow M, & Cooke HJ. (1978). The identification of a repeated DNA sequence involved in the karyotype polymorphism of the human Y chromosome. *Cytogenetic and Genome Research*, 21(1–2), 19–32. https://doi.org/10.1159/000130875

McKenzie WH, Hostetter TL, & Lubs HA. (1972). Y family study: Heritable variation in the length of the human Y chromosome. *American Journal of Human Genetics*, 24(6 Pt 1), 686.

Meselson M, Stahl FW, & Vinograd J. (1957). EQUILIBRIUM SEDIMENTATION OF MACROMOLECULES IN DENSITY GRADIENTS. *Proceedings of the National Academy of Sciences*, 43(7), 581–588. https://doi.org/10.1073/pnas.43.7.581

Meyne J, Goodwin EH, & Moyzis RK. (1994). Chromosome localization and orientation of the simple sequence repeat of human satellite I DNA. *Chromosoma*, 103(2), 99–103. https://doi.org/10.1007/bf00352318

Miga KH. (2019). Centromeric Satellite DNAs: Hidden Sequence Variation in the Human Population. *Genes*, 10(5), 352. https://doi.org/10.3390/genes10050352

Mitchell AR, Beauchamp RS, & Bostock CJ. (1979). A study of sequence homologies in four satellite DNAs of man. *Journal of Molecular Biology*, 135(1), 127–149. https://doi.org/10.1016/0022-2836(79)90344-9

Mitchell AR, Gosden JR, & Ryder OA. (1981). Satellite DNA relationships in man and the primates. *Nucleic Acids Research*, 9(14), 3235–3249. https://doi.org/10.1093/nar/9.14.3235

Mitchell AR, Seuanez HN, Lawrie SS, Martin DE, & Gosden JR. (1977). 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(4), 345–358. https://doi.org/10.1007/bf00288618

Moyzis RK, Albright KL, Bartholdi MF, Cram LS, Deaven LL, Hildebrand CE, Joste NE, Longmire JL, Meyne J, & Schwarzacher-Robinson T. (1987). Human chromosome-specific repetitive DNA sequences: novel markers for genetic analysis. *Chromosoma*, 95(6), 375–386. https://doi.org/10.1007/bf00333988

Nakahori Y, Mitani K, Yamada M, & Nakagome Y. (1986). A human Y-chromosome specific repeated DNA family (DYZ1) consists of a tandem array of pentanucleotides. *Nucleic Acids Research*, 14(19), 7569–7580. https://doi.org/10.1093/nar/14.19.7569

Ninomiya K, Adachi S, Natsume T, Iwakiri J, Terai G, Asai K, & Hirose T. (2020). LncRNA-dependent nuclear stress bodies promote intron retention through SR protein phosphorylation. *The EMBO Journal*, 39(3), e102729. https://doi.org/10.15252/embj.2019102729

Ninomiya K, Iwakiri J, Aly MK, Sakaguchi Y, Adachi S, Natsume T, Terai G, Asai K, Suzuki T, & Hirose T. (2021). m6A modification of HSATIII lncRNAs regulates temperature-dependent splicing. *The EMBO Journal*, 40(15), e107976. https://doi.org/10.15252/embj.2021107976

Nogalski MT, & Shenk T. (2020). HSATII RNA is induced via a noncanonical ATM-regulated DNA damage response pathway and promotes tumor cell proliferation and movement. *Proceedings of the National Academy of Sciences*, 117(50), 31891–31901. https://doi.org/10.1073/pnas.2017734117

Nogalski MT, Solovyov A, Kulkarni AS, Desai N, Oberstein A, Levine AJ, Ting DT, Shenk T, & Greenbaum BD. (2019). A tumor-specific endogenous repetitive element is induced by herpesviruses. *Nature Communications*, 10(1), 90. https://doi.org/10.1038/s41467-018-07944-x

Nurk S, Koren S, Rhie A, Rautiainen M, Bzikadze AV, Mikheenko A, Vollger MR, Altemose N, Uralsky L, Gershman A, Aganezov S, Hoyt SJ, Diekhans M, Logsdon GA, Alonge M, Antonarakis SE, Borchers M, Bouffard GG, Brooks SY, … Phillippy AM. (2021). The complete sequence of a human genome. *BioRxiv*, 2021.05.26.445798. https://doi.org/10.1101/2021.05.26.445798

Paar V, Basar I, Rosandic M, & Gluncic M. (2007). Consensus Higher Order Repeats and Frequency of String Distributions in Human Genome. *Current Genomics*, 8(2), 93–111. https://doi.org/10.2174/138920207780368169

Pardue ML, & Gall JG. (1970). Chromosomal Localization of Mouse Satellite DNA. *Science*, 168(3937), 1356–1358. https://doi.org/10.1126/science.168.3937.1356

Pech M, Igo-Kemenes T, & Zachau HG. (1979). Nucleotide sequence of a highly repetitive component of rat DNA. *Nucleic Acids Research*, 7(2), 417–432. https://doi.org/10.1093/nar/7.2.417

Petes TD. (1980). Unequal meiotic recombination within tandem arrays of yeast ribosomal DNA genes. *Cell*, 19(3), 765–774. https://doi.org/10.1016/s0092-8674(80)80052-3

Philippsen P, Streeck RE, & Zachau HG. (1974). Defined Fragments of Calf, Human, and Rat DNA Produced by Restriction Nucleases. *European Journal of Biochemistry*, 45(2), 479–488. https://doi.org/10.1111/j.1432-1033.1974.tb03573.x

Pita M, Gosálvez J, Gosálvez A, Nieddu M, López-Fernández C, & Mezzanotte R. (2010). A Highly Conserved Pericentromeric Domain in Human and Gorilla Chromosomes. *Cytogenetic and Genome Research*, 126(3), 253–258. https://doi.org/10.1159/000251962

Podugolnikova OA, & Korostelev AP. (1980). The quantitative analysis of polymorphism on human chromosomes 1,9,16, and Y. IV. Heterogeneity of a normal population. *Human Genetics*, 54(2), 163–169.

Prosser J, Frommer M, Paul C, & Vincent PC. (1986). Sequence relationships of three human satellite DNAs. *Journal of Molecular Biology*, 187(2), 145–155. https://doi.org/10.1016/0022-2836(86)90224-x

Prosser J, Reisner AH, Bradley ML, Ho K, & Vincent PC. (1981). Buoyant density and hybridization analysis of human DNA sequences, including three satellite DNAs. *Biochimica et Biophysica Acta (BBA) - Nucleic Acids and Protein Synthesis*, 656(1), 93–102. https://doi.org/10.1016/0005-2787(81)90031-9

Rizzi N, Denegri M, Chiodi I, Corioni M, Valgardsdottir R, Cobianchi F, Riva S, & Biamonti G. (2003). Transcriptional activation of a constitutive heterochromatic domain of the human genome in response to heat shock. *Molecular Biology of the Cell*, 15(2), 543–551. https://doi.org/10.1091/mbc.e03-07-0487

Sanger F, Nicklen S, & Coulson AR. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences of the United States of America*, 74(12), 5463–5467. https://doi.org/10.1073/pnas.74.12.5463

Saunders GF, Hsu TC, Getz MJ, Simes EL, & Arrighi FE. (1972). Locations of a Human Satellite DNA in Human Chromosomes. *Nature New Biology*, 236(69), 244–246. https://doi.org/10.1038/newbio236244a0

Sauter G, Moch H, Carroll P, Kerschmann R, Mihatsch MJ, & Waldman FM. (1995). Chromosome-9 loss detected by fluorescencein situ hybridization in bladder cancer. *International Journal of Cancer*, 64(2), 99–103. https://doi.org/10.1002/ijc.2910640205

Sawamura K, Yamamoto MT, & Watanabe TK. (1993). Hybrid lethal systems in the Drosophila melanogaster species complex. II. The Zygotic hybrid rescue (Zhr) gene of D. melanogaster. *Genetics*, 133(2), 307–313. https://doi.org/10.1093/genetics/133.2.307

Schmid CW, & Deininger PL. (1975). Sequence organization of the human genome. *Cell*, 6(3), 345–358. https://doi.org/10.1016/0092-8674(75)90184-1

Schmid M, Guttenbach M, Nanda I, Studer R, & Epplen JT. (1990). Organization of DYZ2 repetitive DNA on the human Y chromosome. *Genomics*, 6(2), 212–218. https://doi.org/10.1016/0888-7543(90)90559-d

Schueler MG, Higgins AW, Rudd MK, Gustashaw K, & Willard HF. (2001). Genomic and Genetic Definition of a Functional Human Centromere. *Science*, 294(5540), 109–115. https://doi.org/10.1126/science.1065042

Sengupta S, Parihar R, & Ganesh S. (2009). Satellite III non-coding RNAs show distinct and stress-specific patterns of induction. *Biochemical and Biophysical Research Communications*, 382(1), 102–107. https://doi.org/10.1016/j.bbrc.2009.02.137

She X, Horvath JE, Jiang Z, Liu G, Furey TS, Christ L, Clark R, Graves T, Gulden CL, Alkan C, Bailey JA, Sahinalp C, Rocchi M, Haussler D, Wilson RK, Miller W, Schwartz S, & Eichler EE. (2004). The structure and evolution of centromeric transition regions within the human genome. *Nature*, 430(7002), 857–864. https://doi.org/10.1038/nature02806

Shen CJ, Wiesehahn G, & Hearst JE. (1976). Cleavage patterns of Drosophila melanogaster satellite DNA by restriction enzymes. *Nucleic Acids Research*, 3(4), 931–952. https://doi.org/10.1093/nar/3.4.931

Singer MF. (1982). Highly Repeated Sequences in Mammalian Genomes. *International Review of Cytology*, 76, 67–112. https://doi.org/10.1016/s0074-7696(08)61789-1

Skaletsky H, Kuroda-Kawaguchi T, Minx PJ, Cordum HS, Hillier L, Brown LG, Repping S, Pyntikova T, Ali J, Bieri T, Chinwalla A, Delehaunty A, Delehaunty K, Du H, Fewell G, Fulton L, Fulton R, Graves T, Hou S-F, … Page DC. (2003). The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. *Nature*, 423(6942), 825–837. https://doi.org/10.1038/nature01722

967 **Slee RB, Steiner CM, Herbert B-S, Vance GH, Hickey RJ, Schwarz T, Christan S,**
968 **Radovich M, Schneider BP, Schindelhauer D, & Grimes BR. (2012). Cancer-associated**
969 **alteration of pericentromeric heterochromatin may contribute to chromosome**
970 **instability. *Oncogene*, 31(27), 3244–3253. https://doi.org/10.1038/onc.2011.502**

971 **Smith GP. (1976). Evolution of Repeated DNA Sequences by Unequal Crossover. *Science*,**
972 **191(4227), 528–535. https://doi.org/10.1126/science.1251186**

973 **Smith HO, & Welcox KW. (1970). A Restriction enzyme from Hemophilus influenzae.**
974 ***Journal of Molecular Biology*, 51(2), 379–391. https://doi.org/10.1016/0022-**
975 **2836(70)90149-x**

976 **Southern EM. (1970). Base Sequence and Evolution of Guinea-pig α-Satellite DNA. *Nature*,**
977 **227(5260), 794–798. https://doi.org/10.1038/227794a0**

978 **Southern EM. (1975a). Detection of specific sequences among DNA fragments separated by**
979 **gel electrophoresis. *Journal of Molecular Biology*, 98(3), 503–517.**
980 **https://doi.org/10.1016/s0022-2836(75)80083-0**

981 **Southern EM. (1975b). Long range periodicities in mouse satellite DNA. *Journal of***
982 ***Molecular Biology*, 94(1), 51–69. https://doi.org/10.1016/0022-2836(75)90404-0**

983 **Stergachis AB, Debo BM, Haugen E, Churchman LS, & Stamatoyannopoulos JA. (2020).**
984 **Single-molecule regulatory architectures captured by chromatin fiber sequencing.**
985 ***Science*, 368(6498), 1449–1454. https://doi.org/10.1126/science.aaz1646**

986 **Sueoka N. (1961). Variation and heterogeneity of base composition of deoxyribonucleic**
987 **acids: A compilation of old and new data. *Journal of Molecular Biology*, 3(1), 31-IN15.**
988 **https://doi.org/10.1016/s0022-2836(61)80005-3**

989 **Swanson EC, Manning B, Zhang H, & Lawrence JB. (2013). Higher-order unfolding of**
990 **satellite heterochromatin is a consistent and early event in cell senescence. *Journal of***
991 ***Cell Biology*, 203(6), 929–942. https://doi.org/10.1083/jcb.201306073**

992 **Szostak JW, & Wu R. (1980). Unequal crossing over in the ribosomal DNA of**
993 **Saccharomyces cerevisiae. *Nature*, 284(5755), 426–430. https://doi.org/10.1038/284426a0**

994 **Tagarro I, Fernández-Peralta AM, & González-Aguilera JJ. (1994). Chromosomal**
995 **localization of human satellites 2 and 3 by a FISH method using oligonucleotides as**
996 **probes. *Human Genetics*, 93(4), 383–388. https://doi.org/10.1007/bf00201662**

997 **Tagarro I, Wiegant J, Raap AK, González-Aguilera JJ, & Fernández-Peralta AM. (1994).**
998 **Assignment of human satellite 1 DNA as revealed by fluorescent in situ hybridization**
999 **with oligonucleotides. *Human Genetics*, 93(2), 125–128.**
1000 **https://doi.org/10.1007/bf00210595**

1001 **Tartof KD. (1974). Unequal mitotic sister chromatid exchange and disproportionate**
1002 **replication as mechanisms regulating ribosomal RNA gene redundancy. *Cold Spring***
1003 ***Harbor Symposia on Quantitative Biology*, 38(0), 491–500.**
1004 **https://doi.org/10.1101/sqb.1974.038.01.053**

1005 **Ting DT, Lipson D, Paul S, Brannigan BW, Akhavanfard S, Coffman EJ, Contino G,**
1006 **Deshpande V, Iafrate AJ, Letovsky S, Rivera MN, Bardeesy N, Maheswaran S, & Haber**
1007 **DA. (2011). Aberrant Overexpression of Satellite Repeats in Pancreatic and Other**
1008 **Epithelial Cancers. *Science*, 331(6017), 593–596. https://doi.org/10.1126/science.1200801**

1009 **Tyler-Smith C. (2009). The Neglected Parts of The Genome: Half a Century of Satellite**
1010 **DNA. In R. K. Mishra (Ed.), *Chromosomes to Genome*. International Pub. House.**

1011 **Unnérus V, Fellman J, & Chapelle A de la. (1967). The Length of the Human Y**
1012 **Chromosome. *Cytogenetic and Genome Research*, 6(3–4), 213–227.**
1013 **https://doi.org/10.1159/000129943**

1014 **Vafa O, & Sullivan KF. (1997). Chromatin containing CENP-A and α-satellite DNA is a**
1015 **major component of the inner kinetochore plate. *Current Biology*, 7(11), 897–900.**
1016 **https://doi.org/10.1016/s0960-9822(06)00381-2**

1017 **Valgardsdottir R, Chiodi I, Giordano M, Rossi A, Bazzini S, Ghigna C, Riva S, & Biamonti**
1018 **G. (2008). Transcription of Satellite III non-coding RNAs is a general stress response in**
1019 **human cells. *Nucleic Acids Research*, 36(2), 423–434.**
1020 **https://doi.org/10.1093/nar/gkm1056**

1021 **Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, Yandell M,**
1022 **Evans CA, Holt RA, Gocayne JD, Amanatides P, Ballew RM, Huson DH, Wortman JR,**
1023 **Zhang Q, Kodira CD, Zheng XH, Chen L, … Zhu X. (2001). The sequence of the human**
1024 **genome. *Science (New York, NY)*, 291(5507), 1304–1351.**
1025 **https://doi.org/10.1126/science.1058040**

1026 **Viegas-Pequignot E, & Dutrillaux B. (1976). Segmentation of human chromosomes induced**
1027 **by 5-ACR (5-azacytidine). *Human Genetics*, 34(3), 247–254.**
1028 **https://doi.org/10.1007/bf00295287**

1029 **Vissel B, Nagy A, & Choo KH. (1992). A satellite III sequence shared by human**
1030 **chromosomes 13, 14, and 21 that is contiguous with alpha satellite DNA. *Cytogenetics***
1031 ***and Cell Genetics*, 61(2), 81–86.**

1032 **Vollger MR, Kerpedjiev P, Phillippy AM, & Eichler EE. (2022). StainedGlass: interactive**
1033 **visualization of massive tandem repeat structures with identity heatmaps.**
1034 ***Bioinformatics*. https://doi.org/10.1093/bioinformatics/btac018**

Warburton PE, Hasson D, Guillem F, Lescale C, Jin X, & Abrusan G. (2008). Analysis of the largest tandemly repeated DNA families in the human genome. *BMC Genomics*, 9(1), 533. https://doi.org/10.1186/1471-2164-9-533

Waring M, & Britten RJ. (1966). Nucleotide Sequence Repetition: A Rapidly Reassociating Fraction of Mouse DNA. *Science*, 154(3750), 791–794. https://doi.org/10.1126/science.154.3750.791

Yandım C, & Karakülah G. (2019a). Dysregulated expression of repetitive DNA in ER+/HER2- breast cancer. *Cancer Genetics*, 239, 36–45. https://doi.org/10.1016/j.cancergen.2019.09.002

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

Yasmineh WG, & Yunis JJ. (1969). Satellite DNA in mouse autosomal heterochromatin. *Biochemical and Biophysical Research Communications*, 35(6), 779–782. https://doi.org/10.1016/0006-291x(69)90690-1

Yunis JJ, & Yasmineh WG. (1971). Heterochromatin, Satellite DNA, and Cell Function: Structural DNA of eucaryotes may support and protect genes and aid in speciation. *Science*, 174(4015), 1200–1209. https://doi.org/10.1126/science.174.4015.1200

Zhu Q, Hoong N, Aslanian A, Hara T, Benner C, Heinz S, Miga KH, Ke E, Verma S, Soroczynski J, Yates JR, Hunter T, & Verma IM. (2018). Heterochromatin-Encoded Satellite RNAs Induce Breast Cancer. *Molecular Cell*, 70(5), 842-853.e7. https://doi.org/10.1016/j.molcel.2018.04.023

Zhu Q, Pao GM, Huynh AM, Suh H, Tonnu N, Nederlof PM, Gage FH, & Verma IM. (2011). BRCA1 tumour suppression occurs via heterochromatin-mediated silencing. *Nature*, 477(7363), 179–184. https://doi.org/10.1038/nature10371