Satellite DNA and related diseases

J. Rich, V. V. Ogryzko, I. V. Pirozhkova

CNRS, University Paris 11; UMR8126: Nuclei & Innovations in Oncology, Gustave-Roussy
114, rue Edouard Vaillant, Villejuif Cedex, 94805

iryna.pirozhkova@gustaveroussy.fr

Satellite DNA, also known as tandemly repeated DNA, consists of clusters of repeated sequences and represents a diverse class of highly repetitive elements. Satellite DNA can be divided into several classes according to the size of an individual repeat: microsatellites, minisatellites, midisatellites, and macrosatellites. Originally considered as «junk» DNA, satellite DNA has more recently been reconsidered as having various functions. Moreover, due to the repetitive nature of the composing elements, their presence in the genome is associated with high frequency mutations, epigenetic changes and modifications in gene expression patterns, with a potential to lead to human disease. Therefore, the satellite DNA study will be beneficial for developing a treatment of satellite-related diseases, such as FSHD, neurological, developmental disorders and cancers.

Keywords: satellite DNA, repeated sequences, frequency mutations, satellite-related diseases.

The non-coding DNA – is it «junk» or functional? Only a tiny percentage of human DNA is coding for proteins, whereas the non-coding DNA, transposons and transposon-derived elements make up the majority of the genome [1]. This fact was first discovered already in the 70's of the last century. Later, the efforts of the Human Genome Project resulted in the estimation of the percentage of non-coding DNA of the human genome as 98–99 %.

This somewhat surprising fact led Susumi Ohno to suggest the notion of «junk DNA», referring to the idea that most genomic DNA has no use for the organism [2]. According to Richard Dawkins, one of the leaders of the pro-«Junk» concept, «the greater part of the genome might as well not be there, for all the difference it makes» [3]. This claim found support in recent work showing that the megabase sized deletion of non-coding sequences in mice has no phenotypic effect [4]. There were, indeed, good reasons (based on the mathematical population genetics), to expect that most of the sequences in a typical eukaryotic genome (except only about 30000 loci, according to Ohno) could not be under se-

lection, and thus could not be important. The concept of «junk DNA» was also supported by the so-called C-value enigma, according to which the genome size does not correlate with the complexity level of the organism [5] – *e. g.*, humans have genomes much smaller than those of some single cell eukaryotes.

However, with the completion of the Human Genome Project and the launching of the ENCODE (The Encyclopedia of DNA Elements) project by the US National Human Genome Research Institute (NHGRI), the prevailing view that portrays most of the human genome as having no function whatsoever, has become less and less sustainable.

The ENCODE project has systematically mapped regions of transcription, transcription factor association, chromatin structure and histone modification. These data allowed researchers to assign the biochemical functions for 80 % of the human genome, in particular to non-coding regions, highlighting an evidence of the functionality of non-protein-coding DNA. One of the demonstrations is the transcription of most non-protein-coding DNA into RNA [6]. Only 4 % of the 65000 RNAs produced by genome are coming from exons [7]. It would not make sense for the organism to waste pre-

cious resources transcribing such DNA, if it were not functional. Moreover, hundreds of non-coding regions of DNA have been recently found to be 100 % conserved between humans and mice (called ultra-conserved elements, UCEs) [8]. The perfect conservation of these regions makes an argument for their evolutionary importance.

Much further evidence of functionality of non-protein-coding DNA has been shown. For example, the introns regulating an alternative splicing [9–12] are involved in gene expression [5, 13, 14] and chromatin organization [15, 16], some pseudogenes are involved in lipid metabolism [17], in the process of protein synthesis [18], and in the RNA interference [19, 20].

The changes in our views on the functional importance of non-coding DNA also concern the «poster child» of the junk DNA concept – the so called satellite DNA, the main subject of this review.

Satellite DNA in genome organization. In humans, recent estimations based on the results of the Human Genome project suggest that about 70 % of the genome is represented by repetitive and repeat-derived DNA elements [21].

The repeated sequences can be divided into two categories (Fig. 1):

– The «interspersed repeated sequences», when the repeating copies are dispersed over the genome.

– The «tandem repeated sequences» (also called satellite DNA), when the repeating copies are adjacent to each other.

Satellite DNA is a diverse class of highly repetitive units, accounting for approximately 10–15 % of all repetitive DNA sequences in the human genome [22]. The size of one tandem repeat unit can vary dramatically, *i. e.*, from one base pair (bp) to several hundred bps. According to the orientation of the repeat units in tandem, two types of repeats can be distinguished: i) direct repeats, which are head-to-tail orientated and ii) inverted repeats, which are head-to-head orientated. However, only direct repeats are frequently found in genomes, whereas the inverted repeats are rare, presumably because they can induce double strand breaks [23].

First thought as an artefact, satellite DNA was discovered in 1961, as additional peaks appeared during fractionation of DNA according to floating density in gradient of caesium salts [24, 25]. Ten years later, the enrichment of highly repetitive satellite DNA in the constitutive heterochromatin regions was described [26]. Satellite DNA can be located in different parts of chromosomes with the preference in telomeric and centromeric regions. The examples of satellite DNA are the telomeres, the centromeres and the ribosomal genes. Despite numerous studies, the functional significance of satellite DNA is still poorly understood.

Classification of satellite DNA. Satellite DNA is divided into several categories according to size, structure and localization. However, a universally accepted classification does not exist so far, and the scheme can vary among authors [23, 27, 28]. According to the size of an individual repeat, satellite DNA can be divided in-
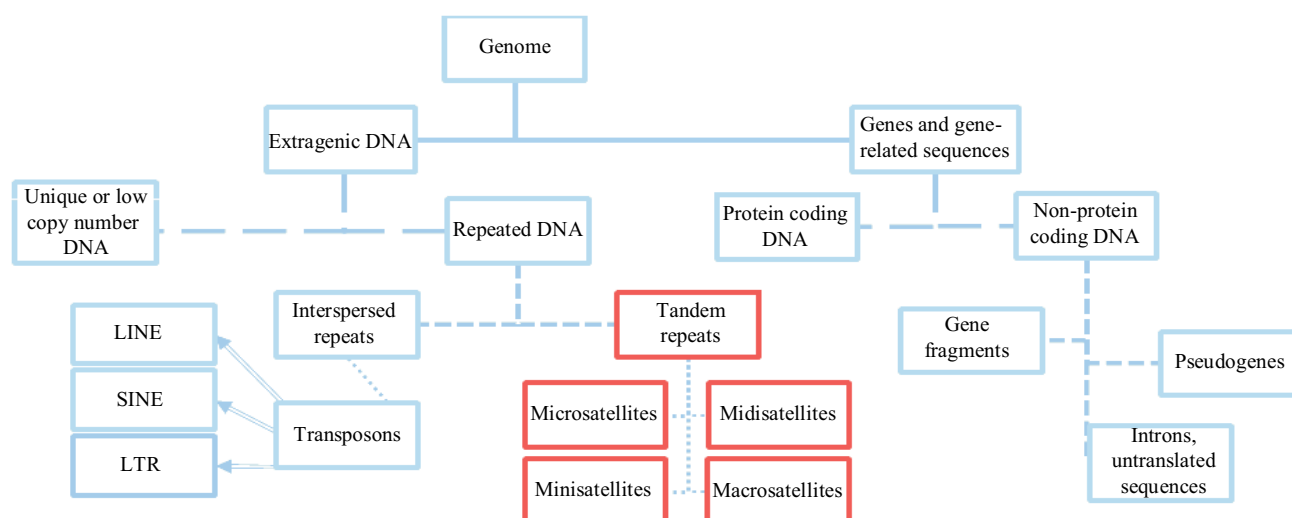

Fig. 1. Organization of eukaryotic genome

to four classes: (1) microsatellites (less than 10 bp per repeat), (2) minisatellites (10–60 bp per repeat), (3) satellites (up to hundreds of bp per repeat) and (4) macrosatellites (several kb per repeat). To avoid potential confusion, we propose naming the third class of satellite repeats as midisatellites, leaving the term of «satellite DNA» to encompass all tandemly repeated DNA.

The most abundant human satellites are the simplest repeats (microsatellites), which are also called simple sequence repeats (SSR), or short tandem repeats (STR) in the literature. These repeats are widely spread in plant and animal genomes; for example, there is, on average, one microsatellite every 2 kb in the human genome [29]. The microsatellites are frequently involved in neurodegenerative diseases caused by the strong expansion of SSR in one gene [30]. One of their main characteristics is a high rate of instability manifested in a loss or gain of repeat units, which is explained by i) the non-homologous recombination between homologous chromosomes and ii) the replication errors [31]. The microsatellites have been shown to emerge as ubiquitous genetic markers for many eukaryotic genomes [32–34]. This could be useful in the evolutionary studies regarding the genetic relationships.

The minisatellites exhibit some similarities with the sequence of lambda-phage (GCTGTGG) and the majority of them are GC-rich with a strong strand asymmetry. Among others, this class of satellite tandem DNA contains the hypermutable repeats, which are flanked by DSB (double-strand break) hot spots and show a high rate of meiotic instability. Because of their length polymorphism, the minisatellites are used for DNA fingerprinting in forensic science for the identification of individuals by their respective DNA profiles. In addition, they have been proposed to serve as markers for genotoxicity [35].

Very little is known about the midisatellite non-coding DNA, which is mostly located in the telomeric or centromeric regions of chromosomes. This kind of satellite DNA is represented by α, β and γ-satellite repeats. Alpha-satellite repeats (ASR), composed of a tandem array of a 171 bp repeat unit, are located at the centromeres of all human chromosomes. ASR play a crucial role in the chromosome segregation of normal and human artificial chromosomes (HACs) [36]. The beta-satellite DNA was described by Waye and Willard, and

is presented by tandem arrays of divergent Sau3A 67–69 pb monomer repeat units [37]. Beta satellite repeats (BSR) show a predominant heterochromatic distribution, which includes the short arm of acrocentric chromosomes and the pericentomeric part of chromosomes 1, 3 and 9 [38, 39]. In these regions, beta satellite arrays are adjacent to LSau 3.3 kb macrosatellites (D4Z4-like repeats). The D4Z4 macrosatellite repeats and BSR are also located in the subtelomeric regions of chromosomes 4 and 10 [40, 41]. Recently, the presence of BSR next to a newly created telomere has been shown to retard its replication timing [42]. The DNA of gamma-satellite repeats (GSR), a tandem array of 220-bp GC-rich repeating units, was identified in the pericentromeric regions of human chromosomes 8, X, and Y. GSR usually form 10–200 kb clusters, flanked by alpha-satellite DNA. It has been proposed that the primary role of the gamma-satellite DNA is to prevent the pericentromeric genes from epigenetic silencing [43].

The macrosatellite repeat (MSR) DNA is the only satellite DNA that could contain an open reading frame (ORF) and thus, could produce protein-coding RNA in every repeat unit, as illustrated by the D4Z4 (DUX4), CT47, RS447, TAF11-Like and PRR20 repeats. In comparison to the telomeric or centromeric satellites, which could be present on many chromosomes, MSR are often specific for only one or two chromosomes. Additionally, MSR are mostly expressed in the germ lines, with the expression being affected by the methylation status of DNA. Moreover, MSR could be associated with both transcriptionally active and silenced chromatin [28, 44]. Other non-coding MSR, such as DXZ4, have been shown to play a mechanical role, and are involved in the formation of MSR chromatin topology [45]. The role of MSR is not well known. Therefore, D4Z4 is probably one of the most studied macrosatellites because of its association with facioscapulohumeral muscular dystrophy (FSHD).

Clinical relevance of satellite DNA. In recent years, much attention has been brought to the role of repeat sequences in various pathologies, such as epilepsy, embryonic lethality and cancers [46, 47]. The increase in the number of copies of repeats in genomic DNA is the single most important cause of nearly 30 hereditary disorders [48]: the X-fragile syndrome [49], the myotonic dystrophy [50], the Huntington's disease [51], the

Friedreich's ataxia [52], *etc*. Scott *et al*. have described the first evidence of the insertion of 18 beta-satellite units in to the gene coding transmembraine serine protease resulting in autosomal recessive deafness [53].

Most of these diseases are due to the microsatellites, whose structural features often result in the disruption of DNA replication, repair and recombination processes, leading to expanded or contracted DNA structures. The microsatellite expansion has been implicated in serious myopathies, as well as in neuromuscular and neurodegenerative hereditary diseases. The majority of the disorders are caused by the expansion of the triplet repeats (CGG)*(CCG), (CAG)*(CTG), (GAA)*(TTC) and (GCN)*(NGC). Nevertheless, diseases can also result from the expansion of a tetranucleotide, a pentanucleotide and even a dodecanucleotide repeat [48].

The microsatellite expansion diseases can result in a gain or/and loss of function. Among disorders caused by gain-of-function mechanism, the main cause is the protein conformation alteration, leading to changes in protein activity or abundance. As an example, the polyglutamine diseases, one of the nine classes of gain-of-function disorders, are due to the expansion of CAG repeat. These include Huntington disease (HD), Kennedy disease and Spinocerebellar ataxia (SCA) [54, 55]. Affected individuals develop nuclear inclusion bodies containing aggregated proteins with expanded polyglutamine stretches.

Alternatively, the repression of the gene transcription, caused by the microsatellite expansion, leads to the loss-of-function mechanism. For example, fragile X mental retardation 1 gene (*FMR1*) is associated with (CGG)$_n$ satellite expansion, resulting in the transcription silencing and the loss of *FMRP*, the protein product of *FMR1* gene [56]. Other examples of loss of protein expression are Jacobsen syndrome [57] and Friedreich's ataxia [58]. The contributions of the microsatellite expansion could be manifested by both gain and loss of function mechanisms as it has been demonstrated for spinocerebellar ataxia type 1 (SCA1) [59].

Intriguingly, for the microsatellite-expansion related pathologies, a correlation has been reported between the size of the repeat tandem and the severity of the disease [60]. This correlation is behind the phenomenon of genetic anticipation, according to which the progressive increase in the repeat number (to be inherited in subsequent generations) results in the increased severity and earlier manifestation of disease [61]. Importantly, the expandable microsatellite repeats responsible for a particular disease are usually located within the affected gene, in either the coding (ORF) or else the noncoding (promoters, introns and UTRs) regions of it [47].

In the cases other than microsatellites, the exact molecular mechanisms of how satellite DNA can be involved in pathology remain largely unexplored, particularly when satellite DNA is not located in the coding region of the affected gene. Among the unresolved issues are: (i) the relative role of satellite DNA presence and (ii) the contribution of genetic and epigenetic factors, *i. e.*, whether structural modifications (tandem size, nucleotide sequence etc.) of satellite DNA provide the starting point, or the epigenetic modifications play the primary role in the mechanism of pathology. The slow progress in this matter could be explained by the methodological difficulties in studying large repeating DNA sequences, as well as by the lack of animal models that could replicate the symptoms of the hereditary satellite-related diseases in humans.

Cancer as a satellite-related disease. «Breaking of satellites' silence» is now a new paradigm in cancerogenesis [62]. While gene-specific loci can be either hypo- or hypermethylated in cancer, the highly repeated DNA sequences are only hypomethylated in this disease. Moreover, the global DNA hypomethylation observed so frequently in cancers is mostly due to satellite DNA. Ehrlich *et al*. were the first to demonstrate the hypomethylation of tandem minisatellite centromeric DNA (namely centromere-adjacent satellite 2, (Sat2)) in breast adenocarcinomas [63], ovarian epithelial tumors [64] and Wilms tumors [65]. A highly significant difference in the methylation level was found in satellite repeats (SATR1 and ARLa) in neurosarcoma [66]. In 76 % of hepatocellular carcinoma cases, the reduction in Sat2 methylation has also been demonstrated [67]. More recently, Zhu *et al*. showed that the loss of the BRCA1 tumour suppressor gene provoked satellite-DNA derepression in breast and ovarian tumours of mice and humans [68]. Satellite DNA hypomethylation has been postulated as the mechanism that underlies the induction of peri-centromeric instability in many human cancers [69]. All together, these results suggest satellite repeats to be the main target for hypomethylation.

Recently, a 40 fold-increase of the ASR tandem sequence expression has been observed for cancers of lung, kidney, ovary, colon, and prostate. More impressively, a 131-fold increase of the pericentromeric satellite Sat2 transcripts has been demonstrated in pancreatic tumors in comparison with normal pancreases [70]. Interestingly, other satellites, such as BSR, GSATII and TARI were not affected.

However, some tandem repeats, NBL2 and D4Z4, could be hypomethylated in some cancers and hypermethylated in others, which could be explained by *de novo* methylation processes [71, 72].

There is no data in the literature concerning the potential mechanisms for the selective demethylation of satellite DNA in cancer. Nevertheless, this phenomenon and the overexpression of satellite transcripts in this disease could potentially be useful as a biomarker for cancer detection and for the evaluation of the efficacy of anti-cancer treatment. In addition, epigenetic cancer therapy directed against global methylation changes needs to be reconsidered towards the targeting of the only affected DNA, in order to avoid side effects. The knowledge of what satellite DNA is implicated in particular cancer will be important in cancer study.

One can expect that the search for links between satellite DNA, tumor suppressors and genomic instability promises new routes for cancer therapy. However, the molecular connections between the satellite hypomethylation and the cancer development remain obscure. The opening of chromatin at satellite repeat regions resulting in the mislocalization of the transcription machinery, could lead to the aberrations in gene regulation with pathological consequences [73]. Additional mechanism explaining the changes in transcriptional regulation may be achieved through the remodelling and looping of chromatin [74, 75], as described in our discussion of FSHD below, or through the involvement of long-noncoding RNA (lncRNA) [76]. It has been shown that repeating sequences can massively produce lncRNA, which can further contribute to the epigenetic changes leading to pathology [70, 77, 78]. In every type of cancer analyzed so far, at least 200 lncRNA have been found to be affected [79]. Among those lncRNA, some important validated candidates for prognostic indicators and/or functionally relevant universal «drivers» and «suppressors» of drug-resistant metastasis have been identified [80]. One can expect that these figures will increase in the near future, thanks to the efforts of the ENCODE project in decrypting 98 % of non-protein-coding DNA of the human genome, including satellite DNA, which aim is to characterize the non-coding transcriptional landscape and to analyze the possible expression of extremely short peptides encoded by lncRNA [81].

The role of satellite DNA in cancer illustrates again that, whereas it was first introduced as «junk» DNA, in fact it plays an important role in genome functioning. Moreover, there are suggestions in the literature that the purifying selection, responsible for the conservation of the DNA satellite sequences in evolution, is due to the oncological consequences of the mutations in satellite DNA [82].

Facioscapulohumeral muscular dystrophy as a satellite-related disease. Another example of clinical relevance of satellite DNA is the FSHD disease, which is associated with the macrosatellite (D4Z4) and beta-satellite (4qA allele) DNA sequences. It is an autosomal dominant muscular dystrophy ranking second after Duchenne muscular dystrophy, with an incidence of 1: 14,000 throughout the world (2010, The FSH Society). FSHD is clinically characterized by a progressive muscular weakness in an up-to-down manner involving face, pectoral girdle, upper limbs, lower limbs and hips [83, 84].

The FSHD locus was mapped in the 1990s by linkage analysis in the subtelomeric region of the long arm of chromosome 4 (4q35) [85–87]. The causal molecular anomaly is the contraction of the macrosatellite tandem size consisting of D4Z4 repeats. In the normal population, the number of copies is polymorphic and varies from 11 to 150. In 90–95 % of the FSHD patients, this number is decreased to 1–10 units [88]. The lower the repeat number, the younger the age of the beginning of the disease and the higher the severity of the disease [89]. The presence of at least one D4Z4 repeat is necessary to develop the disease, as patients with 4qter deletion have no FSHD symptoms [90].

D4Z4 is a macrosatellite repeat of 3303 bp length (Fig. 2). Each unit contains LSau and hhspm3 interspersed repeats and an open reading frame of 1173 bp, named DUX4, which contains two homeobox domains [91, 92]. LSau is a Long Sau3A-repeated element of
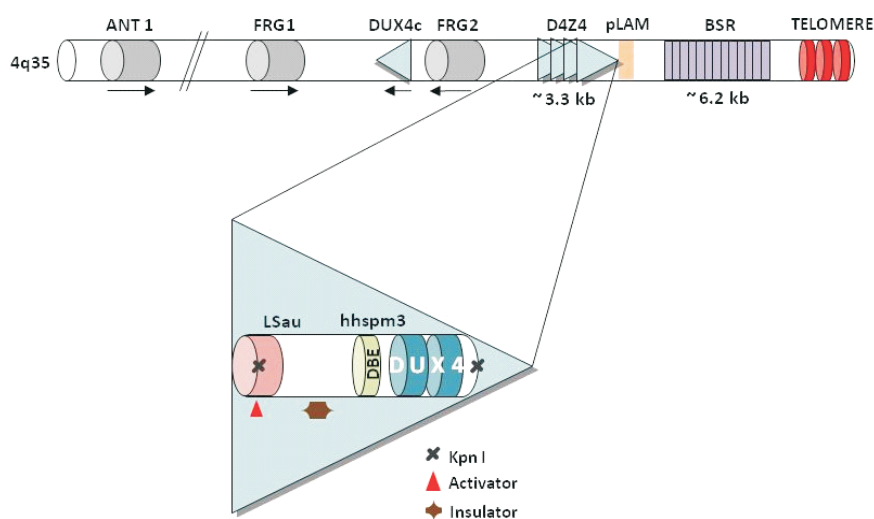
Fig. 2. Structure of 4q35 locus and D4Z4 repeat. Each 3303 bp repeat is flanked by KpnI sites and contains following regions: LSau; hhspm3; DBE and DUX4 ORF (positioned, respectively 1–340 bp, 1313–1780 bp, 1584–1611 bp and 1792–3063 bp from the 5' KpnI site) [91, 92, 94]. The positions of two *cis*-elements: an activator [96] and an insulator [105] are also indicated (1–170 bp and 460–1197 bp). Here, DBE (D4Z4 Binding Element), BSR (beta-satellite repeat), ANT1 (Adenine Nucleotide Translocator gene1), FRG1/2 (FSHD Region Gene 1/2)

340 bp, whereas hhspm3 (a human DNA insert showing sperm-specific hypomethylation Sp-0.3-8) is a GC-rich low copy repeat sequence of 467 bp. The open reading frame has neither introns nor a polyadenylation site in the D4Z4 repeat unit and is preceded by a promoter (box TACAA) located 149 bp upstream [93]. DBE (D4Z4 Binding Element), located 181 bp upstream of DUX4, is a binding site for the YY1 transcriptional repressor, HMGB2 (High-Mobility Group Protein B2) and nucleolin and was shown to be responsible for epigenetic repression [94, 95]. In addition, the presence of a strong enhancer was demonstrated in each D4Z4 repeat [96]. Both the transcriptional repressor DBE and the transcriptional activator were shown to mediate the transcriptional control of 4q35 genes.

The 4qter and 10qter regions possess 99 % of homology, which extends over more than 200 kb [97]. The 10qter located repeats can be distinguished by the presence of a Blnl restriction site [97]. There is a repeat exchange between 10qter and 4qter regions in 20 to 30 % of the population [98], which creates 4q/10q hybrid sequences. FSHD is specific to chromosome 4 because the contraction in 10q26 is not associated with the disease.

The second satellite DNA sequence associated with FSHD is a BSR tandem of 6.2 kb length (4qA allele) in 4q35 locus downstream of D4Z4 [40]. The 4qA allele is present in about half of a normal population. In contrast to the 4qA, the 4qB allele does not contain the BSR tandem and the reduction of D4Z4 MSR tandem does not result in the manifestation of the disease. Evolutionary analysis by Van Geel *et al*. indicates that the 4qA allele

is older than the 4qB [99]. Moreover, evolution-wise, FSHD appears as a very young disease, present only in humans, as the linked D4Z4-BSR clusters from the FSHD region were found only in chimpanzees, and even this primate has never been found to suffer from FSHD [100].

The FSHD-like, or FSHD2, type of the disease represents 5 % of FSHD and is characterised by the high frequency of sporadic cases (70 %) and the absence of macrosatellite contraction in 4q. Nevertheless, the presence of the BSR 4qA allele remains a necessary condition for FSHD2 development [101]. Recently, we have demonstrated that the BSR fragment from the 4qA allele possesses the properties of a transcriptional activator [102]. These observations strongly suggest the significance of the BSR presence in FSHD development.

Macrosatellites can provoke epigenetic changes. *Chromatin*. First studies showed that the D4Z4 macrosatellite has in most cases the features of «unexpressed euchromatin» [103]. Some specific changes in post-translational modifications (PTM) of histones in FSHD were described [104]. In healthy individuals, D4Z4 has both the heterochromatic (trimethylation of H3K9 and H3K27) and euchromatic marks (dimethylation of H3K4 and acetylation of H3). The FSHD patients present a loss of H3K9 trimethylation in both chromosomes 4q and 10q.

In addition to the changes in histone PTMs, several studies demonstrated that the reduced number of D4Z4 macrosatellite repeats in FSHD myoblasts was associa-
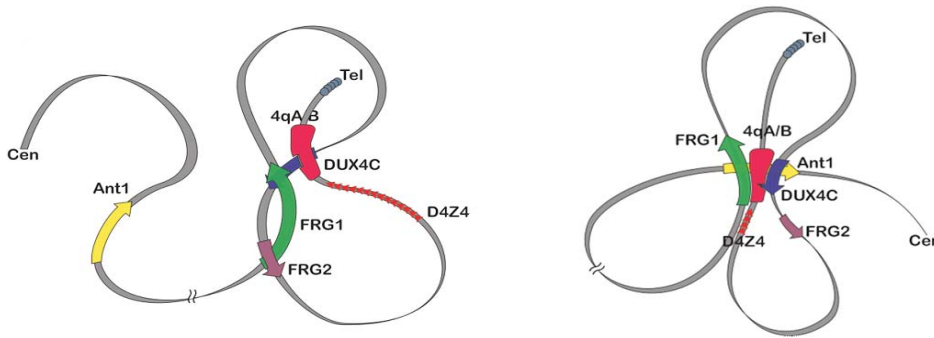
Fig. 3. Architecture of 4q35 FSHD locus. The reduction of macrosatellite tandem provokes the drastic changes in chromatin organization. The 3C analysis of chromatin organization of control (left) and FSHD (right) myoblasts [102]

ted with a global alteration of the chromatin organization in the 4q35 region [74, 75]. Moreover, our work has provided evidence about the potential role of the 4qA BSR marker in chromatin remodeling in FSHD (Fig. 3) [102]. These results have been obtained by the Chromosome Conformation Capture (3C) methodology, which evaluates the spatial proximity of two given genomic fragments based upon their relative propensity to be ligated *in vitro*.

Later on, Ottavini *et al*. showed that the D4Z4 macrosatellite repeat acts as a CTCF insulator protecting the adjacent genes, activated in pathology, from their heterochromatization in FSHD patients (Fig. 2) [105]. The loss of this feature was observed with the increase of the number of D4Z4 repeat units. This is an example of how a change in tandem repeat copy number (i) leads to a switch of its function from repressor to insulator and (ii) provokes dramatic downstream biological effects.

DNA methylation. The significant hypomethylation of D4Z4 CpG dinucleotides was observed in FSHD1 patients compared with the healthy individuals [106]. FSHD1 patients show hypomethylation in the contracted D4Z4 allele, with the methylation degree depending on the number of repeats. In particular, the hypomethylation is more significant in the patients with short D4Z4 tandem (3–6 repeat units) than in the patients with moderate size of MSR (6–9 repeat units). This observation suggests a correlation between the severity of disease, the number of D4Z4 repeats and their methylation level [107]. On the other hand, hypomethylation might contribute to the disease independent of contraction, as the FSHD2 patients without a decrease in the number of D4Z4 repeat units also present a strong hypomethylation of the satellite on chromosomes 4q and

10q [106, 108]. In both cases (FSHD1 and FSHD2), the hypomethylation takes place in the macrosatellite region (D4Z4) and does not extend to the adjacent region towards centromere. Nothing is known about the methylation status of the adjacent region extending towards telomere.

The D4Z4 hypomethylation may provide the missing link between DNA changes and transcriptional derepression in FSHD. Whatever the exact nature of this mechanism is, the D4Z4 hypomethylation in FSHD individuals strongly supports a central role of epigenetic events in the pathogenic pathway of FSHD.

Satellite DNA: therapeutic targets? Several publications have highlighted the existence of crosstalk between the MSR DUX4 expression and the differentiation state of cells. The *MSR* gene is expressed in germ line and embryonic cells and is epigenetically repressed in somatic cells [44, 109]. In FSHD patients, it has been proposed that the DUX4 transcript is transiently expressed at a pre-myoblast stage in the affected regions of skeletal muscles and possibly among certain subsets of muscle satellite cells. This could result (i) in the upregulation of the expression of adjacent genes and (ii) in the dampening of the expression of many muscle lineage-associated genes, during regenerative myogenesis. This, in turn, could decrease the efficiency of the late stages of regenerative myogenesis or affect the muscle function in a manner consistent with the usually slow progression of FSHD.

Recently, another mechanism of the MSR DUX4 expression has been proposed. It consists in the expression of a DUX4 full length stable RNA, which is transcribed from the last unit of D4Z4, which includes the 3'UTR region containing two facultative introns and a

255

polyadenylation site. The canonical polyadenylation site is only present in the pLAM sequence associated with BSR of the 4qA allele [41, 93]. This long DUX4 transcript was found in half of examining biopsies samples from FSHD patients. Additionally, the high level of this long DUX4 transcript was observed in special pools of muscle cells representing 0.1 % of FSHD culturing muscles [110]. According to these two models of FSHD development, DUX4 expression is considered as the main therapeutic target for FSHD.

Further on, Dixit and collaborators reported that, in the muscles of FSHD patients, the transcription of DUX4 was associated with an increase of a homeodomain transcription factor PITX1 [93]. They showed that the DUX4 protein can act as a transcriptional activator of PITX1. Another study showed the existence of partial transcripts of small RNAs (mi/siRNA) resulting from DUX4 [109]. These findings provide additional therapeutic targets for FSHD.

More recently, however, the DUX4-centric paradigm of FSHD development has been challenged. The inappropriate expression of the DUX4 stable transcript that takes place in FSHD pathogenesis was observed in FSHD fibroblasts [110], which are not affected in FSHD, as well as in muscles and myogenic cells of some unaffected individuals without D4Z4 deletion [111]. These somewhat controversial observations suggest that the role of DUX4 expression in FSHD and its potential role as a therapeutic target require a more thorough exploration.

As far as the second satellite repeat (BSR) actor in the FSHD development is concerned, little is known about the functional role of the 4qA allele. Previously, we have shown that a 1.5 kbp fragment from 4qA allele contains a transcriptional activator [102], allowing us (i) to suggest that the presence of BSR could influence the transcriptional activation of adjacent or juxtaposed sequences and (ii) to propose that the protein activator of 4qA would be a new therapeutic target for FSHD.

Altogether, the recent progress in the studies of FSHD, including our own data, suggests that the specific satellite DNA, macrosatellite (D4Z4) and beta-satellite (4qA) sequences are essential for the development of this disease and could be considered as potential targets for the gene therapy in FSHD treatment.

Conclusions. Much experimental evidence reviewed in this manuscript, indicates that the study of satellite DNA promises to provide important insights into human diseases, as they are related to transcription control and often have an impact on chromatin structure and genomic instability [112]. Despite many technical challenges due to their repeated nature, the manipulation of satellite DNA might have great therapeutic potential in the treatment of FSHD, neurological, developmental disorders and cancers. These arguments warrant further investigations on various mechanisms regarding satellite DNA functions in the genome.

Сателітна ДНК і пов'язані хвороби

Ж. Річ, В. В. Огризько , І. В. Пірожкова

Резюме

Сателітна ДНК, також відома як тандемно повторювана ДНК, складається з кластерів повторюваних послідовностей, об'єднаних у широкий клас часто повторюваних елементів. Сателітні ДНК можна розділити на декілька класів залежно від розміру окремого повтора: мікросателітні, мінісателітні, мідісателітні і макросателітні ДНК. Сателітну ДНК спочатку розглядали як «сміттєву». Лише зовсім недавно таку концепцію було переглянуто і наразі сателітну ДНК відносять до ДНК, якій притаманні різні функції. Крім того, присутність у геномі повторюваних послідовностей пов'язана з високою частотою мутацій, епігенетичними змінами і модифікаціями в профілі експресії генів, що потенційно може призвести до різних патологій. Таким чином, вивчення сателітної ДНК буде корисним при розробці терапії, спрямованої на лікування захворювань, пов'язаних із сателітною ДНК, таких як м'язова дистрофія FSHD, неврологічні патології, хвороби, обумовлені порушеннями розвитку, та онкологічні захворювання.

Ключові слова: сателітна ДНК, повторювані послідовності, частота мутацій, захворювання, пов'язані із сателітною ДНК.

Сателлитная ДНК и сопутствующие заболевания

Ж. Рич, В. В. Огрызько, И. В. Пирожкова

Резюме

Сателлитная ДНК, также известная как тандемно повторяющаяся ДНК, состоит из кластеров повторяющихся последовательностей, объединенных в широкий класс часто повторяющихся элементов. Сателлитную ДНК можно разделить на несколько классов в зависимости от размера отдельного повтора: микросателлитная, минисателлитная, мидисателлитная и макросателлитная ДНК. Сателлитную ДНК первоначально рассматривали как «мусорную». Только совсем недавно эта концепция была пересмотрена, в результате чего сателлитную ДНК относят к ДНК, обладающей различными функциями. Кроме того, присутствие в геноме повторяющихся последовательностей связано с высокой частотой мутаций, эпигенетическими изменениями и модификациями в профиле экспрессии генов, что потенциально может привести к различным патологиям. Таким образом, изучение сателлитной ДНК будет полезно при разработке тера-

пии, направленной на лечение заболеваний, таких как мышечная дистрофия FSHD, неврологические патологии, болезни, обусловленные нарушениями развития, и онкологические заболевания.

Ключевые слова: сателлитная ДНК, повторяющиеся последовательности, частота мутаций, заболевания, связанные с сателлитной ДНК.

REFERENCES

1. *Goldenfeld N, Woese C*. Life is physics: evolution as a collective phenomenon far from equilibrium. *Annu Rev Condens Matter Phys*. 2011;**2**:375-99.

2. *Ohno S*. So much «junk» DNA in our genome. *Brookhaven Symp Biol*. 1972;**23**:366–70.

3. *Dawkins R*. The greatest show on Earth: the evidence for evolution. 2010; 480 p.

4. *Nobrega MA, Zhu Y, Plajzer-Frick I, Afzal V, Rubin EM*. Megabase deletions of gene deserts result in viable mice. *Nature*. 2004;**431**(7011):988–93.

5. *Hoeppner MP, White S, Jeffares DC, Poole AM*. Evolutionarily stable association of intronic snoRNAs and microRNAs with their host genes. *Genome Biol Evol*. 2009;**1**:420–8.

6. *Pheasant M, Mattick JS*. Raising the estimate of functional human sequences. *Genome Res*. 2007;**17**(9):1245–53.

7. *Wright FA, Lemon WJ, Zhao WD, et al*. A draft annotation and overview of the human genome. *Genome Biol*. 2001;**2**(7): RESEARCH0025.

8. *Bejerano G, Pheasant M, Makunin I, et al*. Ultraconserved elements in the human geno me. *Science*. 2004;**304**(5675):1321–5.

9. *Minovitsky S, Gee SL, Schokrpur S, Dubchak I, Conboy JG*. The splicing regulatory element, UGCAUG, is phylogenetically and spatially conserved in introns that flank tissue-specific alternative exons. *Nucleic Acids Res*. 2005;**33**(2):714–24.

10. *Hui J, Hung LH, Heiner M, et al*. Intronic CA-repeat and CA-rich elements: a new class of regulators of mammalian alternative splicing. *EMBO J*. 2005;**24**(11):1988–98.

11. *Nakaya HI, Amaral PP, Louro R, et al*. Genome mapping and expression analyses of human intronic noncoding RNAs reveal tissue-specific patterns and enrichment in genes related to regulation of transcription. *Genome Biol*. 2007;**8**(3):R43.

12. *Barash Y, Calarco JA, Gao W, et al*. Deciphering the splicing code. *Nature*. 2010;**465** (7294):53–9.

13. *Baskerville S, Bartel DP*. Microarray profiling of microRNAs reveals frequent coexpression with neighboring miRNAs and host genes. *RNA*. 2005;**11**(3):241–7.

14. *Alotaibi H, Yaman E, Salvatore D, et al*. Intronic elements in the Na$^+$/I-symporter gene (NIS) interact with retinoic acid receptors and mediate initiation of transcription. *Nucleic Acids Res*. 2010; **38**(10):3172– 85.

15. *Lavrov SA, Kibanov MV*. Noncoding RNAs and chromatin structure. *Biochemistry (Mosc)*. 2007;**72**(13):1422–38.

16. *Rodriguez-Campos A, Azorin F*. RNA is an integral component of chromatin that contributes to its structural organization. *PLoS One*. 2007;**2**(11):e1182.

17. *Sorge J, Gross E, West C, Beutler E*. High level transcription of the glucocerebrosidase pseudogene in normal subjects and patients with Gaucher disease. *J Clin Invest*. 1990;**86**(4): 1137–41.

18. *Schmutzler C, Gross HJ*. Genes, variant genes, and pseudogenes of the human tRNA(Val) gene family are differentially expressed in HeLa cells and in human placenta. *Nucleic Acids Res*. 1990;**18**(17):5001–8.

19. *Watanabe T, Totoki Y, Toyoda A, et al*. Endogenous siRNAs from naturally formed dsRNAs regulate transcripts in mouse oocytes. *Nature*. 2008;**453**(7194):539–43.

20. *Korneev SA, Park JH, O'Shea M*. Neuronal expression of neural nitric oxide synthase (nNOS) protein is suppressed by an antisense RNA transcribed from an NOS pseudogene. *J Neurosci*. 1999; **19**(18):7711–20.

21. *de Koning AP, Gu W, Castoe TA, Batzer MA, Pollock DD*. Repetitive elements may comprise over two-thirds of the human genome. *PLoS Genet*. 2011;**7**(12):e1002384.

22. *Turnpenny PD, Ellard S*. Emery's Elements of Medical Genetics. Philadelphia, Elsevier. 2012; 464 p.

23. *Richard GF, Kerrest A, Dujon B*. Comparative genomics and molecular dynamics of DNA repeats in eukaryotes. *Microbiol Mol Biol Rev*. 2008;**72**(4):686–727.

24. *Kit S*. Equilibrium sedimentation in density gradients of DNA preparations from animal tissues. *J Mol Biol*. 1961;**3**:711–6.

25. *Sueoka N*. Compositional correlation between deoxyribonucleic acid and protein. *Cold Spring Harb Symp Quant Biol*. 1961;**26**: 35–43.

26. *Yunis JJ, Yasmineh WG*. Heterochromatin, satellite DNA, and cell function. Structural DNA of eukaryotes may support and protect genes and aid in speciation. *Science*. 1971;**174**(4015):1200–9.

27. *Tremblay DC, Alexander G Jr, Moseley S, Chadwick BP*. Expression, tandem repeat copy number variation and stability of four macrosatellite arrays in the human genome. *BMC Genomics*. 2010;**11**:632.

28. *Balog J, Miller D, Sanchez-Curtailles E, et al*. Epigenetic regulation of the X-chromosomal macrosatellite repeat encoding for the cancer/testis gene *CT47*. *Eur J Hum Genet*. 2012;**20**(2): 185–91.

29. *Picardo M, Ottaviani M, Camera E, Mastrofrancesco A*. Sebaceous gland lipids. *Dermatoendocrinol*. 2009;**1**(2):68–71.

30. *Lebre AS, Jamot L, Takahashi J, Spassky N, et al*. Ataxin-7 interacts with a Cbl-associated protein that it recruits into neuronal intranuclear inclusions. *Hum Mol Genet*. 2001;**10**(11):1201–13.

31. *Payseur BA, Nachman MW*. Microsatellite variation and recombination rate in the human genome. *Genetics*. 2000;**156**(3):1285–98.

32. *Galaeva MV, Fayt VI, Sivolap YuM*. Genetic diversity of the gene pool of bread winter wheat alleles of microsatellite locus Xgwm 182-5D associated with frost resistance. *Biopolym Cell*. 2013; **29**(6):487–92.

33. *Slishchuk GI, Volkova N.E, Sivolap YuM*. Bioinformatics analysis of the secondary structure of maize whp1 gene intron 1 transcripts. *Biopolym Cell*. 2012;**28**(2):156–60.

34. *Kravchenko SA, Livshits LA*. Nature and origin of germline mutations in human tandem repeated loci. *Biopolym Cell*. 2007;**23** (3):188–201.

35. *Vergnaud G, Denoeud F*. Minisatellites: mutability and genome architecture. *Genome Res*. 2000;**10**(7):899–907.

36. *Rudd MK, Mays RW, Schwartz S, Willard HF*. Human artificial chromosomes with alpha satellite-based *de novo* centromeres show increased frequency of nondisjunction and anaphase lag. *Mol Cell Biol*. 2003;**23**(21):7689–97.

37. *Waye JS, Willard HF*. Human beta satellite DNA: genomic organization and sequence definition of a class of highly repetitive tandem DNA. *Proc Natl Acad Sci USA*. 1989;**86**(16):6250–4.

38. *Meneveri R, Agresti A, Marozzi A, et al*. GC-rich heterochromatic blocks. *Gene*. 1993;**123**(2):227–34.

39. *Cardone MF, Ballarati L, Ventura M, et al*. Evolution of beta satellite DNA sequences: evidence for duplication-mediated repeat amplification and spreading. *Mol Biol Evol*. 2004;**21**(9): 1792–9.

40. *Lemmers RJ, de Kievit P, Sandkuijl L, et al.* Facioscapulohumeral muscular dystrophy is uniquely associated with one of the two variants of the 4q subtelomere. *Nat Genet.* 2002;**32**(2): 235–6.

41. *Lemmers RJ, van der Vliet PJ, Klooster R, et al.* A unifying genetic model for facioscapulohumeral muscular dystrophy. *Science.* 2010;**329**(5999):1650–3.

42. *Arnoult N, Schluth-Bolard C, Letessier A, et al.* Replication timing of human telomeres is chromosome arm-specific, influenced by subtelomeric structures and connected to nuclear localization. *PLoS Genet.* 2010;**6**(4):e1000920.

43. *Kim JH, Ebersole T, Kouprina N, et al. H*uman gamma-satellite DNA maintains open chromatin structure and protects a transgene from epigenetic silencing. *Genome Res.* 2009;**19**(4):533–44.

44. *Geng LN, Yao Z, Snider L, et al. D*UX4 activates germline genes, retroelements, and immune mediators: implications for facioscapulohumeral dystrophy. *Dev Cell.* 2012;**22**(1):38–51.

45. *Chadwick BP.* DXZ4 chromatin adopts an opposing conformation to that of the surrounding chromosome and acquires a novel inactive X-specific role involving CTCF and antisense transcripts. *Genome Res.* 2008;**18**(8):1259–69.

46. *Djian P.* Evolution of simple repeats in DNA and their relation to human disease. *Cell.* 1998;**94**(2):155–60.

47. *Mirkin SM.* Expandable DNA repeats and human disease. *Nature.* 2007;**447**(7147):932–40.

48. *Mirkin SM.* DNA structures, repeat expansions and human hereditary disorders. *Curr Opin Struct Biol.* 2006;**16**(3):351–8.

49. *Santoro MR, Bray SM, Warren ST.* Molecular mechanisms of fragile X syndrome: a twenty-year perspective. *Annu Rev Pathol.* 2012;**7**:219–45.

50. *Brook JD, McCurrach ME, Harley HG, et al.* Molecular basis of myotonic dystrophy: expansion of a trinucleotide (CTG) repeat at the 3' end of a transcript encoding a protein kinase family member. *Cell.* 1992;**69**(2):385.

51. *A novel* gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. The Huntington's Disease Collaborative Research Group. *Cell.* 1993;**72** (6):971–83.

52. *Klockgether T.* Update on degenerative ataxias. *Curr Opin Neurol.* 2011;**24**(4):339–45.

53. *Scott HS, Kudoh J, Wattenhofer M, et al.* Insertion of beta-satellite repeats identifies a transmembrane protease causing both congenital and childhood onset autosomal recessive deafness. *Nat Genet.* 2001;**27** (1):59–63.

54. *Zoghbi HY, Orr HT.* Glutamine repeats and neurodegeneration. *Annu Rev Neurosci.* 2000;**23**:217–47.

55. *Nakamura K, Jeong SY, Uchihara T, et al.* SCA17, a novel autosomal dominant cerebellar ataxia caused by an expanded polyglutamine in TATA-binding protein. *Hum Mol Genet.* 2001;**10** (14):1441–8.

56. *Warren ST, Sherman SL.* The fragile X syndrome. *The Metabolic and Molecular Bases of Inherited Disease* Eds CR Scriver, AL Beaudet, D Valle, *et al.* New York: McGraw-Hill Companies, 2001; Vol. I:1257–90.

57. *Schlotterer C, Harr B.* Microsatellite Instability.–eLS: Chichester, John Wiley & Sons Ltd, 2004.

58. *Nelson DL, Orr HT, Warren ST.* The unstable repeats – three evolving faces of neurological disease. *Neuron.* 2013;**77**(5): 825–43.

59. *Zoghbi HY, Orr HT.* Pathogenic mechanisms of a polyglutamine-mediated neurodegenerative disease, spinocerebellar ataxia type 1. *J Biol Chem.* 2009;**284**(12):7425–9.

60. *Harper PS, Harley HG, Reardon W, Shaw DJ.* Anticipation in myotonic dystrophy: new light on an old problem. *Am J Hum Genet.* 1992;**51**(1):10–6.

61. *Richards RI, Sutherland GR.* Heritable unstable DNA sequences. *Nat Genet.* 1992;**1**(1):7–9.

62. *Burgess DJ.* Chromosome instability: Tumorigenesis via satellite link. *Nat Rev Cancer.* 2011;**11**(3):158.

63. *Narayan A, Ji W, Zhang XY, Marrogi A, Graff JR, Baylin SB, Ehrlich M.* Hypomethylation of pericentromeric DNA in breast adenocarcinomas. *Int J Cancer.* 1998;**77**(6):833–8.

64. *Qu G, Dubeau L, Narayan A, Yu MC, Ehrlich M.* Satellite DNA hypomethylation *vs.* overall genomic hypomethylation in ovarian epithelial tumors of different malignant potential. *Mutat Res.* 1999;**423**(1–2):91–101.

65. *Qu GZ, Grundy PE, Narayan A, Ehrlich M.* Frequent hypomethylation in Wilms tumors of pericentromeric DNA in chromosomes 1 and 16. *Cancer Genet Cytogenet.* 1999;**109**(1):34–9.

66. *Feber A, Wilson GA, Zhang L, et al.* Comparative methylome analysis of benign and malignant peripheral nerve sheath tumors. *Genome Res.* 2011;**21**(4):515–24.

67. *Wong N, Lam WC, Lai PB, Pang E, Lau WY, Johnson PJ.* Hypomethylation of chromosome 1 heterochromatin DNA correlates with q-arm copy gain in human hepatocellular carcinoma. *Am J Pathol.* 2001;**159**(2):465–71.

68. *Zhu Q, Pao GM, Huynh AM, et al.* BRCA1 tumour suppression occurs via hete- rochromatin-mediated silencing. *Nature.* 2011; **477**(7363):179–84.

69. *Molecular* Genetics of Liver Neoplasia. Eds XW Wang, JW Grisham, SS Thorgeirsson. Springer, 2010; 399 p.

70. *Ting DT, Lipson D, Paul S, et al.* Aberrant overexpression of satellite repeats in pancreatic and other epithelial cancers. *Science.* 2011;**331**(6017):593–6.

71. *Nishiyama R, Qi L, Tsumagari K, Weissbecker K, et al.* A DNA repeat, NBL2, is hypermethylated in some cancers but hypomethylated in others. *Cancer Biol Ther.* 2005;**4**(4):440–8.

72. *Katargin AN, Pavlova LS, Kisseljov FL, Kisseljova NP.* Hypermethylation of genomic 3.3-kb repeats is frequent event in HPV-positive cervical cancer. *BMC Med Genomics.* 2009; **2**:30.

73. *Ehrlich M.* DNA hypomethylation in cancer cells. *Epigenomics.* 2009;**1**(2):239–59.

74. *Petrov A, Pirozhkova I, Carnac G, Laoudj D, Lipinski M, Vassetzky YS.* Chromatin loop domain organization within the 4q35 locus in facioscapulohumeral dystrophy patients versus normal human myoblasts. *Proc Natl Acad Sci USA.* 2006; **103**(18): 6982–7.

75. *Bodega B, Ramirez GD, Grasser F, et al.* Remodeling of the chromatin structure of the facioscapulohumeral muscular dystrophy (FSHD) locus and upregulation of FSHD-related gene 1 (FRG1) expression during human myogenic differentiation. *BMC Biol.* 2009;**7**:41.

76. *Perkel JM.* Visiting «noncodarni»a. *Biotechniques.* 2013;**54**(6): 301, 303–4.

77. *Cabianca DS, Casa V, Gabellini D.* A novel molecular mechanism in human genetic disease: a DNA repeat-derived lncRNA. *RNA Biol.* 2012;**9**(10):1211–7.

78. *Cabianca DS, Casa V, Bodega B, Xynos A, Ginelli E, Tanaka Y, Gabellini D.* A long ncRNA links copy number variation to a polycomb/trithorax epigenetic switch in FSHD muscular dystrophy. *Cell.* 2012;**149**(4):819–31.

79. *Gibb EA, Vucic EA, Enfield KS, et al.* Human cancer long noncoding RNA transcriptomes. *PLoS One.* 2011;**6**(10):e25915.

80. *Jandial R.* Metastatic cancer: clinical and biological perspectives. Austin, Landes Bioscience, 2013; 312 p.

81. *Banfai B, Jia H, Khatun J, et al.* Long noncoding RNAs are rarely translated in two human cell lines. *Genome Res.* 2012;**22**(9): 1646–57.

82. *Khurana E, Fu Y, Colonna V, Mu XJ, et al.* Integrative annotation of variants from 1092 humans: application to cancer genomics. *Science*. 2013;**342**(6154):1235587.

83. *Padberg GW, Lunt PW, Koch M, Fardeau M.* Diagnostic criteria for facioscapulohumeral muscular dystrophy. *Neuromuscul Disord*. 1991;**1**(4):231–4.

84. *Tawil R, Figlewicz DA, Griggs RC, Weiffenbach B.* Facioscapulohumeral dystrophy: a distinct regional myopathy with a novel molecular pathogenesis. FSH Consortium. *Ann Neurol.* 1998;**43** (3):279–82.

85. *Wijmenga C, Frants RR, Brouwer OF, Moerer P, Weber JL, Padberg GW.* Location of facioscapulohumeral muscular dystrophy gene on chromosome 4. *Lancet*. 1990;**336**(8716):651–3.

86. *Sarfarazi M, Wijmenga C, Upadhyaya M, et al.* Regional mapping of facioscapulohumeral muscular dystrophy gene on 4q35: combined analysis of an international consortium. *Am J Hum Genet*. 1992;**51**(2):396–403.

87. *Wijmenga C, Hewitt JE, Sandkuijl LA, et al.* Chromosome 4q DNA rearrangements associated with facioscapulohumeral muscular dystrophy. *Nat Genet*. 1992;**2** (1):26–30.

88. *Lunt PW.* 44th ENMC International Workshop: Facioscapulohumeral Muscular Dystrophy: Molecular Studies 19–21 July 1996, Naarden, The Netherlands. *Neuromuscul Disord*. 1998;**8**(2): 126–30.

89. *Lunt PW, Jardine PE, Koch M, et al.* Phenotypic-genotypic correlation will assist genetic counseling in 4q35-facioscapulohumeral mus- cular dystrophy. *Muscle Nerve Suppl*. 1995;**2**:S103–9.

90. *Tupler R, Berardinelli A, Barbierato L, et al.* Monosomy of distal 4q does not cause facioscapulohumeral muscular dystrophy. *J Med Genet*. 1996;**33**(5):366–70.

91. *Hewitt JE, Lyle R, Clark LN, et al.* Analysis of the tandem repeat locus D4Z4 associated with facioscapulohumeral muscular dystrophy. *Hum Mol Genet.* 1994;**3** (8):1287–95.

92. *Gabriels J, Beckers MC, Ding H, et al.* Nucleotide sequence of the partially deleted D4Z4 locus in a patient with FSHD identifies a putative gene within each 3.3 kb element. *Gene*. 1999; **236**(1):25–32.

93. *Dixit M, Ansseau E, Tassin A, et al.* DUX4, a candidate gene of facioscapulohumeral muscular dystrophy, encodes a transcriptional activator of PITX1. *Proc Natl Acad Sci USA*. 2007;**104** (46):18157–62.

94. *Gabellini D, Green MR, Tupler R.* Inappropriate gene activation in FSHD: a repressor complex binds a chromosomal repeat deleted in dystrophic muscle. *Cell*. 2002;**110**(3):339–48.

95. *Cabianca DS, Casa V, Gabellini D.* A novel molecular mechanism in human genetic disease: a DNA repeat-derived lncRNA. *RNA Biol*. 2012;**9**(10):1211–7.

96. *Petrov A, Allinne J, Pirozhkova I, Laoudj D, Lipinski M, Vassetzky YS.* A nuclear matrix attachment site in the 4q35 locus has an enhancer-blocking activity *in vivo*: implications for the facioscapulo-humeral dystrophy. *Genome Res*. 2008;**18**(1):39–45.

97. *Deidda G, Cacurri S, Piazzo N, Felicetti L.* Direct detection of 4q35 rearrangements implicated in facioscapulohumeral muscular dystrophy (FSHD). *J Med Genet*. 1996;**33**(5):361–5.

98. *Rossi M, Ricci E, Colantoni L, et al.* The Facioscapulohumeral muscular dystrophy region on 4qter and the homologous locus on 10qter evolved independently under different evolutionary pressure. *BMC Med Genet*. 2007;**8**:8.

99. *van Geel M, Dickson MC, Beck AF, et al.* Genomic analysis of human chromosome 10q and 4q telomeres suggests a common origin. *Genomics*. 2002;**79**(2):210–7.

100. *Giussani M, Cardone MF, Bodega B, Ginelli E, Meneveri R.* Evolutionary history of linked D4Z4 and Beta satellite clusters at the FSHD locus (4q35). *Genomics*. 2012;**100**(5):289–96.

101. *de Greef JC, Lemmers RJ, Camano P, et al.* Clinical features of facioscapulohumeral muscular dystrophy 2. *Neurology*. 2010; **75**(17):1548– 54.

102. *Pirozhkova I, Petrov A, Dmitriev P, Laoudj D, Lipinski M, Vassetzky Y.* A functional role for 4qA/B in the structural rearrangement of the 4q35 region and in the regulation of FRG1 and ANT1 in facioscapulohumeral dystrophy. *PLoS One*. 2008;**3** (10):e3389.

103. *Jiang G, Yang F, van Overveld PG, Vedanarayanan V, van der Maarel S, Ehrlich M.* Testing the position-effect variegation hypothesis for facioscapulohumeral muscular dystrophy by analysis of histone modification and gene expression in subtelomeric 4q. *Hum Mol Genet*. 2003;**12**(22):2909–21.

104. *Zeng W, de Greef JC, Chen YY, et al. S*pecific loss of histone H3 lysine 9 trimethylation and HP1gamma/cohesin binding at D4Z4 repeats is associated with facioscapulohumeral muscular dystrophy (FSHD). *PLoS Genet*. 2009;**5**(7):e1000559.

105. *Ottaviani A, Rival-Gervier S, Boussouar A, et al.* The D4Z4 macrosatellite repeat acts as a CTCF and A-type lamins-dependent insulator in facio-scapulo-humeral dystrophy. *PLoS Genet*. 2009; **5**(2):e1000394.

106. *van Overveld PG, Lemmers RJ, Sandkuijl LA, et al.* Hypomethylation of D4Z4 in 4q-linked and non-4q-linked facioscapulohumeral muscular dystrophy. *Nat Genet*. 2003;**35**(4):315–7.

107. *van Overveld PG, Enthoven L, Ricci E, et al.* Variable hypomethylation of D4Z4 in facioscapulohumeral muscular dystrophy. *Ann Neurol*. 2005;**58**(4):569–76.

108. *de Greef JC, Lemmers RJ, van Engelen BG, et al.* Common epigenetic changes of D4Z4 in contraction-dependent and contraction-independent FSHD. *Hum Mutat*. 2009;**30**(10):1449–59.

109. *Snider L, Asawachaicharn A, Tyler AE, et al.* RNA transcripts, miRNA-sized fragments and proteins produced from D4Z4 units: new candidates for the pathophysiology of facioscapulohumeral dystrophy. *Hum Mol Genet*. 2009;**18**(13):2414–30.

110. *Snider L, Geng LN, Lemmers RJ, et al.* Facioscapulohumeral dystrophy: incomplete suppression of a retrotransposed gene. *PLoS Genet*. 2010;**6**(10):e1001181.

111. *Jones TI, Chen JC, Rahimov F, et al.* Facioscapulohumeral muscular dystrophy family studies of DUX4 expression: evidence for disease modifiers and a quantitative model of pathogenesis. *Hum Mol Genet.* 2012;**21** (20):4419–30.

112. *Jurka J, Kapitonov VV, Kohany O, Jurka MV.* Repetitive sequences in complex genomes: structure and evolution. *Annu Rev Genomics Hum Genet*. 2007;**8**:241–59.