Modifying the modifier: discovering mechanisms of SMCHD1 mediated chromatin repression

Goossens, R.

Appendix

English summary
Nederlandse samenvatting
List of publications
Curriculum Vitae
Acknowledgements

English Summary

Epigenetic regulation of gene expression by chromatin modifiers is one of the fundamental cellular processes that allow the different cell types in the body to develop from the totipotent embryonic stem cells. While almost all cells have an identical set of genetic information, vastly diverse transcriptional programs need to be active to form different tissues. The different epigenetic marks that chromatin modifiers can deposit (write), interact with (read) or remove (erase) can either facilitate expression as seen in euchromatin, or prohibit expression as found in heterochromatin. A properly functioning chromatin environment allows cells to express the coding and non-coding RNAs necessary for its unique cellular identity and function, but in a flexible manner to adapt to changes in the environment as needed. However, when this epigenetic control mechanism becomes compromised, such as by mutations in these chromatin modifiers, it can lead to the development of disease. An example of such epigenetic disease is facioscapulohumeral muscular dystrophy (FSHD), in which the chromatin structure of the D4Z4 macrosatellite repeat is compromised. The loss of a repressive D4Z4 chromatin structure either by contraction of the repeat to a size of 1-10 D4Z4 units (FSHD1), or by mutations in D4Z4 chromatin repressors (FSHD2), results in inappropriate expression of the *DUX4* gene from the repeat in skeletal muscle, which is considered the root cause of FSHD.

FSHD type 2 is the rare form of FSHD (5%), and is most often caused by mutations in the chromatin modifier SMCHD1, and rarely by mutations in the chromatin modifiers DNMT3B or LRIF1. In FSHD2 patients, all D4Z4 repeats in the genome on chromosomes 4q and on 10q are epigenetically derepressed and have lost their high levels of DNA methylation. However, only when a mutation in either of these chromatin repressors is combined with a D4Z4 repeat of moderate size (8-20 units), this can lead to DUX4 expression and development of FSHD.

In either FSHD type the result is the same: chromatin relaxation of the D4Z4 repeat results in expression of the transcription factor DUX4 from the most distal D4Z4 unit, but only when this repeat ends with a somatic poly-adenylation signal for *DUX4* which is polymorphic in the population. DUX4 is a transcription factor involved in zygotic genome activation in cleavage stage embryos, and should not be expressed in muscle cells. In FSHD, DUX4 activates in myonuclei a range of target genes that are important for embryonic development, which disturbs the muscle cell program. The disturbed muscle homeostasis inflicted by DUX4 causes apoptosis, leading to muscle wasting in the patient. In this thesis, we studied the functionality of SMCHD1, and aimed to understand the DUX4 repressive processes in which SMCHD1 is involved. Furthermore, we gathered information on the different roles that SMCHD1 fulfills, such as X-chromosome inactivation in female cells and telomere maintenance.

In **chapter 2**, we describe how loss of one copy of *SMCHD1* in patients suffering from 18p-deletion syndrome can be a risk factor for developing FSHD. In these patients, the short arm of chromosome 18, where among other genes *SMCHD1* resides, is (partially) deleted in one of the chromosome copies. By only being able to express half the amount of SMCHD1 that healthy people do, these 18p- patients are haploinsufficient similar to FSHD2 patients with a mutation in *SMCHD1* which disrupts the open reading frame. We show that when 18p- patients also have an FSHD-permissive D4Z4 repeat (i.e. 8-20 copies and a *DUX4*

polyadenylation signal), 18p- patients indeed express DUX4 and exhibit the molecular and clinical hallmarks of FSHD.

In **chapter 3** we investigated the *SMCHD1* gene in two FSHD families where no mutations could be found in the coding region of *SMCHD1*. We found that in these families the mutation resides in intronic sequences of *SMCHD1*, respectively in introns 13 and 34. These variants introduce a non-canonical splice site and cause part of the intronic sequences to be incorporated into the mature transcript. These exonized fragments of the intronic sequence cause a frameshift of the reading frame and premature stop codons, prohibiting the production of SMCHD1 protein from these mutant alleles. We obtained muscle cells from the proband of the *SMCHD1* intron 34 mutation family, and used CRISPR/Cas9 genome editing technology to remove a part of the intron, including the pathogenic variant. This genome editing successfully restored SMCHD1 expression to near-endogenous levels, and effectively abolished expression of *DUX4*. Together, chapter 2 and chapter 3 consolidate the critical dose-dependent role of SMCHD1 in D4Z4 repression.

Proteins rarely function just by themselves; in general they act in complex with other proteins. These interacting proteins can regulate (part of) their activity, or direct them to sites of action. For SMCHD1, relatively little was known about the proteins it interacts with to fulfill its different roles. In **chapter 4**, we employ an advanced quantitative proteomics technique called SILAC Mass Spectrometry to study the interaction partners of SMCHD1 in a model cell line, U2OS. We identify a robust set of novel interacting proteins, and manage to individually validate the interaction of SMCHD1 and a selection of known chromatin modifiers. We specifically studied the interaction of SMCHD1 with RUVBL1, a protein with known ATPase and helicase activity, which has been shown previously to be present in D4Z4 repeat chromatin. We find that loss of RUVBL1 in FSHD muscle cells indeed leads to transcriptional upregulation of *DUX4*, but the exact multimeric complex in which it operates remains to be identified. Another interesting protein that interacts with SMCHD1 is EZHIP, a protein that is able to inhibit the activity the polycomb repressive complex 2 which is responsible for deposition of the repressive H3K27me3 mark. Although EZHIP likely plays no role in muscle cells, it might be involved in SMCHD1-dependent regulation of *DUX4* in early embryogenic development.

Finally, in **Chapter 5** we investigate the post-translational modification of SMCHD1 by SUMOylation. Post-translational modifications are small molecules that are added and removed from mature proteins, thereby e.g. regulating their activity, stability and subcellular localization. This gives cells a rapid method of controlling protein behavior and signaling. We show that SMCHD1 is highly SUMOylated on lysine 1374, but we find no clear direct molecular effect of inhibiting SMCHD1 SUMOylation. We also show that the SENP5 deSUMOylating enzyme can remove SUMO from SMCHD1 and observe that knocking down SENP5 in FSHD muscle cells inhibits *DUX4* expression. Interestingly, manipulating cellular SUMOylation by the small molecule SUMO inhibitor ML-792 or knock down of UBA2 results in *DUX4* expression in muscle cells. Finally, we performed a thorough investigation of the molecular characteristics of an SMCHD1 variant identified in an FSHD2 patient which results in the lack of the SUMOylation site K1374. While this SMCHD1 variant is pathogenic and the patient cells express DUX4, the main defect in SMCHD1 seems not to be related to the lack of the main SUMO acceptor site, but rather to result in SMCHD1 haploinsufficiency.