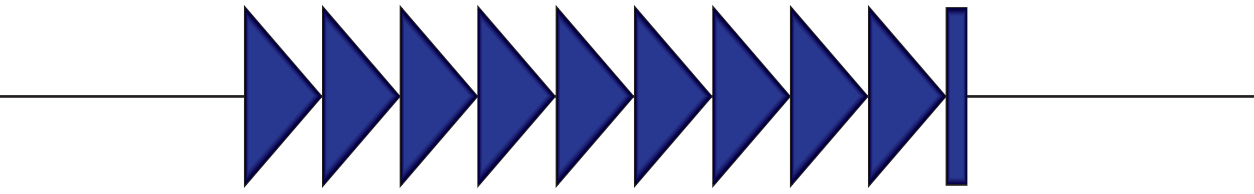
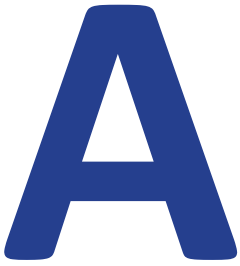
Genetics and epigenetics of repeat derepression in human disease

Thijssen, P.E.



Appendices



Summary

The repressive chromatin organization of repetitive DNA is pivotal for maintaining cell homeostasis. Defects in repeat silencing are a hallmark of the two genetic diseases studied in this thesis: FSHD and ICF syndrome. In FSHD, the D4Z4 macrosatellite repeat gets derepressed through an *in cis* or *in trans* mechanism, leading to sporadic expression of the toxic transcription factor *DUX4*. ICF is a severe immunodeficiency syndrome, characterized by derepression of (peri-) centromeric satellite repeats and is genetically heterogeneous. Mutations in multiple genes lead to a highly similar phenotype, which suggests these genes to act in common or converging pathways. The studies described in this thesis aimed at better understanding the genetic and epigenetic mechanisms underlying these diseases.

In **chapter 2** we have studied a possible correlation between D4Z4 chromatin compaction in somatic cells, measured by relative abundance of H3K9me3 over H3K4me2 at the D4Z4 repeat, and disease severity. Overall, FSHD derived somatic cells are characterized by a reduced chromatin compaction at D4Z4, which is especially pronounced in FSHD2 cells. Although trending in fibroblasts, we did not find a significant correlation between the chromatin compaction of D4Z4 and the age corrected clinical severity. Strikingly, the muscle pathology score in the *vastus lateralis* muscle showed a clear correlation with clinical severity score.

Telomere position effects (TPE) could affect the regulation of the subtelomeric D4Z4 repeat, but we could not investigate the 4q subtelomere directly due to its repetitive nature. Therefore, in **chapter 3** we have studied the chromatin regulation at two single copy subtelomeres (7q and 11q), as a model for subtelomeres in general, in response to telomere shortening and cellular senescence. Subtelomeres are characterized by a loss of H3K9me3 and CpG methylation upon telomere shortening and senescence. In contrast, we did not detect transcriptional activation of subtelomeric loci at the transcriptional and chromatin level. Silencing of the studied subtelomeres is most likely maintained through increased levels of markers for facultative heterochromatin, mediated by PRC2. Both subtelomeres under study, as well as D4Z4, showed similar regulation, with the exception of TERRA promoters. Overall, the 7q TERRA promoter showed similar regulation to the more proximal subtelomere, whereas the 11q promoter was distinct from its more proximal subtelomere, in concordance with the observed difference in TERRA transcription from 7q and 11q.

Having established the generic principles of subtelomeric chromatin structure in human cells, in **chapter 4** we focused on the chromatin regulation of D4Z4 by SMCHD1 in control and FSHD derived myotubes. This study established a pivotal role for *SMCHD1* in both genetic forms of the disease. In FSHD cells, the activation of *DUX4* during *in vitro* muscle cell differentiation coincides with a decreased level of SMCHD1 at D4Z4. More importantly, moderate overexpression of SMCHD1 resulted in silencing of DUX4 in both FSHD1 and FSHD2 cells. Ectopic inhibition of SMCHD1 expression in control cells leads to the accumulation of PRC2 and the PRC2 mediated H3K27me3 marker for facultative heterochromatin at D4Z4. Increased abundance of PRC2 and H3K27me3 at

D4Z4, and sensitivity to PRC2 inhibition, was also observed in FSHD2 patient derived myotubes, however not in FSHD1 myotubes. In conclusion, although SMCHD1 is able to silence DUX4 in both FSHD1 and FSHD2 derived myotubes, the epigenetic regulation of the D4Z4 repeat in both forms of the disease seem to be distinguished by aberrant Polycomb mediated chromatin regulation, which is only observed in FSHD2.

Over the past decades, the absence of evidence for sporadic *DUX4* activation resulted in numerous studies focusing on proximal gene dysregulation upon D4Z4 contraction as a potential disease mechanism. Models were presented in which D4Z4 could act as an insulator element to prevent TPE spreading in proximal direction, or in which the chromatin structure of D4Z4 itself could affect the transcriptional regulation of nearby genes through spreading or looping mechanisms. In **chapter 5** we show that the FSHD specific dysregulation of *FRG2*, the gene closest to the D4Z4 repeat array, is a direct consequence of DUX4 protein activity, rather than any of the above proposed models. This further strengthens the disease mechanism revolving around *DUX4* while at the same time giving a molecular explanation for a longstanding enigmatic observation of *FRG2* activation in FSHD patients.

In **chapter 6** we show that the epigenetic regulation of D4Z4 and its derepression upon contraction can be recapitulated in the mouse, independent from its subtelomeric location. We have generated a transgenic FSHD mouse model, carrying an FSHD1 sized D4Z4 repeat array of 2.5 units (D4Z4-2.5), and a transgenic “control” mouse carrying a 12.5 unit containing D4Z4 repeat (D4Z4-12.5). Our data shows that the D4Z4-2.5 mouse model, as compared to the D4Z4-12.5 mouse, recapitulates key aspects of FSHD: robust *DUX4* expression in the germ line, chromatin derepression of the repeat in somatic cells accompanied by sporadic *DUX4* activation in skeletal muscle and activation of (mouse specific) DUX4 target genes. Intriguingly, the D4Z4-2.5 mouse does not develop an overt muscle phenotype, suggesting that the consequences of DUX4 expression are not conserved between man and mouse. Nevertheless, our model can serve great purpose to assess therapeutic interventions aiming at inhibiting the mis-expression of *DUX4* in FSHD.

Finally, in **chapter 7** we describe the identification of two new ICF syndrome disease genes, *CDCA7* (ICF3) and *HELLS* (ICF4). Herewith we identify two novel chromatin modifiers of repetitive DNA structures in humans, and therefore potentially important factors for the regulation of D4Z4 and *DUX4*. For both newly identified genes, as well as for the ICF2 gene *ZBTB24*, we show a role in the maintenance of CpG methylation at centromeric repeats. The identification of two additional ICF disease genes renders new possibilities for mechanistic studies into the disease mechanism behind ICF syndrome, and to study the epigenetic regulation of (macrosatellite) repeats in general.

The progress made in understanding the (epi-)genetic mechanisms underlying FSHD and ICF syndrome offers new possibilities for translational and mechanistic studies focusing on chromatin organization of repetitive DNA. The common feature of repeat derepression in both diseases offers potential to study commonalities and differences of different mechanisms affecting repeat silencing during development and differentiation.