Short arms of human acrocentric chromosomes and the completion of the human genome sequence

Stylianos E. Antonarakis1,2,3

1Department of Genetic Medicine and Development, University of Geneva Medical Faculty, 1211 Geneva, Switzerland; 2Foundation Campus Biotech, 1202 Geneva, Switzerland; 3Medigenome, Swiss Institute of Genomic Medicine, 1207 Geneva, Switzerland

The complete, ungapped sequence of the short arms of human acrocentric chromosomes (SAACs) is still unknown almost 20 years after the near completion of the Human Genome Project. Yet these short arms of Chromosomes 13, 14, 15, 21, and 22 contain the ribosomal DNA (rDNA) genes, which are of paramount importance for human biology. The sequences of SAACs show an extensive variation in the copy number of the various repetitive elements, the full extent of which is currently unknown. In addition, the full spectrum of repeated sequences, their organization, and the low copy number functional elements are also unknown. The Telomere-to-Telomere (T2T) Project using mainly long-read sequence technology has recently completed the assembly of the genome from a hydatidiform mole, CHM13, and has thus established a baseline reference for further studies on the organization, variation, functional annotation, and impact in human disorders of all the previously unknown genomic segments, including the SAACs. The publication of the initial results of the T2T Project will update and improve the reference genome for a better understanding of the evolution and function of the human genome.

Five human chromosomes have been named as acrocentrics by the 1960 Denver nomenclature report (Lejeune et al. 1960); these are Chromosomes 13, 14, 15, 21, and 22. In these chromosomes, the centromere appeared to be at one end of the chromosomal structure, and from the Greek word ακρο (“summit or extreme tip”), they were named acrocentric. It was recognized, however, that these acrocentric chromosomes do have a short arm (p-arm according to the cytogenetics nomenclature), which peculiarly was of variable size: from almost invisible in some of the acrocentric chromosomes of certain individuals to quite a sizable chromosomal material in other individuals (for more historical notes, see Levan et al. 1964). Cytogeneticists have recognized that the short arms of the five acrocentric chromosomes contain similar sequences and that even a complete deletion of one of these short arms such as in the translocation t(21p;21p) does not have any phenotypic consequence, presumably because the remaining short arms are functionally sufficient. In this review, I refer to the short arms of acrocentric chromosomes as SAACs. In the early description of the SAACs, cytogeneticists recognized three distinct regions: the proximal short arm adjacent to the centromere, the satellite stalk, and the more distal-telomeric region, named satellite. These three regions roughly correspond to the three chromosomal bands of the SAACs p11, p12, and p13, respectively. During the 1960s and 1970s, cytogeneticists drew the acrocentric chromosomes as shown in Figure 1.

Nucleolar organizing regions and rDNA genes

The term nucleolar organizing region (NOR) was first coined by Barbara McClintock in 1934 in plant chromosomes (McClintock 1934). The NORs are chromosomal regions (DNA sequences) around which the nucleoli form; these chromosomal regions map on the SAACs in humans and contain tandem repeats (TRs) of ribosomal RNA genes (rDNA or RNR genes). It has been estimated that humans have approximately 300 copies of RNR genes in the diploid genome distributed on the SAACs, namely, Chromosomes 13, 14, 15, 21, and 22 (Henderson et al. 1972; Schmickel 1973; Stults et al. 2008; Floutsakou et al. 2013).

During metaphase, the upstream binding transcription factor (UBTF, also known as UBF), an HMG-box protein, binds to the RNR gene–containing region of each acrocentric chromosome and provides a marker for “active” NORs. These NORs are the content of the satellite stalks or chromosomal band p12. The binding of UBTF marks the secondary constriction and allows the staining with silver nitrate (AgNORs). The AgNOR staining is a method to visualize the NORs with AgNORs. The AgNOR staining is a method to visualize the NORs with AgNORs. The AgNOR staining has been used extensively in cytogenetic studies and diagnosis of chromosomal abnormalities (Fig. 2).

The unit of rDNA in the SAACs is an “operon” of ∼13 kb, which contains the gene for 45S transcribed by the RNA polymerase I (Pol I). The mapping of the 45S rDNA to the five acrocentric chromosomes was performed in the early 1970s (Henderson et al. 1972). The names of RNR1, RNR2, RNR3, RNR4, and RNR5 have been given to them for the tandem clusters on chromosomes 13p, 14p, 15p, 21p, and 22p, respectively (OMIM 180450, 180451, 180452, 180453, and 180454).

The 45S rRNA transcript is processed into 18S, 5.8S, and 28S rRNA components. The 18S and 5.8S are separated by the internal transcribed spacer 1 (ITS1) sequence; the 5.8S and 28S, by the ITS2 sequence. Each unit has a 5′ and a 3′ external transcribed spacer (ETS). The tandem units of the 45S rDNA are connected by a 32-kb intergenic spacer (IGS) sequence (Sylvestre et al. 1986; Gonzalez and Sylvester 1995). The 18S rRNA is part of the small 40S subunit of ribosomes, whereas the 28S, 5.8S, and 5S RNAs are incorporated in the large 60S subunit of the eukaryotic...
snRNA genotypes of the D group chromosomes. (a) Long arm of Chromosome 1 showing the satellite stalk. (b) Short arm of Chromosome 1 with the satellite stalk. (c) P arm of Chromosome 1 with the satellite stalk. (d) Metaphase chromosome spread with nine differentially stained satellite stalks, which appear as dark areas above the centromere of the acrocentric chromosomes. Arrow points to a D group chromosome, which lacks a satellite staining region. Figure reprinted with permission from Howell et al. (1975).
Figure 6 provides a schematic representation of inter- and intra-chromosomal exchanges in the SAACs.

A recent study using high-coverage short-read genome sequence (which apparently provides more reliable results than low-coverage sequencing) has estimated the copy number of the 45S genes of the diploid genome in different individuals. The coverage was approximately 35× for Chromosome 1, which was used as a metric. The samples tested were from 86 individuals with European ancestry and 77 with African ancestry, respectively. The copy number of 18S ranged from 115 to 550 copies with a mean of 293 copies and standard deviation of 70. Similar copy number estimates have been obtained for the 5.8S (minimum 113, maximum 562, mean 295, SD 83) and 28S (minimum 83, maximum 426, mean 226, SD 64). In this data set, there was a strong correlation of the values for these three genes ($\rho = 0.97$).

Thus, on average there are approximately 30 copies of the rDNA genes in each acrocentric chromosome; however, it remains to be seen if there are substantially unequal numbers of rDNA copies in each acrocentric chromosome. This estimation agrees with earlier estimates from the 1970s (Schmickel 1973). The mean number of each of these genes was higher in the samples of African ancestry than that of European; for example, the mean for the 18S gene was 322 and 282 in Africans and Europeans, respectively (Hall et al. 2021).

Repeated sequences in SAACs other than rDNAs

Several different arrays of TRs have been found that map in the SAACs. Most of them are called satellite repeats, and they provide a major challenge in the assembly of sequence reads. The major classes of these repeated sequences in SAACs that are included in the GRCh38 (hg38) genome assembly are (1) the alpha satellite with a repeat unit of 171 nucleotides (nt; these sequences account for ~2.58% of the human genome) (Waye and Willard 1987); (2) the beta satellite (Waye and Willard 1989) with a repeat unit of 68 nt, accounting for 0.02% of the genome; (3) the gamma satellite with a repeat unit of 220 nt, composing 0.13% of the genome; (4) the HSAT1 with a 42-nt repeat unit, accounting for 0.12% of the genome; (5) the HSAT2 and HSAT3 of a repeat unit of 5 nt, which account for 1.42% of the genome (Jones et al. 1973; Altemose et al. 2014); (6) the ACRO1 of a repeat unit of 147 nt, representing 0.01% of the genome; and (7) the CER of 96-nt unit, accounting for 0.008% of the diploid genome. (Levy et al. 2007). All of these sequences together with the addition of the rDNA gene repeats have been estimated to compose at least 3.54% of the diploid genome or equivalent to 56 megabases, a genome equivalent to a small-size chromosome such as Chromosome 19.

The alpha satellite sequences comprise most of the centromeric regions and spread over the SAACs. The technological advances of long-read sequences and the improvement of the computational methods have provided the opportunity to better understand the composition and architecture of the satellite repeats (Jain et al. 2018). An important and expected observation is the extensive variability of the structure, composition, and length (copy number) of the various satellite repeat arrays in different individuals. For example, the alpha satellites of a pair of X Chromosomes vary in length from 1 Mb to 3.5 Mb; in addition, the composition of the subfamilies of alpha satellite sequences also varies (Miga et al. 2014, 2020; Miga 2015). Figure 7 provides a schematic representation of the extensive variability of the satellite repeats.

In a recent study of more than 800 individuals, the alpha satellite had an estimate of 3.1% median value of the diploid genome, with a range between 1% and 5% (Miga et al. 2014).
Genes other than rDNAs on SAACs

A protein-coding gene named transmembrane phosphatase with tensin homology (TPTE) has been identified on Chromosome 21p. The predicted polypeptide of 551 amino acids encoded by TPTE has significant homology with tensin and auxilin domains, cyclin-G associated kinase (GAK) and the tumor suppressor PTEN. The gene contains 24 exons and spans 87 kb; the mRNA is ∼2.5 kb. On Chromosome 21 the gene maps between the D21Z1 and D21Z4 repeats, and the orientation of the transcript is from the centromere (21cen) to telomere (21pter). Monochromosomal cell hybrids showed the presence of homologous sequences on Chromosomes 13, 15, 22, and Y. The estimated number of copies in the haploid human genome was seven in males and six in females. The gene is highly expressed in testis (Chen et al. 1999; Guipponi et al. 2000). The mouse homologous gene maps to mouse Chromosome 8, which shows synteny to human Chromosome 13q14.2-q21 between NEK3 and SUGT1. The syntenic region on the human chromosome contains a partial, highly divergent copy of TPTE, now considered a pseudogene named TPTE2P2, that is likely to represent the ancestral copy from which all the other copies of TPTE arose through duplication events (Guipponi et al. 2001). Several alternative spliced isoforms have been identified. In the original description, the Chr 21 and Chr 13 transcripts were described as producing a functional transcript, whereas that of Chr 22 was transcribed under the control of an LTR and was predicted not to code for a peptide with transmembrane domains (Tapparel et al. 2003). In the current GRCh38 assembly of the UCSC Genome Browser, the Chr 21 gene sequence is listed as TPTE and that of Chr 22 as TPTEP1, and it is shown as mapping in the centromeric region. TPTE2 on Chr 13 is also shown as mapped to the 13q next to the centromere. Several pseudogenes of TPTE2, namely, TPTE2P1, TPTE2P2, TPTE2P3, TPTE2P5, and TPTE2P6, all map in the middle of Chr 13q, whereas TPTE2P4 maps on Chr Y. The update of the centromeric and short arm regions of the acrocentric chromosomes using long-read sequencing and de novo assembly will clarify the exact mapping positions and determine the potential positional variability of these genes. One of the objectives of the T2T Project is to complete the sequence and analysis of the acrocentric chromosomes in the near future.

There are several additional potential gene sequences on the SAACs. Interestingly, in the Genome Browser, these sequences are shown only on Chr 21p, probably because there are more studies on BAC sequences of Chr 21p than any other chromosome. Most of these sequences are likely to also map on other acrocentric p-arms too, and a more accurate mapping will await the results of the T2T Project (Miga et al. 2020; Logsdon et al. 2021; https://sites.google.com/ucsc.edu/t2tworkinggroup). These include members of the BAGE family of genes (Boël et al. 1995), tektin pseudogenes, long intergenic non-protein-coding RNA 1667 (LINC01667), microRNA 3156-3 (MIR3156-3), small nucleolar RNA, H/ACA box 70 (SNORA70); the list is not exhaustive. The expression of some of these genes has been tested in the GTEx project (https://gtexportal.org/home/). The TPTE gene is exclusively expressed in testis with an average level of 136 transcripts per million (TPM); the BAGE2 gene is also expressed in the testis with an average level of 2.9 TPM. Similarly, the LINC01667 gene is expressed in testis with an average level of 10.4 TPM, whereas the TEKT4P2 pseudogene is ubiquitously expressed with levels ranging from 1.5 to 5.8 TPM in different tissues.

The older sequencing efforts of 1.1 Mb of Chr 21p using BACs and short reads has identified five potential gene models in a BAC

Figure 5. Satellite associations of the short arms of human acrocentric chromosomes. (Top) Various types of satellite association from different cells. (Middle) Human nucleolar organizer chromosomes: satellite associations. (Bottom) Localization of rDNA (green) and centromeres in cells that divided in the presence of topoisomerase inhibitor ICRF-193, showing inter-chromosomal rDNA linkage. Figures reproduced with permission from Ferguson-Smith et al. (1961), Denton et al. (1976), and Potapova et al. (2019), respectively.
proximal to the rDNA repeats. These gene models GM9, GM11, GM10, GM12, and GM28 were shown transcription signals by RT-PCR in a panel of 24 tissues (Lyle et al. 2007).

Clinical significance and disorders related to sequences within the SAACs

Robertsonian translocations

Robertsonian translocations are special kinds of chromosomal translocations in which two acrocentric chromosomes are joined with breakpoints and junction sequences in their short arms. These were first described by W.R.B. Robertson in 1916 in insect speciation (Robertson 1916). Robertsonian translocations could occur among nonhomologous chromosomes (e.g., Chromosomes 14 and 21) or among homologous chromosomes (between, e.g., Chromosomes 21). Robertsonian translocations are among the most common chromosomal rearrangements in humans. A series of cytogenetic studies of more than 110,000 karyotypes of both spontaneous abortions and liveborn individuals showed a frequency of Robertsonian translocations of 0.9–1.2 per 1000 individuals (Hamerton et al. 1975; Jacobs 1981; Nielsen and Wohlert 1991). The most common is the Robertsonian translocation between Chromosomes 13 and 14, which accounts for ~75% of all Robertsonian translocations. The translocations t13;21, t14;21, t15;21, and t21;22 in the parents provide an increased risk for trisomy in the offspring. Furthermore, a t21;21 carrier could only produce children with trisomy 21. Chromosome 13 is involved in 81% of Robertsonian translocations, whereas Chromosomes 14, 15, 21, and 22 are involved in 89%, 9%, 14%, and 6% of such translocations, respectively (Therman et al. 1989). These translocations could occur in oogonial/spermatogonial mitosis, in meiotic prophase I, in the zygote, or in the early postzygotic mitotic divisions. Robertsonian translocations likely occur because of the similarity/identity of sequences in the SAACs and inappropriate recombination events between two nonhomologous chromosomes during the satellite associations. The fusion, nonhomologous Robertsonian translocation chromosome usually has two centromeres (dicentric) in ~90% of the cases (Blouin et al. 1994). In many instances of homologous Robertsonian translocations, dicentric chromosomes have also been found (Shaffer et al. 1991). Of note is that the homologous Robertsonian translocation could be either isochromosomes or the fusion of two different homologs (Blouin et al. 1994). An early study has concluded that the breakpoints in Robertsonian translocations occur preferentially in repetitive DNA, which is located between the satellite III and the rDNA (Gravholt et al. 1992). In a similar later study, most of the breakpoints in Robertsonian translocations 13;14 and 14;21 were between repetitive sequences TRI-6 (subfamily of satellite I) and rDNA on Chromosome 13p, between TRS-47 and TRS-63 (subfamilies of satellite III) on Chromosome 14, and between TRI-6 and rDNA on Chromosome 21 (Page et al. 1996).

The true nature of Robertsonian translocations, the exact recombination mechanism, and the phenotypic correlations could
be elucidated using the genomic structure from the results of the T2T Project.

Variation of rDNAs
The variation in the total number of the rDNA copies is possibly associated with phenotypic variation and/or disease states, particularly in individuals with an extremely low or high number of rDNA copies (Kampen et al. 2020). In addition, the SNVs of the resulting rDNA copies may result in heterogeneous ribosomes (Parks et al. 2018) with variation in their function. Genomic instability of the rDNA has been reported in Bloom syndrome and ataxia-telangiectasia (Parks et al. 2018), both conditions with increased cancer risk. Somatic losses of rDNA copies have been described in several tumor types (Xu et al. 2017; Udugama et al. 2018; Wang and Lemos 2019). Thus, the nucleotide and copy number variation of rDNA may become an important contributor to phenotypic variation.

Methylation of rDNA, as well as rRNA sequences
Several studies have shown that a considerable fraction of the copies of the rDNAs are methylated. The methylation level of rDNA is strongly associated with age, linking the biological age to nucleolar biology (Wang and Lemos 2019). In another study, methylation of the rDNA transcription unit including the upstream control element (UCE), core promoter, 18S rDNA, and 28S rDNA in human sperm also significantly increased with donor’s age (Potabattula et al. 2020). The rRNA is also heavily modified, and among these modifications, the most prominent is the methylation of adenosine at position 6. The biological significance of this and other modifications is under investigation in human disorders, including cancer (Barbieri and Kouzarides 2020).

Expression of satellite repeats and nearby genes
Satellite sequences could be transcribed, and the functional consequences of this transcription are inadequately studied and poorly understood. Alpha-satellite expression, for example, occurs through RNA polymerase II–dependent transcription. Single-molecule fluorescence in situ hybridization (smFISH) detects alpha-satellite RNA transcripts in intact human cells. The levels of alpha-satellite RNA vary across cell lines and over the cell cycle (Bury et al. 2020). The topology of this transcription and the spatial relationship of the nucleolus and centromeres need further study.

Ectopic insertion of satellite repeats
The insertion of 18 monomeric (\sim \sim 68\text{-bp}) beta-satellite repeat units was inserted in the TMPRSS3 gene caused one form of autosomal recessive congenital deafness (DFNB10) (Scott et al. 2001). Thus, sequences from the SAACs could “jump” and reinsert in a protein-coding gene outside of these SAACs. The mobile nature of repetitive sequences on SAACs is well documented (Farrell et al. 1993). Circular extrachromosomal molecules present in many eukaryotic cells, small polydisperse circular DNAs (spcDNA), may contain beta-satellites (Assum et al. 1993) or other repeats on the SAACs (Gaubatz 1990); these are likely to be produced by unequal homologous recombination between or within repetitive sequences. The insertion into TMPRSS3 in the DFNB10 family may have arisen by recombination of spcDNA containing beta-satellites with a region of minimal homology spanning exon 11 of TMPRSS3. The complete sequence of the SAACs will provide insights into the events of ectopic insertion of certain sequences.

The completion and analyses of the SAAC sequences through the T2T Project; promises and challenges
The use of long-read sequences of tens or hundreds of kilobases (for review, see Logsdon et al. 2020) now provides the opportunity for the first time to investigate the structure of the SAAC. Currently, the v1.0 T2T assembly that includes the completed acrocentric chromosomes has been deposited in GenBank (https://www.ncbi.nlm.nih.gov/assembly/GCA_009914755.2). The recent publication using long-read and strand-specific sequencing technologies to study SVs in a cohort of 32 genomes did not resolve the sequences of the SAACs (Ebert et al. 2021). Computational methods for de

Figure 8. Schematic representation of the two alternative hypotheses regarding some DNA sequences in the SAACs. The alternative on the left depicts the situation in which there are no chromosome-specific sequences in the short arms; the alternative on the right includes chromosome-specific sequences shown as bars of different colors in the short arms of each acrocentric chromosome.
novel assembly have been developed to capture and visualize the complexity and the variability of the sequence assemblies. The genomes of hydatidiform moles have been used to assemble the complete haplotype of DNA sequences (Steinberg et al. 2014). Such moles originate from a single sperm that has undergone postmeiotic chromosome duplication; these genomes are, therefore, uniformly homozygous for one set of alleles. The hydatidiform mole CHM13, with stable chromosomal content in culture, is used for the telomere-to-telomere sequence of Chromosomes X and 8 (Miga et al. 2020; Logsdon et al. 2021). The genomes of the moles simplify the establishment of the single and continued haplotype without the complication of the presence of the second homologous chromosome.

The promises and expectations of the T2T collaborative project regarding the SAACs could be briefly summarized below:
1. Establish the linear structure and nucleotide composition of these five SAAC regions (Chromosomes 13p, 14p, 15p, 21p, 22p).
2. Identify candidate transcribed sequences in the SAACs (both coding and noncoding) and elucidate their copy number and mapping in one or more genomic locations.
3. Identify other functional genomic elements of low-copy number and initially define their potential functional significance.
4. Identify, within the SAACs, sequences specific to each acrocentric chromosome. It is not clear if such sequences exist, and the T2T Project could provide the knowledge infrastructure for further population-based investigations. Figure 8 provides the two extreme alternative hypotheses regarding the existence of chromosome specific sequences within the SAACs.
5. Identify novel classes of repetitive elements localized primarily in the SAACs. The structure of these elements may inform potential function or may provide the reagents for further studies. In addition, a revision and update of the nomenclature of the different satellite sequences will greatly facilitate the communications in the genomic communities.
6. Provide a more accurate total length of the whole genome and each chromosome. An indication of the common variation of the chromosome length could be also provided after the T2T sequence of the genomes of different individuals from a wide variety of geoethnic ancestry.

The potential challenges include the following:
1. It is possible that there is no such thing as chromosome-specific short arm sequences, because of the extensive exchanges of sequences at the cellular and population level. The study of the variation of sequences within the SAACs will require extensive population studies using long-read methodologies that should become financially affordable, computational methods available to the scientific community, and sharing of data in appropriate databases.
2. Development of methods to establish continuous haplotypes in the context of the diploid genome are needed so that the extent of the sequence variability could be better understood, as well as the dynamics of recombinational exchanges and other mechanisms generating structural sequence variation.
3. The functional analysis of the low-copy sequences, and the high-copy repeats will be important in order to understand the potential involvement of the SAACs in human cellular and organismal biology.
4. That the involvement of SAACs in the phenotypic diversity may include Mendelian and complex traits is a medical objection that cannot be underestimated. The current knowledge of SAAC sequences in disease states is abysmally low, and it is expected that some phenotypic traits could be related or caused by pathogenic variability in some of the SAAC sequences.

Competing interest statement
The author declares no competing interests.

Acknowledgments
I thank Drs. Adam Phillippy of the National Human Genome Research Institute of the National Institutes of Health and Evan Eichler of the University of Washington, Seattle, for the expert and constructive comments on this manuscript. I also thank the ChildCare Foundation for partial financial support, as well as the three expert reviewers for the opportunity to improve the original manuscript.

References
Altemose N, Miga KH, Maggioni M, Willard HF. 2014. Genomic characterization of large heterochromatic gaps in the human genome assembly. PLoS Comput Biol 10: e1003628. doi:10.1371/journal.pcbi.1003628
Assaf F, Fink T, Steinbeisser T, Fiesel KJ. 1993. Analysis of human extrachromosomal DNA elements originating from different β-satellite subfamilies. Hum Genet 91: 489–495. doi:10.1007/BF00217778
Barbieri I, Kouzarides T. 2020. Role of RNA modifications in cancer. Nat Rev Cancer 20: 303–322. doi:10.1038/s41555-020-0253-2
Blouin JL, Binkert F, Antonarakis SE. 1994. Biparental inheritance of chromosome 21 polymorphic markers indicates that some Robertsonian translocations t(21;21) occur postzygotically. Am J Med Genet 49: 363–368. doi:10.1002/ajmg.1320490333
Boel P, Wildmann C, Sensi ML, Brasseur R, Renaudt JC, Couille P, Boon T, van der Bruggen P. 1995. BAGE: a new gene encoding an antigen recognized on human melanomas by cytotoxic T lymphocytes. Immunity 2: 167–175. doi:10.1016/S1074-7613(00)80053-0
Bury L, Moodie B, Ly J, McKay LS, Miga KH, Cheesman IM. 2020. α-Satellite RNA transcripts are repressed by centromere–nucleolus associations. eLife 9: e59770. doi:10.7554/eLife.59770
Chen H, Bossier C, Morris MA, Scott HS, Gao A, Baiocchi A, Antonarakis SE. 1999. A testis-specific gene, TPTE, encodes a putative transmembrane tyrosine phosphatase and maps to the pericentromeric region of human chromosomes 21 and 13, and to chromosomes 15, 22, and Y. Hum Genet 105: 389–404. doi:10.1007/s004390051122
Denton TE, Howell WM, Barrett JV. 1976. Human nucleolar organizer chromosomes: satellite associations. Chromosoma 55: 81–84. doi:10.1007/BF00288330
Ebert P, Audano PA, Zhu Q, Rodriguez-Martín B, Ponsby D, Bonder MJ, Sulovari A, Eberl J, Zhou W, Serra Mari R, et al. 2021. Haplotype-diversified human genomes and integrated analysis of structural variation. Science 372: eaab7117. doi:10.1126/science.aba7117
Farrell SA, Winsor EJ, Markovic VD. 1993. Moving satellites and unstable chromosome translocations: clinical and cytogenetic implications. Am J Med Genet 46: 715–720. doi:10.1002/ajmg.1320460624
Ferguson-Smith MA, Handmaker SD, Hopkins ABJ. 1961. Observations on the satellite human chromosomes. Lancet 277: 638–640. doi:10.1016/S0140-6736(61)91655-5
Floutsakou I, Agrawal S, Nguyen TT, Seoighe C, Ganley AR, McStay B. 2013. Shared genomic architecture of human nucleolar organizer regions. Nat Rev Genet 14: 710–713. doi:10.1038/nrg34914.113
Friedrich U, Caprani M, Niebuhr E, Therkelsen AJ, Jørgensen AL. 1996. Extreme variant of the short arm of chromosome 15. Mutat Res 359: 101–110. doi:10.1016/0165-173X(95)81850-8
Fujisawa Y, Naito S, Nishizaki K, Honda M, Ohshima M, Kato T, et al. 2014. A testis-specific gene, TPTE, encodes a putative transmembrane tyrosine phosphatase and maps to the pericentromeric region of human chromosomes 21 and 13, and to chromosomes 15, 22, and Y. Hum Genet 105: 389–404. doi:10.1007/s004390051122
Gaubatz JW. 1990. Extrachromosomal circular DNAs and genomic sequence plasticity in eukaryotic cells. Mutat Res 237: 271–292. doi:10.1016/0270-5537(89)80098-5
Gimelli G, Porro E, Santi F, Scappaticci S, Zuffardi O. 1996. “Jumping” satellites in three generations: a warning for paternity tests and prenatal diagnosis. Hum Genet 98: 315–318. doi:10.1007/BF00295297
Gonzalez IL, Sylvester JE. 1995. Complete sequence of the 43-kb human β-satellite DNA repeat: analysis of the intergenic spacer. Genomics 27: 320–328. doi:10.1006/geno.1995.1049
instability in ATRX-mutated cancers. Proc Natl Acad Sci 115: 4737–4742. doi:10.1073/pnas.1720391115

van Sluis M, Gailin MO, McCarter JGW, Mangan H, Grob A, McStay B. 2019. Human NORs, comprising rDNA arrays and functionally conserved distal elements, are located within dynamic chromosomal regions. Genes Dev 33: 1688–1701. doi:10.1101/gad.331892.119

Wang M, Lemos B. 2019. Ribosomal DNA harbors an evolutionarily conserved clock of biological aging. Genome Res 29: 325–333. doi:10.1101/gr.241745.118

Warner JR. 1999. The economics of ribosome biosynthesis in yeast. Trends Biochem Sci 24: 437–440. doi:10.1016/S0968-0004(99)01460-7

Waye JS, Willard HF. 1987. Nucleotide sequence heterogeneity of α satellite repetitive DNA: a survey of alphoid sequences from different human chromosomes. Nucleic Acids Res 15: 7549–7569. doi:10.1093/nar/15.18.7549

Waye JS, Willard HF. 1989. Human β satellite DNA: genomic organization and sequence definition of a class of highly repetitive tandem DNA. Proc Natl Acad Sci 86: 6250–6254. doi:10.1073/pnas.86.16.6250

Xu B, Li H, Perry JM, Singh VP, Unruh J, Yu Z, Zakari M, McDowell W, Li L, Gerton JL. 2017. Ribosomal DNA copy number loss and sequence variation in cancer. PLoS Genet 13: e1006771. doi:10.1371/journal.pgen.1006771