npg

RESEARCH HIGHLIGHT

The activatory long non-coding RNA DBE-T reveals the epigenetic etiology of facioscapulohumeral muscular dystrophy

Miguel Vizoso[1], Manel Esteller[1, 2, 3]

[1]Cancer Epigenetics and Biology Program (PEBC), Bellvitge Biomedical Research Institute (IDIBELL), 08907 L'Hospitalet, Barcelona, Catalonia, Spain; [2]Department of Physiological Sciences II, School of Medicine, University of Barcelona, Barcelona, Catalonia, Spain; [3]Institucio Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Catalonia, Spain

Facioscapulohumeral muscular dystrophy (FSHD) is a neuromuscular disorder often considered to be the third most common muscular dystrophy. Deletions reducing the copy number of the D4Z4 repeat in the distal end of the 4q arm are the main genetic cause of the disease. The recently highlighted research has identified a transcriptional activatory long non-coding RNA involved in the disease that acts through the recruitment of ASH1L, a protein belonging to the Trithorax family.

Facioscapulohumeral muscular dystrophy (FSHD) is a neuromuscular disorder transmitted in an autosomal dominant fashion with almost complete penetrance, affecting 1 in 20 000 individuals worldwide. The main clinical symptoms of patients with FSHD include progressive weakness of the shoulder and facial muscles, as well as of the lower extremities (distal and proximal) in later stages of the disease. Non-muscle tissue phenotypic consequences that occur very frequently are hearing loss (~75% patients) and retinal telangiectasia (~60% patients). Second-ary, less frequent clinical manifestations include respiratory insufficiency, cardiac conduction defects, learning difficulties and epilepsy.

The genetic region associated with FSHD is localized in the telomeric region of chromosome 4. Like many other genomic locations, it contains a repetitive sequence, in particular, a macrosatellite tandem array called D4Z4, which is highly variable in copy number among the population. In normal individuals, the copy number ranges from 11 to ~150 D4Z4 repeats while most affected individuals carry at least one chromosome with fewer than 11 repeats. The term macrosatellite is indicative of the large size of the affected genomic region. We should bear in mind not only that each D4Z4 repeat spans 3.3 kb but also that there are several protein-coding genes closely located with these repetitions. Interestingly, these genes have great myopathic potential (*ANT1/SLC25A4*, *FRG1/2*, or *DUX4*), which reinforces the notion that 4q35 locus is the Rosetta Stone of FSHD syndrome.

Since the 1990s, interest in the genetic basis of FSHD has centered on the deletions at the 4q35 locus that affect the D4Z4 repeat copy number. Significantly, these deletions were seen to be associated with chromatin architecture. The disruption caused by them induces critical chromatin modifications, whereby their loss produces changes in chromatin conformation, making it more accessible to transcription. These findings tell us that FSHD is not only a genetic disease but also one with an important epigenetic component. A recent paper published in *Cell* has proposed a new model that sets out the epigenetic basis of FSHD disease, and offers new and interesting discoveries from research in this area. Cabianca *et al*. [1] have turned their attention to a molecular world that has been poorly investigated until now. They describe a novel, long non-coding RNA (lncRNA), known as DBE-T, arising from repetitive sequences, that plays an important role in human disease through its involvement in gene regulation. This elegant study provides one of the first descriptions of a lncRNA acting as a master regulator of a retrogene, *DUX4*, that has conserved its ORF intact during mammalian evolution, and is involved in the etiology of an important human disorder. The lncRNA regulates how the chromatin structure of 4q35 locus changes its conformation from a heterochromatic/close state to a more euchromatic/open condition. They dissect the D4Z4

Correspondence: Manel Esteller
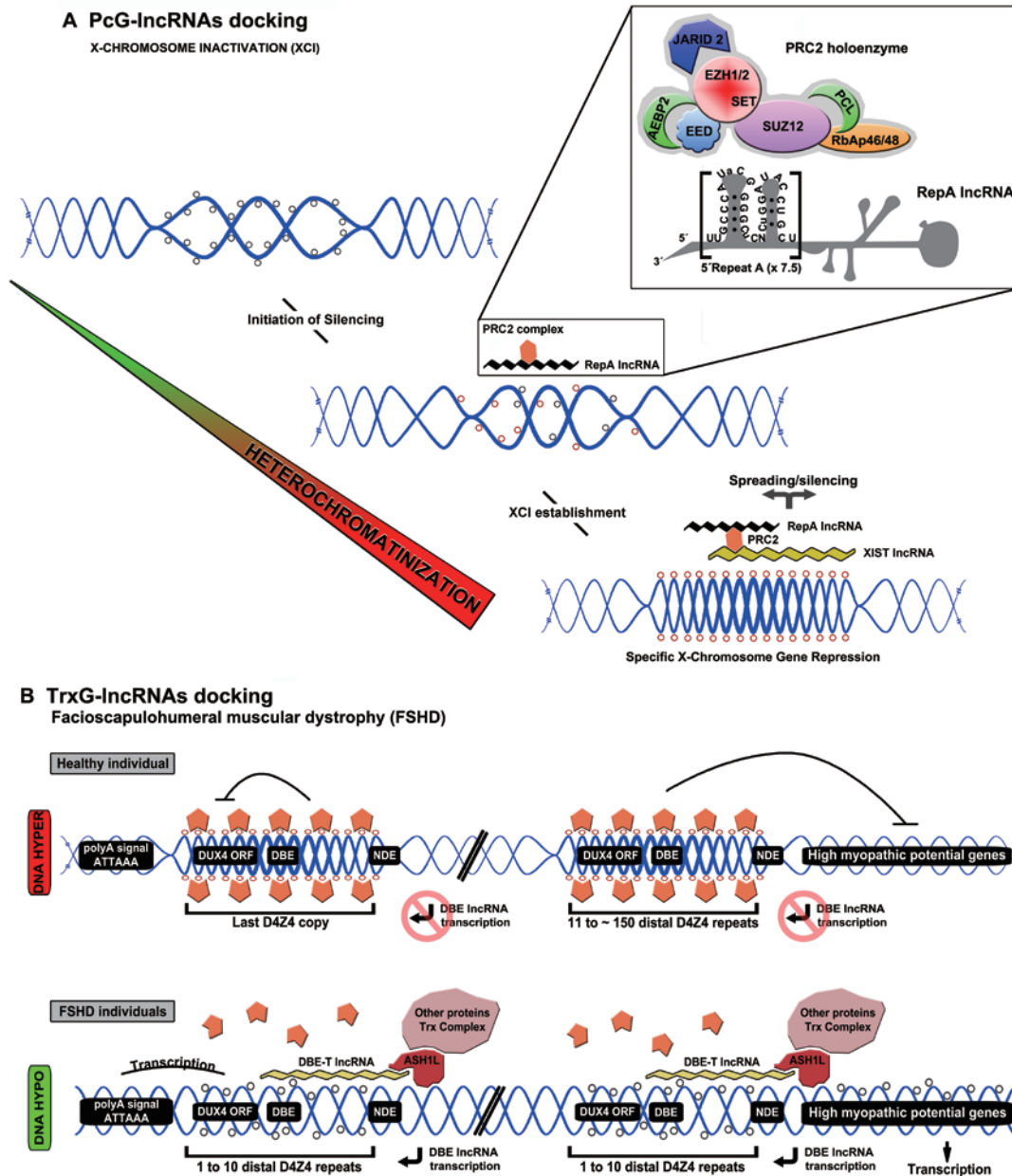E-mail: mesteller@idibell.cat

Figure 1 (A) Role of XIST/RepA lncRNAs in X-chromosome inactivation. Initially, chromatin is relaxed and open to transcription, indicated by open grey circles representing active histone marks such as histone 3 lysine 4 dimethylation (H3K4m2) or H4 acetylation. Initiation of heterochromatinization requires transcription of RepA lncRNA, which recruits PRC2 protein (orange rhombus). This interaction is magnified to highlight the importance of knowing which regions of the lncRNAs and proteins play a role in it. In this case, the 7.5 tandem repeats of two stem-loop structures (5′ repeat A) seem to be important. They are also present in the XIST transcript and interact with SUZ12 protein. Thus, a switch between activatory and repressive histone marks begins, resulting in trimethylation of H3K27. Finally, XIST lncRNA spreads along the X chromosome and propagates PRC2 silencing throughout the X chromosome. **(B)** Role of DBE-T transcript in FSHD disease. In healthy individuals, DNA hypermethylation of 4q35 locus is prevalent and there is generalized PRC2 coating over the chromatin strands. Open red circles represent H3K27m3. This is possible due to the many tandem repeat copies of the D4Z4 macrosatellite, which also inhibits DBE-T transcription. This scenario averts *DUX4* gene transcription from the final D4Z4 repeat and other myopathic genes (*ANT1* or *FRG1/2*), preventing the development of the FSHD syndrome. If the number of D4Z4 repeats is reduced below 11 copies, the epigenetic scenario changes dramatically, decreasing levels of DNA methylation, histone-repressive marks (e.g., H3K27m3; open red circles), and PRC2 attachment along the chromatin (orange polygons). However, this permissive environment enables the transcription of DBE-T lncRNA, which specifically recruits the ASH1L protein belonging to the Trx group. This is responsible for histone 3 lysine 36 dimethylation (H3K36m2) and histone 3 lysine 4 trimethylation (H3K4m3) activatory marks (open grey circles).

repeated unit into its specific putative functional elements and analyze the enrichment of their epigenetic marks. It is clear that the epigenetic marks at the 4q35 locus differ between healthy and affected individuals. In normal subjects, the high number of D4Z4 repeats allows the attachment of polycomb-repressive complex 2 (PRC2) to the DNA, promoting the maintenance of repressed chromatin through histone 3 lysine 27 trimethylation (H3K27m3), DNA methylation, and histone deacetylation marks. Conversely, in FSHD patients, reduction in D4Z4 repeats diminishes the presence of PRC2, causing the chromatin to open, and resulting in DBE-T transcription.

There are several examples of lncRNAs interacting with proteins of the PcG complex, which are major players in the cell-specific recruitment of PRC2 and regulators of PcG target genes. The specific gene targeting performed by lncRNAs may involve linear base pairing with target sequences. Base-to-base complementarity presents itself as the most powerful mechanism for directing proteins against regulatory DNA elements. However, there is also a growing belief that the tertiary structure of the lncRNAs may be the key to achieving this [2]. There are several examples of PRC2-recruiting lncRNAs. In mammals, Xist lncRNA coats one of the X chromosome *in cis*, then the RepA lncRNA recruits PRC2, interacting with Xist and spreading across the chromosome, finally resulting in histone methylation (H3K27m3) and heterochromatin formation [3] (Figure 1A). Another lncRNA, known as HOTAIR, was shown to recruit the PRC2 complex and to silence 40 kb of the *HOXD* locus [4]. It is also worth noting that more than 90% of all human DNA is transcribed and that the vast majority of these transcripts are of the non-protein-coding type [5], derived from antisense

transcription, which is more widespread than previously believed. Importantly, antisense transcription implies the existence of nascent RNA molecules that can modify the epigenetic landscape of promoter regions, changing the DNA methylation profiles of promoters in genes with CpG canonical islands or recruiting PRC2 complexes to the vicinity of the transcription start sites (TSS) of their associated genes (e.g., *Kcnq1ot1* lncRNA and the *KCNQ1* gene) [6].

In contrast to the aforementioned lncRNAs, and highlighting one of the most significant discoveries of Cabianca *et al.* [1], DBE-T is an example of a lncRNA that interacts with the PcG antagonist protein complex, the Trithorax group (Trx). Therefore, the DBE-T transcript is defined as a lncRNA that coordinates the activation of certain PRC2 target-repressed genes, such as *DUX4*, at the 4q35 locus (Figure 1B). The de-repression of this transcription factor results in significant cell toxicity. It is accumulated into the nucleus where it is involved in emerin relocalization and caspase 3 and/or 7 induction, resulting in increased cell death [7]. Moreover, its expression is correlated with lower *MyoD* expression levels, preventing normal cell signaling, and contributing to the final FSHD phenotype [8]. In particular, DBE-T interacts with one of the Trx components, the ASH1L protein, which has also been studied in *Drosophila*, demonstrating its involvement in the activation of Hox gene expression when it is recruited to Hox regulatory elements by lncRNAs [9]. Wang *et al.* [10] described the first activatory lncRNA in 2011: *HOTTIP*, a long intergenic non-coding RNA (lincRNA) involved in the production of *HOXA* genes. Cabianca *et al.* [1] have identified the first lncRNA linked to the Trithorax activation complex in relation to a particular disorder. A long list of lncRNAs involved in human disease

lies ahead of us to be discovered [11].

References

1 Cabianca DS, Casa V, Bodega B, *et al*. A long ncRNA links copy number variation to a Polycomb/Trithorax epigenetic switch in FSHD muscular Dystrophy. *Cell* 2012; **149**:819-831.

2 Margueron R, Reinberg D. The polycomb complex PRC2 and its mark in life. *Nature* 2011; **469**:343-349.

3 Lee JT. Lessons from X-chromosome inactivation: long ncRNA as guides and tethers to the epigenome. *Genes Dev* 2009; **23**:1831-1842.

4 Rinn JL, Kertesz M, Wang JK, *et al*. Functional demarcation of active and silent chromatin domains in human Hox loci by noncoding RNAs. *Cell* 2007; **129**:1311-1323.

5 Birney E, Stamatoyannopoulos JA, Dutta A, *et al*. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* 2007; **447**:799-816.

6 Pandey RR, Mondal T, Mohammad F, *et al*. Kcnq1ot1 antisense noncoding RNA mediates lineage-specific transcriptional silencing through chromatin-level regulation. *Mol Cell* 2008; **32**:232-246.

7 Kowaljow V, Marcowycz A, Ansseau E, *et al*. The DUX4 gene at the FSHD1A locus encodes a proapoptotic protein. *Neuromuscul Disord* 2007; **17**:611-623.

8 Bosnakovski D, Xu Z, Gang EJ, *et al*. An isogenic myoblast expression screen identifies DUX4-mediated FSHD-associated molecular pathologies. *EMBO J* 2008; **27**:2766-2779.

9 Sánchez-Elsner T, Gou D, Kremmer E, *et al*. Noncoding RNAs of trithorax response elements recruit *Drosophila* Ash1 to Ultrabithorax. *Science* 2006; **311**:1118-1123.

10 Wang K, Yang YW, Liu B, *et al*. A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression. *Nature* 2011; **472**:120-124.

11 Esteller M. Non-coding RNAs in human disease. *Nat Rev Genet* 2011; **12**:861-874.