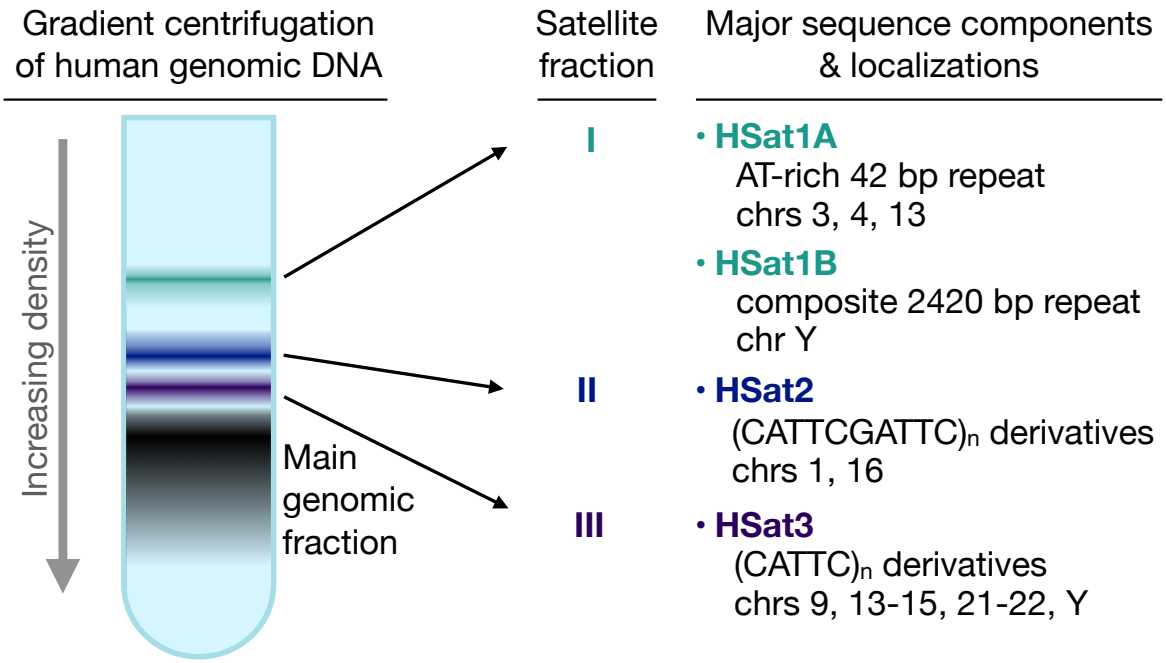

Figure 1. The origins of the classical human satellites. A schematic depicts early experiments in which high-molecular-weight human genomic DNA was fractionated by ultracentrifugation in cesium gradients. Three robust "satellite" fractions (I, II, and III) could be separated from the main genomic fraction,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

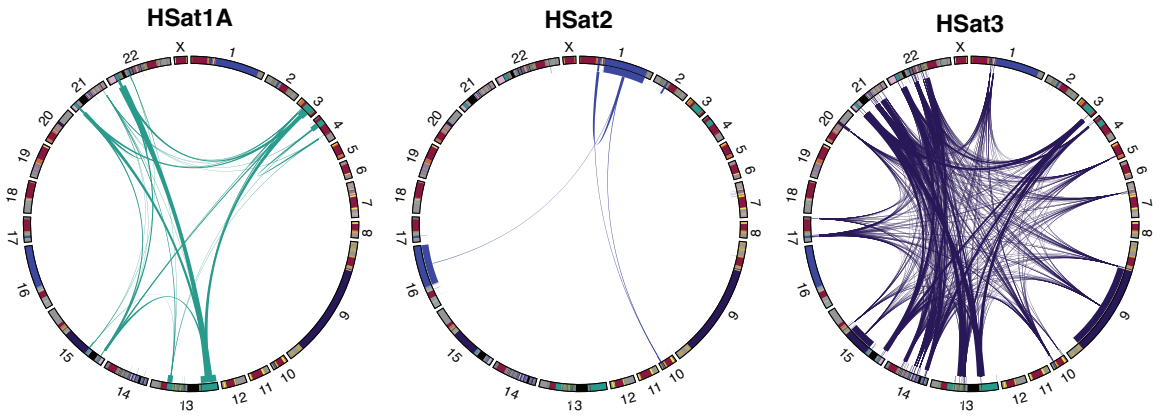

Figure 2. Annotation of HSat1-3 in a complete T2T genome assembly. Circos plots show color-coded annotations of the pericentromeric regions of all chromosomes in the T2T-CHM13 assembly. The locations of HSat1A, HSat2, or HSat3 arrays are highlighted in an inner concentric ring. In the middle of each 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].

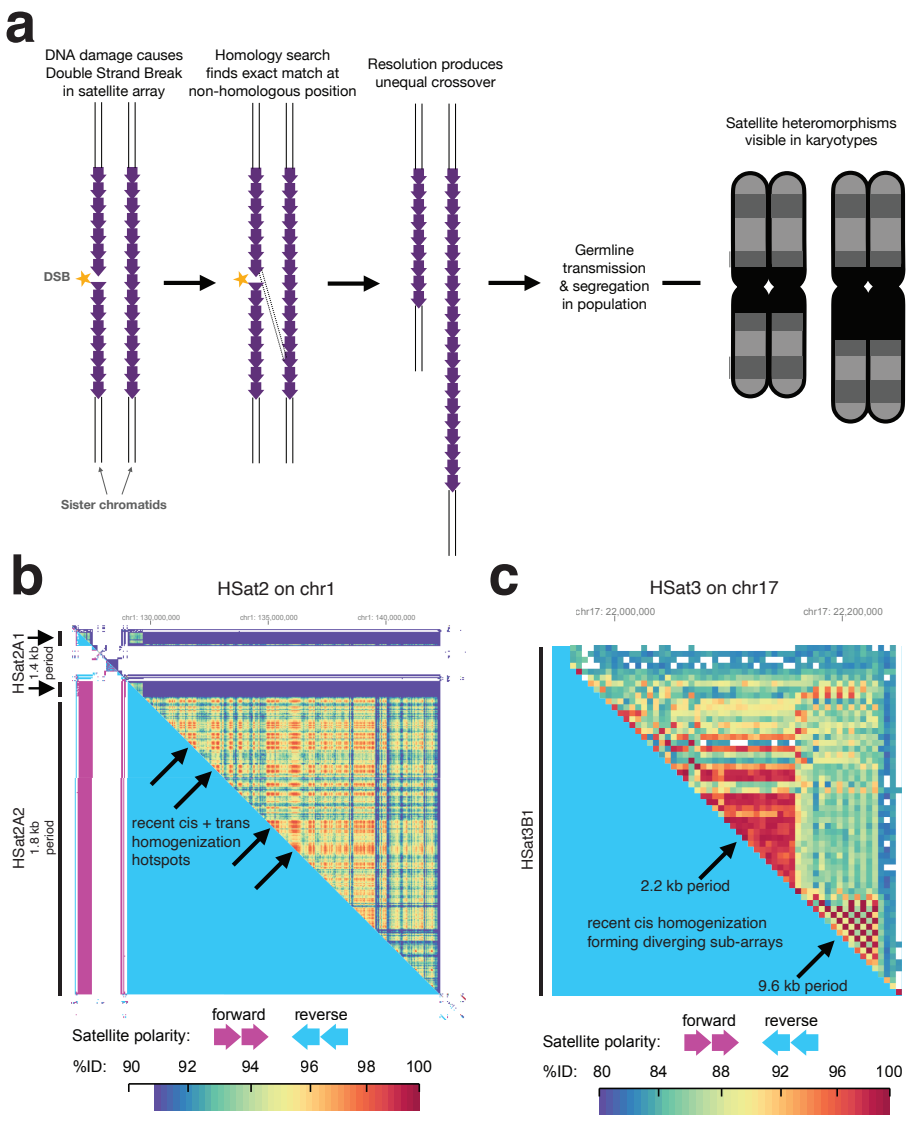

Figure 3. Mechanisms driving the evolution of HSat1-3. a) A schematic illustrating the expansion and contraction of tandem repeat arrays by an unequal non-meiotic crossover mechanism. **b**) A StainedGlass plot (Vollger et al. 2022) illustrates pairwise sequence relationships across the large HSat2 array 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.

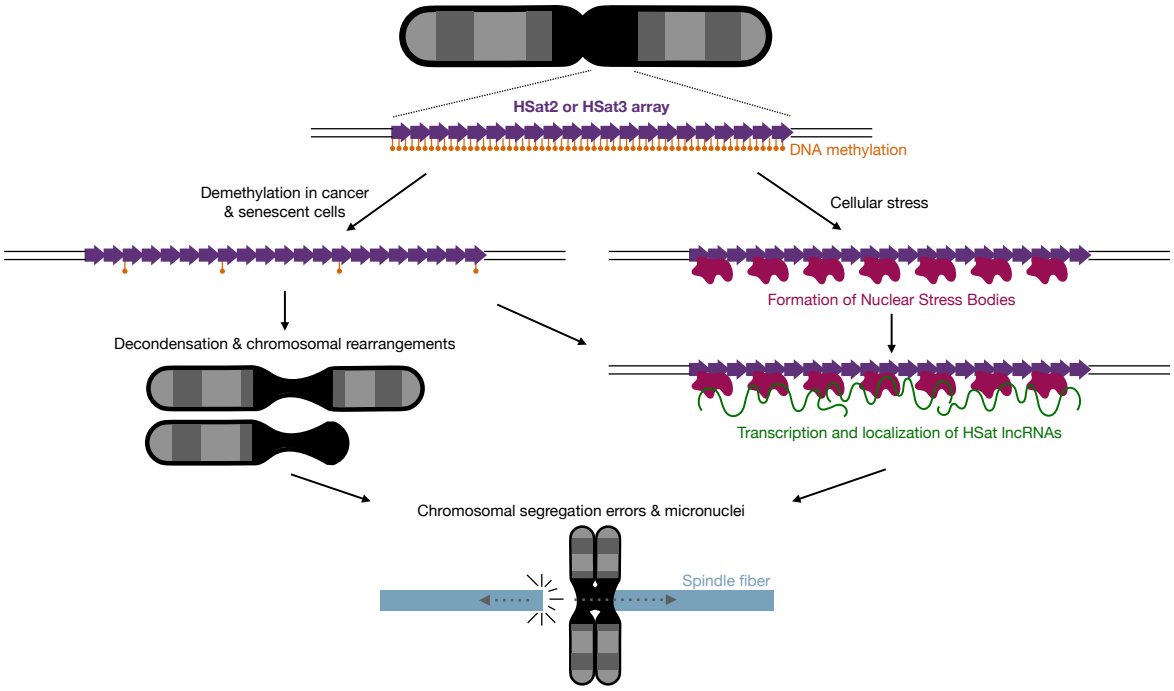

Figure 4. Roles for HSat arrays in stress responses, cancer, and senescence. At the top is a schematic of a typical HSat2 or HSat3 array in a healthy cell. The DNA is heavily methylated and the region becomes highly condensed in mitosis. In senescent cells and certain cancer cells, genome misregulation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