UNIVERSITA' DEGLI STUDI DI MILANO

PhD Course in Molecular and Cellular Biology

XXXII Cycle

Role of NF-YA isoforms in mouse Embryonic Stem Cells and Myoblasts differentiation

Debora Libetti

PhD Thesis

Scientific Tutor: Prof. Roberto Mantovani

Academic Year: 2016 – 2019

PART I

PART I

1. Abstract (English)

The Nuclear Factor Y (NF-Y) is a Transcription Factor (TF) composed by three different subunits, NF-YA, NF-YB and NF-YC, all necessary to recognize and bind the CCAAT box, a DNA-regulatory region highly enriched in active enhancers in human and mouse Embryonic Stem cells (ESCs).

NF-YB and NF-YC subunits have a histone-like domain (related to core histones H2B and H2A respectively) that associate forming a dimer required to interact with the NF-YA subunit. In particular, NF-YA has an evolutionary conserved domain in the C-terminal part responsible for NF-YB/NF-YC interaction and DNA binding, while the N-terminal domain is characterized by a glutamine rich (Q- rich) trans-activation domain. NF-YA comes in two isoforms, the long (NF-YAl) and the short (NF-YAs), differing for 28 amino acids coded by exon 3. Mouse ESCs preferentially express the short isoform, but after differentiation by Embryoid bodies (EBs) formation, a complete switch from the short to the long isoform is observed. Moreover, the progenitor muscle C2C12 cells express the long isoform, but after complete differentiation into muscle tissue, NF-YA protein is not detected.

The first project of my PhD thesis aims to shed light on the function of NF-YAs and NF-YAl in controlling cell differentiation by generating mES and C2C12 cell lines with a genomic deletion of exon 3 through the CRISPR/Cas9 Nickase system. Two isolated homozygous clones for each cell line were obtained and they expressed only the NF-YA short isoform (NF-YAl-KO), as expected.

Both cell lines, after exon 3 deletion, maintained the same morphology of wild type cells, but after differentiation stimuli, NF-YAl-KO clones and wild type cells showed different responses. In mESCs, NF-YAl-KO clones maintained the typical stem cell morphology with high levels of stemness genes and low levels of differentiation markers, compared to the wild type cells. In C2C12, after differentiation induction, wild type cells originated the classical myotubes, while the NF-YAl-KO clones maintained the myoblast identity. These results were supported by a low expression levels of muscle-associated genes in NF-YAl-KO clones, compared to wild type cells. These data confirm a different role of the two NF-YA isoforms in stemness maintenance and in particular suggest that the long one has a pivotal role during the differentiation process.

The second project has been inspired by the study of NF-YA protein 3D structure that highlighted the presence of two features typical of CPPs in the evolutionary conserved

domain. Previous studies have shown that the GST-TAT-NF-YA short fusion protein stimulated Hematopoietic Stem cells (HSCs) growth enhancing cell proliferation. Moreover, mESCs transfected with the GST-TAT-NF-YA short fusion protein maintained their pluripotency identity even after Leukemia Inhibitory Factor (LIF) withdrawal, which is necessary for mESCs to maintain stemness identity in culture media. The work presented in this thesis demonstrated the NF-YA capability to enter cells and to translocate into nuclei in a TAT-independent manner. Moreover, the differentiation induction of myoblast C2C12 cells after NF-YA short recombinant protein transfection inhibited myotubes formation demonstrating the functionality of the transduced recombinant protein.

1.1 Abstract (Italian)

Il Fattore Nucleare Y (NF-Y) è un Fattore Trascrizionale (TF) composto da tre diverse subunità, NF-YA, NF-YB e NF-YC, tutte necessarie per riconoscere e legare il CCAAT box, regione regolatrice del DNA altamente arricchita in Enhancer attivi di cellule staminali embrionali (ESC) sia di uomo che di topo.

Le subunità NF-YB e NF-YC hanno un dominio simile agli istoni H2B e H2A rispettivamente, e si associano formando un dimero necessario per interagire con la subunità NF-YA. In particolare, NF-YA ha un dominio evolutivamente conservato nella parte C-terminale responsabile dell'interazione con il dimero NF-YB/NF-YC e del legame al DNA, mentre il dominio N-terminale è caratterizzato da un dominio di trans-attivazione ricco in glutammine (Q-rich). La proteina di NF-YA è presente in due isoforme, la lunga (NF-YAl) e la corta (NF-YAs), che differiscono per 28 amminoacidi codificati dall'Esone 3. Le cellule staminali embrionali di topo esprimono preferenzialmente l'isoforma corta della proteina, ma dopo differenziamento mediante la formazione degli Embryoid Bodies (EBs), si osserva un completo cambio di espressione dall'isoforma corta a quella lunga. Inoltre, le cellule muscolari progenitrici C2C12 esprimono l'isoforma lunga della proteina, ma dopo un completo differenziamento in tessuto muscolare, la proteina di NF-YA non viene più espressa.

Il primo progetto della mia tesi di dottorato mira a far luce sulla funzione di NF-YAs e NF-YAl nel controllo del differenziamento cellulare. Per fare ciò, ho generato delle linee cellulari murine di ES e C2C12 con una delezione a livello genomico dell'Esone 3 attraverso il sistema

CRISPR/Cas9 Nickase. Ho isolato due cloni omozigoti per ciascuna linea cellulare che, come previsto, esprimono esclusivamente l'isoforma corta di NF-YA (NF-YAl-KO).

Entrambe le linee cellulari, dopo la delezione dell'Esone 3, mantengono la stessa morfologia delle cellule wild type (wt), ma dopo l'induzione del differenziamento, i cloni NF-YAl-KO mostrano risposte diverse rispetto alle cellule wt. Nel caso delle ESC, i cloni NF-YAl-KO mantengono la tipica morfologia delle cellule staminali con alti livelli di espressione di geni di staminalità e bassi livelli di marcatori di differenziamento, confrontate alle cellule wt. Nelle C2C12, dopo l'induzione del differenziamento, le cellule wt danno origine ai classici miotubi mentre i cloni NF-YAl-KO mantengono l'identità di mioblasti. Tali risultati sono confermati dal un basso livello di espressione dei geni associati al differenziamento muscolare nei cloni NF-YAl-KO rispetto alle cellule wt. Questi dati confermano un diverso ruolo delle due isoforme di NF-YA nel mantenimento della staminalità ed in particolare suggeriscono che l'isoforma lunga sembra avere un ruolo fondamentale durante il processo di differenziamento.

Il secondo progetto è stato ispirato dallo studio della struttura 3D della proteina di NF-YA che ha evidenziato la presenza di due caratteristiche tipiche dei Cell Penetrating Peptides (CPPs) nel dominio evolutivamente conservato. Studi precedenti hanno dimostrato che la proteina di fusione GST-TAT-NF-YAs stimola la crescita delle cellule staminali ematopoietiche (HSC) aumentando la proliferazione cellulare. Inoltre, le cellule staminali embrionali di topo (mESC) trasfettate con la proteina di fusione GST-TAT-NF-YAs possono mantenere la loro pluripotenza anche in assenza del Leukemia Inhibitory Factor (LIF), necessario nel terreno di coltura affinché le mESC mantengano la loro identità staminale. Il lavoro presentato in questa tesi ha dimostrato la capacità di NF-YA di entrare nelle cellule e traslocare nei nuclei in modo indipendente dalla proteina di fusione TAT. Inoltre, la trasfezione dei mioblasti C2C12 con la proteina ricombinante NF-YAs, in terreno che ne stimola il differenziamento, inibisce la formazione di miotubi dimostrando la funzionalità della proteina ricombinante una volta trasdotta.

2. Aims of the Thesis

The Nuclear Factor Y (NF-Y) is a pioneer transcription factor that specifically bind the CCAAT box regulating different genes networks such as those involved in cell cycle progression and proliferation, cell signaling, DNA repair, metabolism and stem pluripotency pathways.

The two projects presented in this PhD thesis contributed to shed light on new biological properties of NF-Y. In particular, it showed its intrinsic capacity to penetrate the cell membrane and the different roles played by the two NF-YA isoforms in stemness maintenance and during the differentiation process.

Cell-Penetrating Peptides (CPPs) are positively charged short peptides of about 5–30 amino acids long with the exclusive capability to penetrate into biological membrane and deliver a wide variety of cargos into cells. CPPs can be used for many diagnostic and therapeutic applications, such as the delivery of fluorescent or radioactive compounds for imaging, the carriage of peptides and proteins for therapeutic application, or the delivery of molecules into induced pluripotent stem cells for lineage specific differentiation [1]. Studies on NF-Y protein tridimensional structure have highlighted the presence of two features typical of CPPs: the presence of two α-helices (A1 and A2) and the abundance of Arginine [2]. These observations have laid the basis to verify if NF-YA can be catalogized into CPPs family, identifying the specific amino acids that confer CPP features and analyzing the functionality of the recombinant protein after cells transfections.

The second project have the objective to highlight the specific functions of the two NF-YA isoforms in the regulation of stemness and differentiation using two different cell lines: the mouse Embryonic Stem cells (ESCs), which are self-renewing Inner Cell Mass (ICM) derived cells with the capability to generate cells of all three germ layers, and the myoblast C2C12 cells, which are muscle cells capable to differentiate into myotubes.

To maintain the pluripotent state, mESCs are grown in vitro in the presence of Leukemia Inhibitory Factor (LIF), an external cytokine that activates the STAT3 signal, necessary to support the undifferentiated state. The involvement of NF-Y in early tissue development was initially demonstrated in Drosophila where the homologous of NF-Y controls the dorsal-ventral patterning. Subsequent studies showed that the deletion of the maternally deposited NF-YA protein in 2-cell mouse embryos drastically decreased DNase I-hypersensitive sites (DHSs), with the majority of embryos arresting to the morula stage [3]. In Hematopoietic Stem

Cells (HSCs), the overexpression of the short protein of NF-YA increased substantially their capacity to repopulate the bone marrow of immunocompromised animals [4]. The analysis of promoters directly bound by NF-Y revealed that it regulates different stem cell genes and more important it works as pioneer transcription factor promoting chromatin accessibility and the binding of specific TFs like SOX2 and OCT4, regulating the expression of key stemness genes like NANOG [5]. Analysis on NF-YA protein levels in mESCs revealed the preferential expression of the short isoform, but after differentiation into Embryoid Bodies, the NF-YA short levels decreased until undetectable levels while the long isoform increased becoming the only expressed isoform [6].

In C2C12 cells, the exclusive expressed isoform of NF-YA is the long one. Previous studies have shown that, after differentiation induction, the overexpression of the short NF-YA protein in C2C12 cells maintained high expression levels of proliferation markers with the inhibition of differentiation [7].

All these data suggest a crucial role of the two NF-YA isoforms in the regulation of stemness and differentiation. In order to determine if the short NF-YA isoform could maintain the stemness identity, mESCs and C2C12 cells were used to generate cell lines with the genomic deletion of the exon 3 by using the CRISPR-Cas9 system in order to check the differentiation capability of this cell lines which express only the short isoform of NF-YA.

3. Introduction

3.1 Transcription and Regulation

Transcription is the first step of gene expression in which a specific DNA sequence is replaced into a complementary RNA molecule. The transcription regulation is the basis of cellular life, controlling the development, proliferation, differentiation, aging and apoptosis of the cells.

Although the control of gene expression is more complex in eukaryotes than in bacteria, they have the same basic principles. As in bacteria, transcription in eukaryotic cells is controlled by proteins that recognize specific regulatory sequences and modulate the activity of the RNA polymerase.

The genomic information of eukaryotic cells is encoded by the DNA that is confined into the nucleus and packaged into chromatin, a complex of macromolecules formed by 146 bps (base-pairs) of duplex DNA (linker DNA) wrapped around the histone octamer composed by H2A, H2B, H3 and H4. This structure represents the basic unit of the chromatin named nucleosome. Nucleosomes are arranged as a linear array along the DNA polymer, which makes them appear as “beads on a string” by electron microscopy, with the string being formed by stretches of linker DNA in between consecutive nucleosomes, representing the primary packing level of the DNA [8].

A fifth histone protein, the H1, binds the linker DNA region between the histone proteins to stabilize the zig-zagged 30-nanometre chromatin fibre, the more compact second packing level.

Although its packaged conformation, chromatin accessibility is necessary to start different cellular process, like transcription and replication. Remodelling of nucleosomes increases the accessibility of DNA sequence elements to regulatory proteins that scan the genome for target sites. This process is controlled by different enzymes which can be divided into nucleosome remodelers and histone modifying enzymes.

Chromatin remodelling is an ATP-dependent process in which the nucleosome structure is disrupted making the DNA accessible to transcription factors or other protein complexes involved in DNA metabolism. The chromatin-remodelling complexes include two main classes of enzymes, the SWI/SNF- and the ISWI- containing complexes [9].

It has been reported that in mammalian cells the chromatin-remodelling complexes can play a crucial role in cancer development: mutations in the human SNF5/INI1 gene, which encodes for the SWI2/SNF2 complexes, contribute to oncogenesis [10].

Chromatin accessibility is also controlled by different covalent modifications of amino acids of the histones: acetylation, phosphorylation, methylation, ubiquitylation, sumoylation and ADP-ribosylation. All these modifications contribute to establish a condensed or decondensed chromatin state leading to transcriptional activation or repression [11].

For transcription initiation, eukaryotic cells require the assembly of RNA polymerase II (RNA pol II) and at least six General Transcription Factors (GTFs): TFIID, TFIIA, TFIIB, TFIIE, TFIIF and TFIIH. The GTFs assemble on the promoter in an ordered fashion to form a Pre-initiation Complex (PIC) [12]. In particular, TFIID, a multi-subunit complex composed by TATA-box binding proteins (TBPs) and at least 8–12 tightly bound subunits, the TBP-associated Factors (TAF), dictates the first step of PIC assembly, which is critical for the rate and efficiency of this process [13].

Transcriptional activity is greatly stimulated by the second class of factors: biochemical and genetic analysis of various model organisms has identified an astounding number of protein factors, known as Transcription Factors (TFs), which control the transcription by binding regulatory-DNA sequences positioned in proximity to the genes (promoters) or elements positioned at distance sites (enhancers) [14]. Transcription factors are composed by two characteristic domains: a DNA-binding domain which specifically recognize gene-regulatory elements, and a trans-activation domain, responsible for the recruitment of coactivators and the transcription machinery to initiate gene-specific transcription [15], [16], [17], [18].

In eukaryotic cells, *Ali H. Brivanlou et al.* [19] classified these key regulatory proteins into two groups on a functional basis: the “Constitutively Active Nuclear Factors”, which likely play a crucial role in the transcription of genes that seem to be always transcribed, and the “Regulatory Transcription Factors”, which indeed regulates the expression of specific genes like cell-type specific genes. In particular, the Constitutively Active Nuclear Factors group includes factors that bind the CCAAT box [20], one of the most widespread regulatory-sequence in the eukaryotic genomes [21].

3.2 The CCAAT-box

The regulation of gene expression is a fine mechanism that orchestrate the cellular life by the combined actions of multiple Transcription Factors, and by the packaging of DNA into chromatin and its modification by methylation. Several DNA sequences are important transcriptional regulators, positioned in proximity or at distance sites from genes promoters, and one of the first cis-acting DNA elements identified is the CCAAT box [14], [22]. It was identified in human, yeast and plant promoters playing an important role in transcriptional regulation.

Bucher P. performing a statistical analysis on over 500 unrelated promotes found that the pentanucleotide CCAAT sequence is present in ∼30% of them, confirming its pivotal role in gene expression regulation [23]. It is usually present between -60 and -100 bps from the Transcription Start Site (TSS) in the forward or reverse orientation [24] and it is extremely conserved, in terms of position and orientation, in different species, as well as the nucleotides flanking the central CCAAT pentanucleotide.

The CCAAT sequence can be found in both TAT-containing and in TATA-less promoters, but despite the canonical promoter representation, which contain both these two DNA regulatory elements, the vast majority of human promoter configurations demonstrate a negative correlation between CCAAT and TATA boxes: the number of double positive promoters is significantly lower than the number of single positive promoters. Moreover, CCAAT box correlates with CpG island promoters, well known areas of genomic regulation [25].

In higher eukaryotes, the pivotal role of this regulatory sequence is confirmed by its presence in developmentally controlled and tissue specific genes, housekeeping and inducible genes. In particular, gene involved in cell cycle progression and regulation, as CCNA, CCNB1, CCB2, CDC2, CDC25A, PLK1, are notably enriched in CCAAT boxes [26], [27]. Surprisingly, these genes usually include multiple CCAAT box in their promoters, lacking the presence of any TATA-box, as in the CCNB2 gene that contains three CCAAT boxes in its proximal promoter region [28].

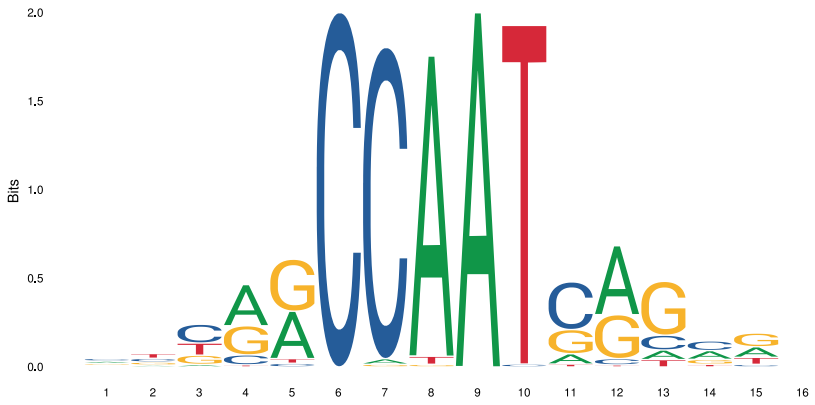


Figure 1. Sequence Logo of NF-Y consensus optimized with ChIP on chip data (g-CCAAT).

Many DNA-binding proteins have been identified as possible CCAAT-interacting polypeptides, as c/EBP (CCAAT/enhancer binding protein) whose binding sites are composed of palindromic repeats that occasionally contain a CCAAT pentanucleotide in the intervening sequence [29], or CTF/NF-1 (CCAAT transcription factor) that recognize a TGG$(N)_6$GCCAA sequence, but a T after the CCAA is not strictly present [30], and none of them shows a strict dependence on this DNA element (Fig. 1).

The only Transcription Factor known to specifically recognize the CCAAT sequence and the flanking nucleotides, both at the 5' and 3' ends, which help in the identification of the CCAAT regulator, is the Nuclear Factor Y (NF-Y) [24].

3.3 The Nuclear Factor Y

The Nuclear Factor Y (NF-Y) was first identified as a factor associated to the conserved Y-box promoter of the major histocompatibility complex class II (MHCII) genes [31]. Subsequently, many other eukaryotic promoters were discovered to be activated by this Transcription Factor.

NF-Y, also known as CBF (CCAAT-binding protein) is a trimeric transcription factor composed by three different subunits, NF-YA, NF-YB and NF-YC, all necessary to specifically bind the CCAAT box. An evolutionary conserved domain, called HAP for the homology with the yeast proteins, is required for CCAAT binding and subunit interaction. NF-YA contains the HAP

domain (HAP2) in the C-terminal portion of the protein, NF-YB in its central region (HAP3) and NF-YC in the N-terminal domain (HAP5) [24] (Fig. 2).

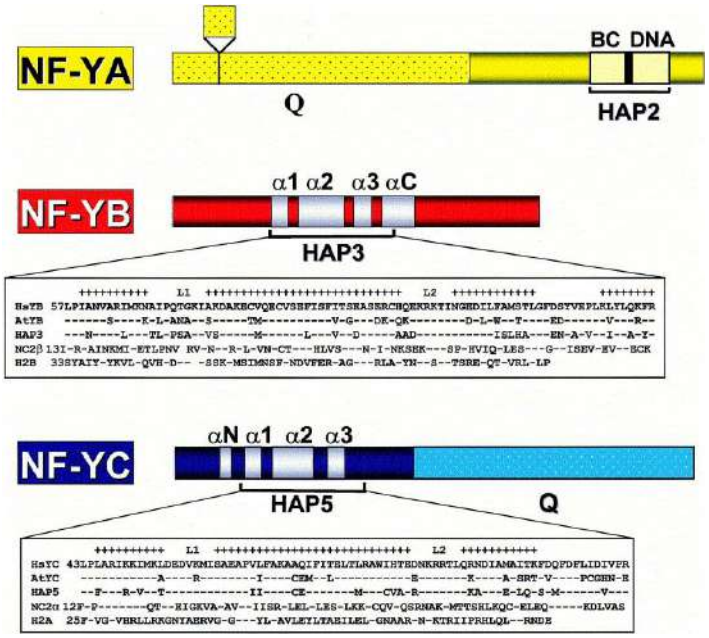


Figure 2. Schematic representation of the NF-Y genes. The yeast homology domains are indicated by brackets. In NF-YB and NF-YC are indicated the positions of the four α-helices of the histone fold domains (white boxes) [24].

NF-YA and NF-YC proteins are composed by a glutamine-rich (Q-rich) domain, at the C-terminus and N-terminus respectively, which is important for the transactivation activity, as assessed by in vitro transcription studies with recombinant proteins and in transfection of mammalian cells with LexA and GAL4 fusions [32], [33].

NF-YB and NF-YC subunits contain conserved histone-fold domains (HFD), related to core histones H2B and H2A respectively, composed by three α-helices separated by short strand regions that associate forming "histone-like" structural heterodimers. The NF-YB/YC tight association is a prerequisite to interact with the NF-YA subunit, which is the responsible for CCAAT recognition with high specificity and affinity [24].

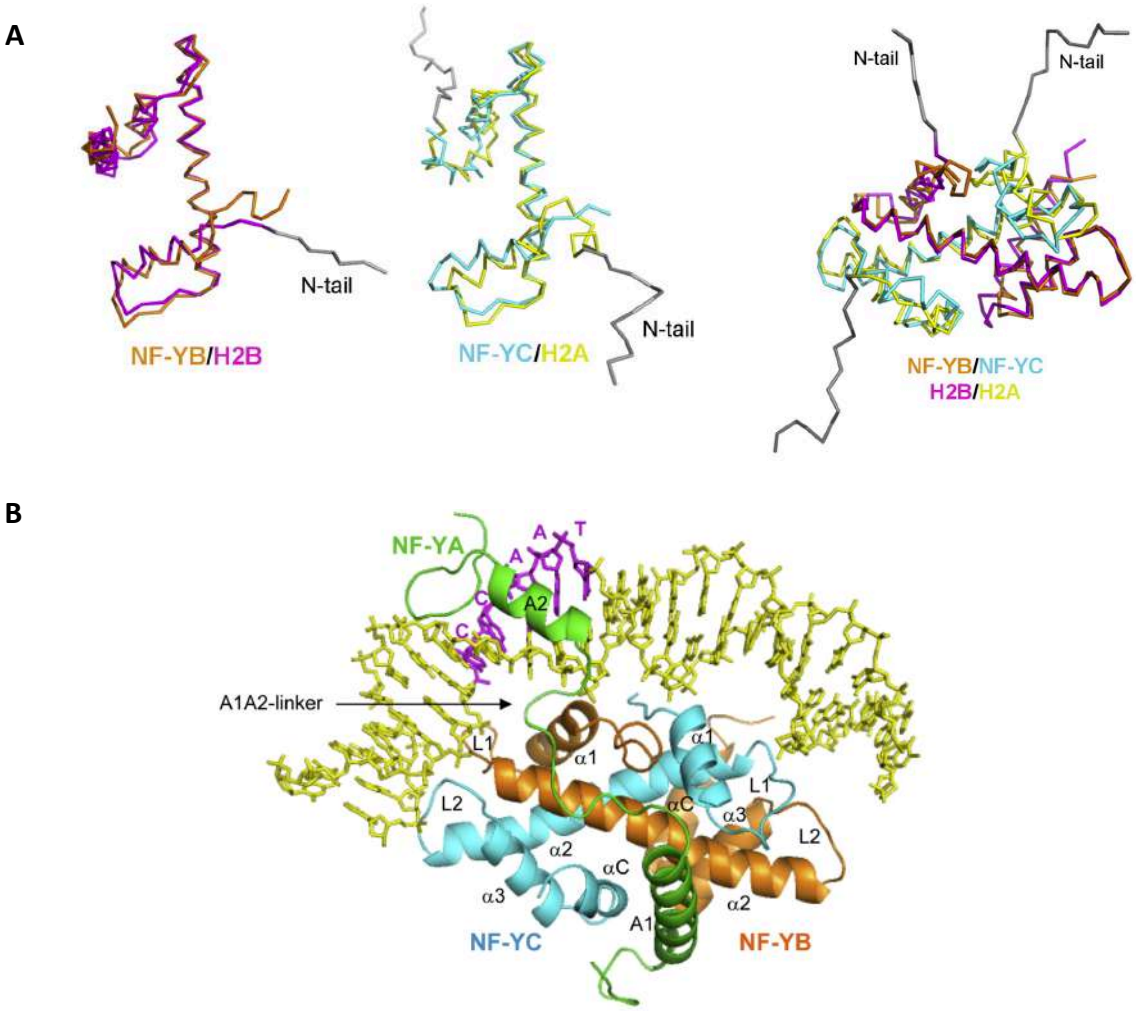


Figure 3. (A): Schematic representation of the single NF-YB and NF-YC subunits compared with the homologous region of the histones H2B and H2A (left panel) and the comparison of the two dimers [2]. **(B):** Representation of the NF-Y trimer bound to the CCAAT box resulting in a total DNA bending of 80° [34].

The NF-Y DNA-binding results in a global DNA bending of 80° that is peculiar of this transcription factor (Fig. 3). This spatial organization can reflect functional consequences on its target promoters. But the role of NF-Y in the regulation of transcription is not only due to NF-Y CCAAT binding: previous studies, both in vivo and in vitro, highlighted several functional and physical interactions with other transcription factors [35].

Maity and Crombrugghe [36] showed that subunits dimerization and its interaction with the NF-YA subunit (which dictates the DNA-binding) depends to the integrity of the histone-fold domain of NF-YB and NF-YC. Moreover, thanks to the conserved histone fold motif of NF-YB

and NF-YC, the NF-Y trimer can interact with other histone like proteins, as different TAFIIs subunits of TFIID. In particular, while TAF11, TAF12 and TAF13-TAF11 can interact with the histone fold domain of the NF-YB/YC dimer, TAF6 and TAF9-TAF6 interact with NF-Y trimer [37].

The general histone fold domain (HFD) structure is characterized by three alpha helices linked by two loops. NF-YB and NF-YC core regions contain an HFD carrying and additional α-helix (α-C) at the C-terminal region that is the responsible for the interaction with other TFs. Indeed, *Izumi et al.* [38] have shown that the α-C of NF-YC is crucial to establish physical interaction between NF-YC and other TFs like c-MYC and p53.

Unlike NF-YB and NF-YC, NF-YA subunit seems to have no homologous proteins. It presents only a homologous domain to the yeast protein HAP2 that is composed by two α-helices, the A1 and A2, linked by a linker loop. The A1 helix is mostly positively charged and contacts extensively the HFD dimer in its negatively charged center, while the A1/A2 linker optimizes electrostatic interactions between NF-YA and the NF-YB/YC dimer allowing the A2 helix to "search" for a CCAAT box, thanks to its flexible conformation.

3.4 NF-Y Alternative splicing

Alternative splicing is a process that allows the human genome to direct the synthesis of many more proteins than would be expected from its 20,000 protein-coding genes. During this process, a particular exon of a gene may be included within or excluded from the final processed messenger RNA (mRNA), producing translated proteins which contains differences in their amino acid sequence, influencing their biological functions.

NF-Y genes, like most of the mammalian genes, undergoes alternative splicing originating different isoforms of the same subunit, except for NF-YB gene that originates a single protein product of 32 kDa. Otherwise, NF-YC gene can originate many protein isoforms not only by alternative splicing but also by the activity of alternative promoters, the other system used by eukaryotic cells to produce different proteins starting from the same DNA sequence.

The NF-YC gene has two alternative promoters, P1 and P2. The P1 promoter represents the principal promoter of NF-YC with housekeep activity, while the P2 promoter is activated following DNA damage in a p53 dependent pathway.

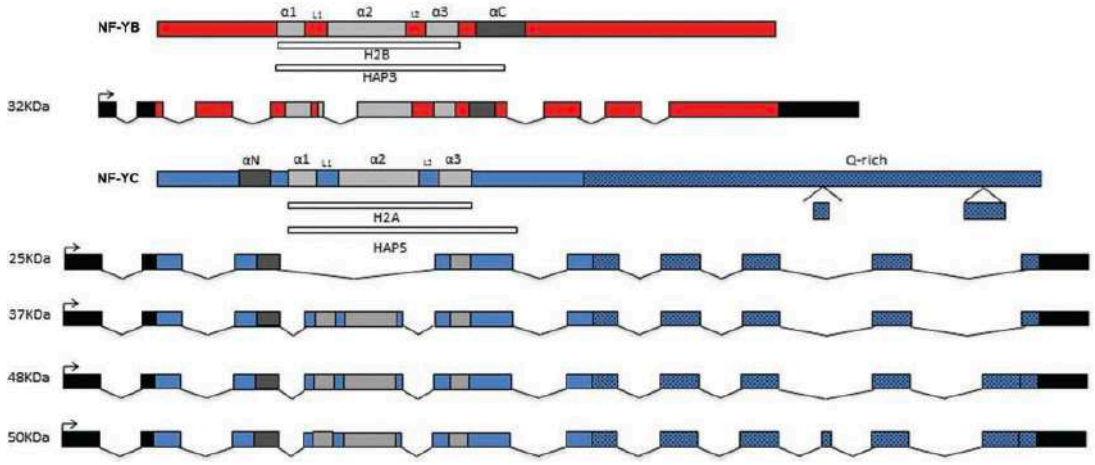


Figure 4. Structure of NF-YB and NF-YC subunits. Functional domains and differential splicing are reported [39].

The exon 8 and exon 9 of NF-YC codified a region of the Q-rich transactivation domain and they are involved by an alternative splicing that originate three different protein products of 37, 48 and 50 kDa. *Ceribelli et al.* [40] highlighted that different cell lines transcribed the same mRNA levels of the three NF-YC isoforms, but analyzing the protein expression level this doesn't correlate with the mRNA level: the relative abundance of each protein isoforms is cell-dependent and, in particular, the 37 and 50 kDa isoforms are mutually exclusive. These data indicate that a strong post-transcriptional mechanism regulate the mRNA stability and the consequent protein translation.

The NF-YA gene is involved by an alternative splicing of the exon 3 that codified a part of the long Q-rich transactivation domain. This exon-skipping event leads to the formation of two protein isoforms, the long (NF-YAl) and the short (NF-YAs), which lacks the 28 amino acids codified by the exon 3, and different cell lines preferentially express the long or the short protein isoform of NF-YA.

Importantly, it has been demonstrated a strong correlation between NF-YA and NF-YC isoforms: cells harboring the short isoform of NF-YA preferentially express the 50 kDa NF-YC isoform while the long isoform of NF-YA is enriched in cells expressing the 37 kDa NF-YC [40]. These data suggest that different cell contests harbor different NF-Y complexes, but how and whether they activate different set of genes is still unclear.

3.5 The NF-YA subunit

The NF-YA subunit represents the regulatory subunit of the trimer dictating the CCAAT-binding specificity. The N-terminus of the protein is composed by a Q-rich transactivation domain while the N-terminus is characterized by an evolutionary conserved domain, which is composed by two alpha helices, the A1, responsible for the interaction with NF-YB/YC dimer, and the A2, involved in DNA-binding and CCAAT recognition.

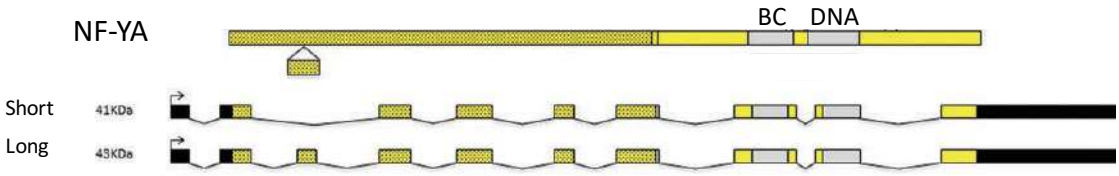


Figure 5. Schematic representation of NF-YA gene and the two isoforms derived from the alternative splicing of the exon 3. The two α-helices responsible for NF-YB/YC interaction and DNA-binding are reported (grey boxes) [39].

While NF-YB and NF-YC proteins are ubiquitously expressed and protein levels are relatively constant during the cell cycle, NF-YA levels fluctuate: in Murine Erythroleukemia (MEL) cells, the NF-YA protein level is maximal during the S phase while decreased in the G2/M phase [28]. The amount of NF-YA protein can be modulated by specific stimuli and cell treatments: while deprivation of serum decreases the expression level of NF-YA protein level in human IMR-90 fibroblasts [41], MMS treatment leads to its increase [42]. Finally, although both immortalized and not transformed cell lines express all the three NF-Y subunits, it was found that post-mitotic normal cells, differentiated myotubes and circulating monocytes do not express NF-YA [43], [44]. These data identified NF-YA as the limiting subunit of the NF-Y trimer.

Differently from the other NF-Y subunits, NF-YA protein levels are regulated by the ubiquitin/proteasome pathway. The estimated half-life of the protein is about two hours in proliferating cells, but after proteasome inhibition, the stability of NF-YA protein increases. Analysis on protein mutations identified four lysine residues in the C-terminal domain, the K283, K289, K292 and K296, as important target of ubiquitination. Moreover, it was observed that two of these lysine residues can be acetylated by the p300 histone acetyl-transferase

[45], highlighting that the competition between acetylation and ubiquitylation on the same lysine residues is a mechanism to regulate NF-YA protein stability.

The two NF-YA protein isoforms are differently expressed in different cell lines: we can observe cell lines that express the two isoforms at same levels, as the PC3 cells, and cells that preferentially express the long isoform, as the U2OS cells, or the short one, like the mouse Embryonic Stem cells (ESCs) [40], [39]. Moreover, *Grskovic et al.* [6] have shown that, during ESCs differentiation, the expression of NF-YAl and NF-YAs changed in opposite directions: the levels of NF-YAl increased with ESCs differentiation by Retinoic Acid (RA) treatment or Embryoid Bodies (EBs) formation, while NF-YAs levels were significantly downregulated with differentiation, suggesting a different role of the NF-YA protein isoforms.

3.6 The Pioneer Transcription Factor NF-Y

Transcription Factors are proteins that detect and bind regulatory sequence in the DNA and target the assembly of protein complexes to control gene expression. Genome-wide location analysis have shown that Transcription Factors occupy a few numbers of their potential DNA-binding sites in the genome [46], [47], indicating that chromatin structure beyond the DNA sequence dictates the binding of Transcription Factors to their target sites.

For many kinds of Factors, cooperative interactions with other Factors can allow nucleosome binding [48], but when the binding of a series of Factors is sequential in time and thus not initially cooperative, it is necessary that a special “Pioneer Transcription Factor” first bind the target site.

Pioneer Transcription Factors are proteins with nucleosome-binding properties that distinguish them from other DNA-binding factors. These Factors can actively help to initiate the assembly of other regulatory factors on the DNA by different strategy; they can open the chromatin locally, position the nucleosomes, enable intrinsic cooperative binding effects among other DNA-binding factors, or they directly recruit other chromatin modifiers and co-regulators. Pioneer factors can have also the passive role to be engaged at a regulatory sequence to require fewer subsequent factors to bind for regulatory activity.

The NF-Y crystal structure reveal that, while the NF-YA subunit dictates the DNA-sequence specificity, the NF-YB/YC dimer interact extensively with the DNA phosphates, in a way that is

extremely similar to H2A/H2B-DNA assembly in the nucleosome [34]. These data suggest a Pioneer role of NF-Y in enhancing chromatin accessibility. Indeed, NF-Y has been found to work as Pioneer Transcription Factor increasing chromatin accessibility and subsequent promoters' activation [49], [50], [5], [3]. In particular, *Falong Lu et al.* [3] depleted from a 2-cell stage mouse embryo the maternal NF-YA protein to analyse the genome activation: they discovered that it is required for the activation of ∼15% of 2-cell activated genes, suggesting that NF-YA is one of the Transcription Factors responsible for Zygotic Genome Activation (ZGA) and introducing the crucial role of NF-Y in stem cells development.

3.7 Mouse Embryonic Stem Cells

Mouse Embryonic Stem Cells (mESCs) are cells derived from the Inner Cell Mass (ICM) of the developing blastocyst. They have the capacity to generate all cell lineages of the developing and adult organism, known as pluripotency, and self-renewing, which means that they can proliferate in the same state.

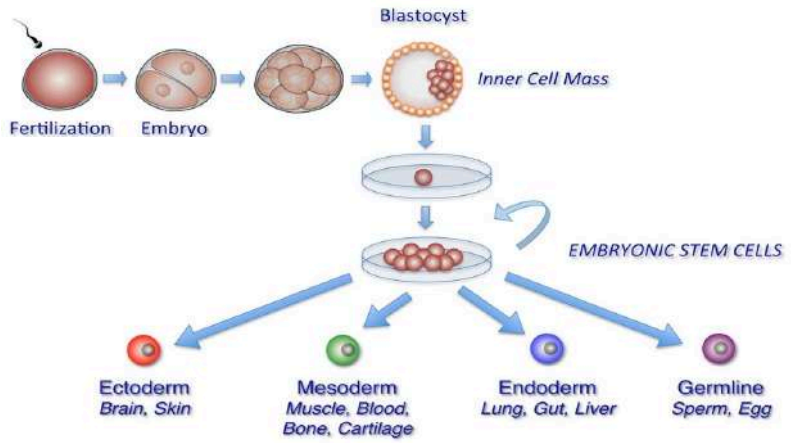


Figure 6. Mouse Embryonic Stem cells derived from the Inner Cell Mass forms cells of all three germ layers.

These characteristic of mESCs are controlled by external and internal signals that maintain cells in the pluripotent state. *In vitro*, Leukemia Inhibitory Factor (LIF) is an external cytokine that supports the undifferentiated state of mESCs by activating STAT3 signal [51], which

directly binds Oct4 and Nanog distal enhancers, modulating their expression to maintain pluripotency of mESCs [52].

Indeed, pluripotency is controlled by TFs that regulate a specific gene expression program by gene activation, establishing a poised state for gene activation in response to developmental stimuli, or contributing by gene silencing. In mammals, TFs are the largest class of proteins representing about 10% of all protein-coding genes, but only a couple of them seems to regulate mESCs by working with other TFs, some of which come from the group of terminal components of developmental signalling pathway.

The pluripotent state is largely controlled by the "core promoter" TFs Oct4, Sox2 and Nanog, and in particular, Oct4 and Nanog have been discovered to be essential for establishing and maintaining a robust pluripotent state being expressed only in stem cells [53], [54]. In mESCs Oct4 bind Sox2 forming a heterodimer that place Sox2 among the key regulators [55], [56].

Somatic cells can be reprogrammed into induced pluripotent stem (iPS) cells by the overexpression of Oct4, Sox2, Klf4 and c-Myc [57]. But if Nanog is not necessary for the reprogramming process, it is essential to promote a stable undifferentiated state [58], and more important it co-occupies most of the sites bound by Oct4/Sox2 heterodimer throughout the ESC genome, many of which are developmentally important homeodomain proteins, suggesting a role of these TFs in repressing these gene to maintain a stable pluripotent state and to undergo normal differentiation [59], [60]. The "core promoter" interacts with coactivators to recruit the RNA polymerase II (RNA-pol II, the responsible for DNA transcription into mRNA in mammalian cells) [61], while c-Myc binds the E-box sequences at core promoter sites recruiting p-TEFb, which phosphorylate the RNA-pol II and its associated pause control factors allowing the release and the fully transcription of the gene [62].

3.7.1 Naïve and primed pluripotent Stem Cell states

Between embryonic day 4 and 5 (preimplantation embryo) the mouse ICM segregates in two compartments: the epiblast and the primitive endoderm in which the epiblast develops to the embryo while the primitive endoderm gives rise to extraembryonic yolk sac tissue. Epiblast cells derived at this transitory stage are considered naïve epiblast and can be propagated *ex vivo* as cell lines that are called mouse Embryonic Stem Cells. Mouse ESCs derived from the

ICM of a preimplantation embryo have an unlimited self-renewal capacity when cultured in vitro under appropriate condition and they can differentiate into tissues of all three germ layers. Moreover, naïve Stem Cells have the intrinsic capacity to contribute to all somatic lineages, including the germline, when injected back into the early embryo, generating a chimera (a single organism that is composed of two or more different populations of genetically distinct cells), demonstrating their pluripotency also *in vivo*.

Mouse Epiblast Stem Cells (mEpiSCs) instead are derived from the epiblast of a post-implantation embryo and they are considered *primed* pluripotent Stem Cells. In this category are included also the human ESCs and iPSCs, although they are derived from a preimplantation embryo. Primed Stem Cells, like the naïve, are capable of unlimited self-renewal and can differentiate into all tissues of all three germ layers *in vitro*, but they are limited in their pluripotency *in vivo* as they cannot give rise to chimeras once injected into an early embryo [63].

Naïve Stem Cells appears with the classical Stem Cell morphology, they are small, compact with a round shape, while the primed are large and grow as a monolayer. Moreover, the naïve Stem Cells survive better to passage as single cells and duplicate in a shorter time [64].

Mouse ESCs can be maintained in the naïve state by culturing in the presence of serum plus LIF, but under serum withdrawal, LIF is not sufficient to prevent the differentiation. This limitation is overcome by using the "2i medium", composed by LIF plus two specific inhibitors: a specific inhibitor of the mitogen-activated protein kinase (ERK1/2) and an inhibitor of the glycogen synthase kinase (GSK3) [65].

The expression profiles of naïve and primed Stem Cells are different: while the pluripotency factors Oct4 and Sox2 are equally expressed, factors like Nanog, Sall4, KLF4 or Esrrb are preferentially expressed or up-regulated only in naïve Stem Cells. Moreover, the Tfe3 protein in naïve Stem Cells is localized into the nucleus, but after the conversion into primed cells, it becomes cytoplasmic with the result downregulation of Esrrb expression [66].

Mouse ESCs can differentiate *in vitro* into EpiSCs by 2i/LIF withdrawal and the addition of activin and Fgf2 [67]. During the differentiation, transitional cells were identified and named epiblast-like cells (EpiLCs). These cells are transient and form from naïve Stem Cells after 48 hours of differentiation induction. EpiLCs do not exhibit naïve pluripotency gene expression with a rapid reduction of about 70% only after 4 hours. Transcriptome analysis revealed high similarity with the preimplantation epiblast while consistent differences with the post-

implantation epiblast: EpiLCs express Fgf5, Otx2 and Oct6 along with Oct4 and Sox2 (typical of a preimplantation epiblast) but not later lineage markers [68].

Interestingly, Oct4 and TFs like Oxt2, induced after differentiation induction, cooperatively regulate the expression of new genes during the transition [68], [69]. Indeed, a link between Oct4 and Otx2 was observed. In the naïve state, Oct4 interacts with Tcf3, Esrrb and Klf4, but after differentiation into EpiLCs, it was found to physically interact with Otx2, Zic2/3 and other TFs [68], and this is explained by the protein levels of the interactors: in the naïve state stemness markers like Esrrb are upregulated while after differentiation their expression decrease with the concomitant expression of early differentiation genes like Otx2, meaning that the levels of gene expression dictates cell identity. In addition, the overexpression of naïve-associated factors like KLF4 can convert the primed Cells into naïve Cells suggesting their importance for the maintenance of the naïve state [67].

3.7.2 Role of NF-YA in pluripotency maintenance

Analysis on mESCs transcriptome to highlight important pathways for stemness identity has identified the CCAAT box as one of the most enriched Transcription Factor Binding Site (TFBS) in active enhancers [6], suggesting a pivotal role of NF-Y in pluripotency regulation.

The involvement of NF-Y in Embryonic development was demonstrated by analysing the NF-YA heterozygous mice that were fertile and normal, while the NF-YA null mice die prior 8.5 d.p.c, suggesting that NF-Y plays a crucial role in early mouse Embryonic development [70].

The involvement of NF-Y in early tissue development was initially suggested by *Stathopoulos et al.* who demonstrated that the homologous of NF-Y controls the dorsal-ventral patterning in Drosophila [71]. In Hematopoietic Stem Cells (HSCs) NF-Y was found as master regulator of self-renewal and differentiation: the retroviral infection of HSCs to overexpress the NF-YAs protein lead to an increase in expression of several HOX genes like HOXB4, a master gene for HSCs, and TFs mediating Notch and Wnt signaling, notoriously crucial in HSCs maintenance. The overexpression of NF-YAs in mouse HSCs increased substantially their capacity to repopulate the bone marrow of immunocompromised animals [4].

Domashenko et al. [72] used an alternative strategy to overexpress the NF-YAs protein: HSCs were transfected using the NF-YAs protein fused to a TAT peptide, a Cell Penetrating Peptide

(CPP) that allows to the fused protein to enter the cell membrane. By adding the recombinant TAT-NF-YAs to the medium, the HSCs were efficiently transfected and it promotes the *ex vivo* proliferation and increases the engraftment of human hematopoietic progenitor cells, suggesting an important role of the NF-YA short isoform in pluripotency regulation.

The analysis of NF-YA mRNA expression levels by Real-Time Reverse Transcriptase PCR (RT-PCR) in mESCs and differentiated mESCs, either treated with RA or induced to form EBs, revealed that the expression of the two isoforms of NF-YA (long and short) changed in opposite directions: mESCs preferentially express the short isoform, but after differentiation the NF-YAs levels significantly decreased while the NF-YAl levels increased [6]. Moreover, TAT-NF-YAs protein transfection of mESCs activated stemness genes containing CCAAT boxes stimulating cell proliferation and maintaining the cells in a pluripotent state after LIF withdrawal [39].

Chromatin Immunoprecipitation (Chip) analysis revealed that NF-Y directly binds to the promoters of genes upregulated in pluripotent cells, as *Sall4* and *Zic3* [6], but more important, bioinformatic analysis highlighted a strong correlation between NF-Y and NANOG with about 30% of Nanog peaks overlapping with NF-Y sites, whereas Nanog promoter is not a NF-Y target but it is regulated by OCT4 and SOX2, suggesting an indirect correlation between NF-Y and NANOG [39]. Indeed, *Oldfield et al.* [5] demonstrated that NF-Y works as pioneer TF promoting chromatin accessibility and OCT4/SOX2 DNA-binding.

Analysing the DHS (DNase I-hypersensitive site) formation in 2-cell embryos, and by depleting the maternally deposited NF-YA protein, about 30% of 2-cell DHS were decreased (enriched in promoter regions), with higher reduction in NF-YA motif-containing promoter DHSs, with the majority of embryos arresting to the morula stage [3].

Taking together, these findings define NF-Y as an essential factor for ESCs pluripotent state regulation.

3.8 Role of NF-YA in muscle differentiation

Skeletal myogenesis is a complex regulated mechanism important for regenerative medicine and tissue repair. To discover the factors that control cellular growth and differentiation, a muscle precursor cell line, the C2C12 cells, is commonly used thanks to its ability to

differentiate into myoblast, then myocyte and finally a multinucleated myotube. This feature enables to discover the transcriptional program that regulates each steps of the differentiation process.

C2C12 cells can easily differentiate into myotubes by changing the culture media from high serum (10%) to low serum (2%), known as *serum starvation*. These precursor cells undergo differentiation expressing muscle-associated genes. The most important regulatory factors during skeletal myogenesis are represented by the Myogenic Regulatory Factors (MRFs), which are composed by MyoD1, Myf5, Myogenin and Mrf4 [73]. This family of TFs binds the E-box, a motif found in promoters of many muscle-specific genes [74], and are differentially expressed during the myogenic pathway: Myf5 is the first MRF expressed during mouse embryonic development, like MyoD1, while Myogenin and Mrf4 are expressed at later stages [75]. Mice devoid of Myf5 show a delay in embryonic myogenesis that instead proceed normally after MyoD1 induction [76]. MyoD1 Knock-out (KO) mice compensate with an increase in the expression level and stability of Myf5 [77], revealing a redundancy of Myf5 and MyoD1 during myogenesis. The analysis of Myogenin KO mice revealed a reduced expression of muscle-genes like Myosine Heavy chain (MYH) and Mrf4, while MyoD1 levels were normal [78]. These data revel that Myf5 and MyoD1 act genetically upstream of Myogenin and Mrf4 to specify myoblast for terminal differentiation, while Myogenin and Mrf4 are more involved in the differentiation process triggering the expression of myotube-specific genes.

The MRFs family cooperate with other TFs, in particular with the myocyte enhancer factor-2 (Mef2), to upregulate specific target genes [79], [80]. The Mef2 family cannot induce myogenesis in transfected fibroblasts, but when co-expressed with the myogenic proteins MyoD1 or Myogenin they dramatically increase the extent of myogenic conversion [81].

Analysing the NF-Y protein levels in different murine adult tissues, Gurtner *et al.* [44] showed that while NF-YB and NF-YC subunits are expressed, the NF-YA subunit is not detectable in skeletal muscle and heart with the consequent loss of NF-Y binding to the CCAAT boxes. Moreover, NF-YA overexpression in C2C12 cells led to a delay in the induction of Myogenin whereas there was no effect on MyoD1 expression, revealing that the down-regulation of NF-YA expression is strictly required for the correct modulation of muscle-specific gene expression during differentiation of muscle cells. Chip analysis revealed that NF-Y is not present on Myogenin promoter, while directly binds the promoter of Mef2D gene, known to activate MyoD1 and Myogenin [7]. The stable overexpression of NF-YAl and NF-YAs in C2C12

clones have different effects on differentiation: while stable NF-YAs-overexpressing cells revealed high expression levels of proliferation markers with the inhibition of differentiation, NF-YAl-overexpressing cells showed a downregulation of cell cycle genes with the enhanced expression of muscle-specific genes like Mef2D, Six1, and Six4 [7].

Another study revealed that the forced expression of NF-YAs protein in C2C12 cells after differentiation induction maintain high expression levels of cyclin A, cyclin B1 and cdk1 with the consequent delay in Myogenin expression [44].

These data confirmed different roles of the two NF-YA protein isoforms suggesting the importance of NF-YAs function suppression to inhibit several cell cycle genes and to initiate the specific differentiation program in myoblast C2C12 cells.

3.9 Cell Penetrating Peptides (CPPs)

The use of recombinant proteins in medical practice starts in the 1980 with the introduction of the recombinant human insulin to treat the diabetes mellitus. The use of these molecules has several advantages with a high specificity and tolerability, but their hydrophilic nature limits their therapeutic value because of their low membrane permeability, preventing their access to intracellular targets.

In 1988, the discovery of the trans-activator of transcription (TAT) protein of HIV-1 as a protein capable to penetrate the cell membrane provides the basis for a new strategy that use peptides, known as Cell-Penetrating Peptides (CPPs) as carriers for intracellular uptake [82].

CPPs are positively charged short peptides with 5-30 amino acids long with the ability to penetrate the cell membrane. They have high transfection efficiency and low toxicity, making them useful for diagnostic and therapeutic applications.

The first CPP identified was the TAT peptide, an RNA-binding protein of 14 kDa that stimulates the HIV-1 gene expression by enhancing the processivity or RNA pol II [83]. The TAT protein is composed by a transcriptional activation domain enriched in cysteine and a hydrophobic arginine-rich motif responsible for RNA-binding. By studying deleted mutants in the N- and C-terminus, *Park et al.* [84] defined the minimal domain required for cell translocation, from the amino acid 49 to 57, which is rich in basic amino acids like lysine and arginine residues.

The attachment of cargo molecules to CPPs enables to transport into cells a wide range of conjugated cargoes like peptides, DNA, short interfering RNA (siRNA) and small drugs [85]. Cargoes can be conjugated by using two different strategies: covalent or non-covalent binding. The covalent conjugation binds the two molecules in a time-consuming process, while in the non-covalent conjugation the two molecules interact by electrostatic interactions [86].
CPPs can be classified based on their origin or their physiochemical properties. Based on the origin, CPPs are divided into chimeric (composed of two or more motifs derived from different peptides), protein-derived or synthetic [85]. Based on their physiochemical properties, CPPs are categorised into Cationic, amphipathic and hydrophobic (less used compared to cationic and amphipathic CPPs) [87].

3.9.1 Internalisation mechanisms of CPPs

Although CPPs are commonly used to transport molecules into cells, the exact mechanism used to penetrate cellular membrane is still obscure. Three main pathways are proposed for the CPP internalisation: the first is based on *direct penetration*, the second is the *endocytosis*, and the third that use the formation of a *transitory structure*.
The internalisation through *direct penetration* is an energy-independent pathway in which the positively charges of the CPP firstly interact with the negatively charged membrane components, leading to a destabilisation of the membrane and the folding of the CPP on the lipid membrane [88].
The *endocytosis* is a natural and energy-dependent pathway occurring in all cells. Thanks to this mechanism, cells can absorb materials from the outside through the plasma membrane that folds inward to carry external elements inside. The endocytosis can be divided in three main classes: pinocytosis (large molecules absorption), phagocytosis (solutes absorption) and receptor-mediated endocytosis. In the receptor-mediated endocytosis, clathrin and caveolin proteins play a key role for the invagination of the membrane and the construction of vesicles: clathrin-coated vesicles have a smaller diameter compared to vesicles coated by caveolin and are more selective, indeed generally the translocation is made through binding the material to specific receptor on cell surface [89].

The formation of a *transitory structure* is based on the construction of an inverted micelle in which at first the CPP is trapped into the hydrophilic environment of the micelle, then the micelle interacts with the membrane components destabilising itself and leading to the release of the CPPs into the cytoplasm [90], [91].

3.9.2 CPPs applications

The presence of cell and tissue barriers prevents many therapeutic drugs to reach their specific intracellular targets. By conjugating drugs with CPPs ca be enhanced their cellular internalization and therapeutic efficacy.

The CPPs-mediated transport of small and large molecules was achieved both *in vivo* and *in vitro* to treat several diseases, especially in cancer. Since traditional chemotherapy usually lacks specificity and may result in toxicity, CPPs appears to be a best strategy to treat oncological pathologies. In particular, the p53 protein induce cell cycle arrest and apoptosis and loss of function of p53 results in malignant advancement [92]. To restore the activity of p53, *Synder et al.* [93] produced a transducible and proteolytically stable peptide resulting in a significant increase in lifespan and generation of disease-free animals [93].

The human papillomavirus (HPV) E2 protein is a DNA-binding protein and the overexpression can induce growth arrest and/or apoptotic cell death in cervical cancer cells. The herpes simplex virus VP22 protein was fused to the E2 protein to induce apoptosis in transiently transfected HPV-transformed cervical carcinoma cell lines [94].

In other cases, CPPs were used to transport enzymes into the cells to make a genomic deletion [95] or to vehicle inhibitors of NF-kB across cell membrane [96].

TAT peptide was also used to transduce NF-YA protein into the human hematopoietic cells to induce the activation of HOXB4 and the consequent proliferation of primitive hematopoietic progenitor cells [72], showing that TAT-peptide therapies can be used as an alternative approach to retroviral Stem Cell therapies and suggesting that the NF-YA protein can be evaluated as a tool for primitive hematopoietic progenitor cells *ex vivo* expansion and therapy.

The use of CPPs were applicated also to target bacteria and fungi inhibiting their growth. This system is of great interest because by binding an antimicrobial agent to a CPPs is possible to overcome the growing problem of bacterial antibiotic resistance.

Another application of CPPs is to transport nucleic acids, which are unable to over-cross the cellular membrane because of them hydrophilic nature. The use of CPPs improves the transfection of nucleic acids exceeding problems derived from classical transfection techniques like viral infection which have high toxicity and low efficiency transfection.

Although CPPs have the advantage to bypass physical cell barriers, the main limitation is the low cell specificity. This limitation can be overpass by binding to the CPP a specific molecule with high affinity for a specific target. *Fang et al.* [97] conjugated the TAT protein with a short A1 peptide, 6 amino acids long, which specifically recognise the vascular endothelial growth factor receptor-1 (VEGFR1). The TAT-A1 selectively penetrates into tumor cells that over-expressed VEGFR1. Moreover, the system successfully transports into cells the siRNA used for the anti-tumor therapy [97].

All these works confirm CPPs as a useful strategy that can be used for many applications by reducing toxicity and enhancing transfection efficiency.

4. Results and Discussion

The transcription factor NF-Y is a trimeric factor composed by the NF-YA, NF-YB and NF-YC subunits which are all necessary to specifically bind the CCAAT box and regulate the transcription of target genes. The NF-YA gene is involved by an alternative splicing that originates two protein isoforms, the long (NF-YAl) and the short (NF-YAs) which lacks the 28 amino acids codified by the exon 3.

Previous studies have shown the pivotal role of NF-YA, and in particular the short isoform, in stem cell marker regulation already at 2-cell stage of mouse development [5], [3]. More interesting, after differentiation into Embryoid Bodies formation, the expression of NF-YAs decreases until a complete switch to the long NF-YA protein [39], suggesting different roles of the two NF-YA protein isoforms in stemness maintenance and during the differentiation process. Moreover, mouse myoblast cells (C2C12 cells), which express only the long NF-YA protein isoform, after protein transfection of NF-YAs completely lost the capability to differentiate into myotubes [98].

To shed light on the specific functions of NF-YAs and NF-YAl proteins in the regulation of self-renewal and differentiation, mouse Embryonic Stem and C2C12 cell lines with the genomic deletion of the exon 3 were generated by using the CRISPR/Cas9 system.

The deletion generates cell lines with the exclusive expression of the short NF-YA protein, which is the presumed stemness isoform of NF-YA. Assuming that the long NF-YA isoform is crucial during the differentiation process while the short protein maintains cells into a stemness state, under external stimuli cells knock-out for NF-YAl (NF-YAl-KO) will show a delay in the differentiation process.

4.1 CRISPR/Cas9 strategy

To generate mouse cell lines with the genomic deletion of exon 3, responsible for the NF-YA long isoform transcription, the CRISPR/Cas9 system was used.

The CRISPR system is an adaptable immune mechanism used by many bacteria to protect themselves from foreign nucleic acids, as viruses or plasmids. The type II CRISPR system from S. pyogenes has been adapted to induce sequence-specific double strand breaks (DSBs) and targeted genome editing.

Two components are necessary to perform genome editing using this engineered system: the Cas9, which is the enzyme that perform the cleavage of double-stranded DNA, and a guide RNA (gRNA) with 20 nt at the 5' end targeting the specific DNA-sequence, using standard RNA-DNA complementarity base-pairing rules. The enzyme cut the targeted-sequence that is immediately downstream to the -NGG PAM sequence.

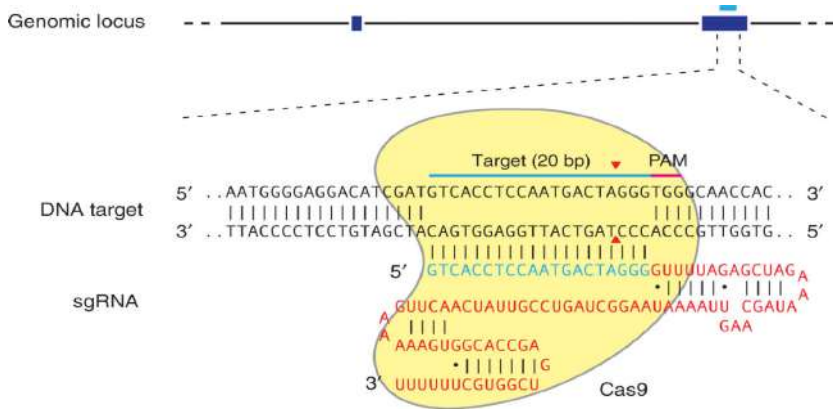


Figure 7. Schematic representation of the gRNA interacting with the Cas9 and targeted DNA sequence [99].

To increase target specificity, the mutant D10A nickase (Cas9n) can be used. This nickase cut only the DNA strand that is complementary to and recognized by the gRNA, making necessary the use of two gRNAs to have a DSB.

The on-line tool https://www.atum.bio/eCommerce/cas9/input, which allows to set different conditions, was used to design gRNAs for the genomic deletion of exon 3. The enzyme set was the Cas9 nickase, which requires the -NGG PAM sequence and two gRNAs to generate a double-strand break minimizing off-target effects. The targeted loci for gRNAs pairs were set with a gap of 8-20 bp for an optimal indels generation [99].

To excide the exon 3, the targeted regions were designed into the two flanking introns. FASTA sequence of 300 bp upstream (for guides A and B) and downstream (for guides C and D) the exon 3 were set into the program. gRNAs less likely to exhibit off target activity (higher score) and closer to the exon 3 were chosen, with a total targeted sequence of about 300 bp (Fig. 8 and Suppl. Fig.1).

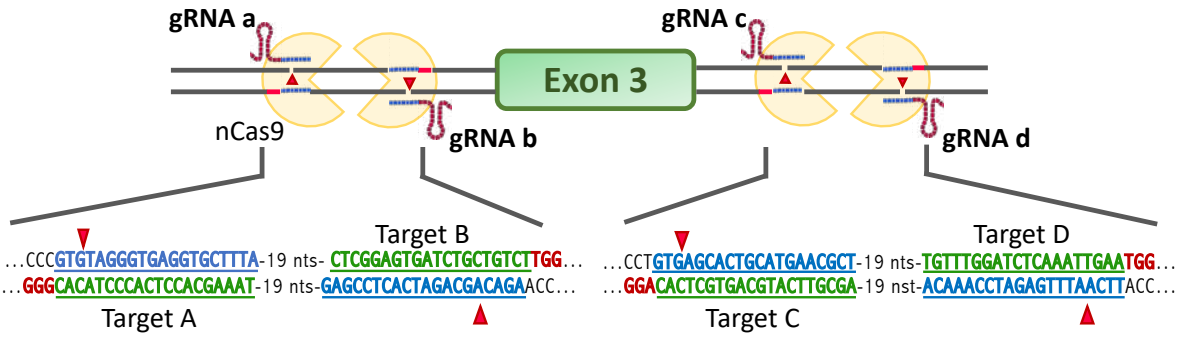


Figure 8. Gene editing strategy for NF-YA exon 3 deletion using the Cas9 nickase and four guide RNAs. The targeted sequence by each guide RNA and the deletion sites are shown.

Another tool was used to assess possible off-target sites for each single guide RNA (https://crispr.cos.uni-heidelberg.de). In general, possible off-targets identified have up to 3 mismatches (MM) in non-coding sequences, except for the gRNA_A which presents a possible off target in a coding sequence with 4 MM, but no common off-target sites were found in paired gRNAs (Suppl. Fig.2).

Two plasmids were generated following the *Sakuma et al.* [100] published method: both pX330A_d10A plasmids codified the Cas9 nickase while gRNAs were inserted unpaired (plasmid_1 carried the guides A and C while the plasmid_2 carried the guides B and D) in order to have enzyme activity only after simultaneous DNA recognition by all four gRNAs.

Plasmids were screened by performing the Polymerase Chain Reaction (PCR) using a forward primer that anneals to the plasmid, and the reverse primer which anneals to the specific inserted gRNA. Finally, plasmids were sequenced and used for cells transfection.

4.1.1 NF-YAl-Knock-out cell lines generation

Two methods were tested to reach high transfection efficiency: Lipofectamine 3000 and Electroporation. By transfecting a GFP plasmid, the efficiency of both methods was valuated and Electroporation showed higher transfection efficiency (~80%) compared to Lipofectamine 3000 (~50%).

For both cell lines, mESCs and C2C12 cells, $1x10^6$ cells were transfected by Electroporation with the two pX330A_d10A plasmids (palsmid_1 and plasmid_2) and then seeded at low density to have single plated cells, leaving them to grow. After 5 days, single colonies were picked and expanded to check for positive NF-YAl-KO clones. For clones screening, three pairs of primers (designed by using the on-line tool http://primer3.ut.ee/) were used for PCR: "primers A" were designed into the two exon 3-flanking Introns (846 pb amplification product), which produces a shorter amplification product in positive NF-YAl-KO clones; "primers B" (482 bp) and "primers C" (163 bp, it can be used in Real-Time PCR) have the forward primer designed into the Intron 4 and the reverse primer into the exon 3, giving an amplification product only in samples negative for the deletion (Fig. 9 and 10).

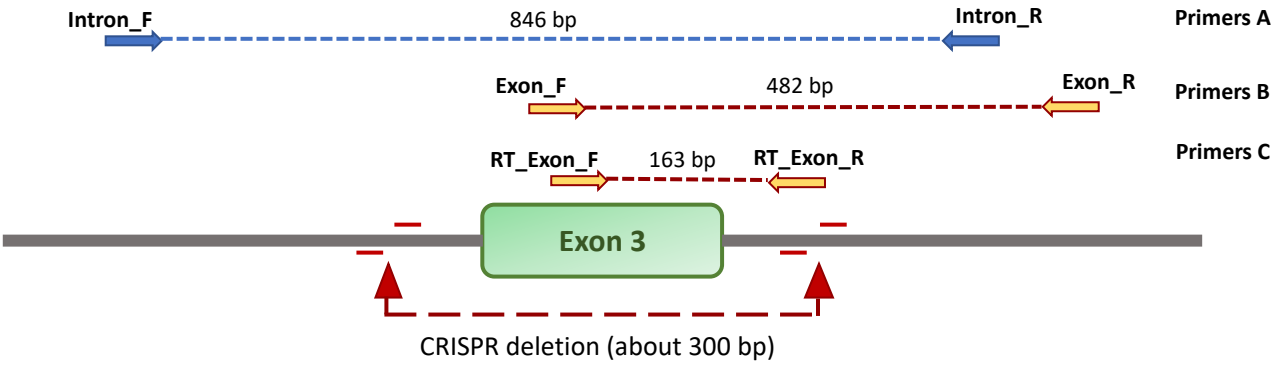


Figure 9. The three primers used to check for positive NF-YAl-KO clones are shown with the specific amplification products highlighted by the dashed lines.

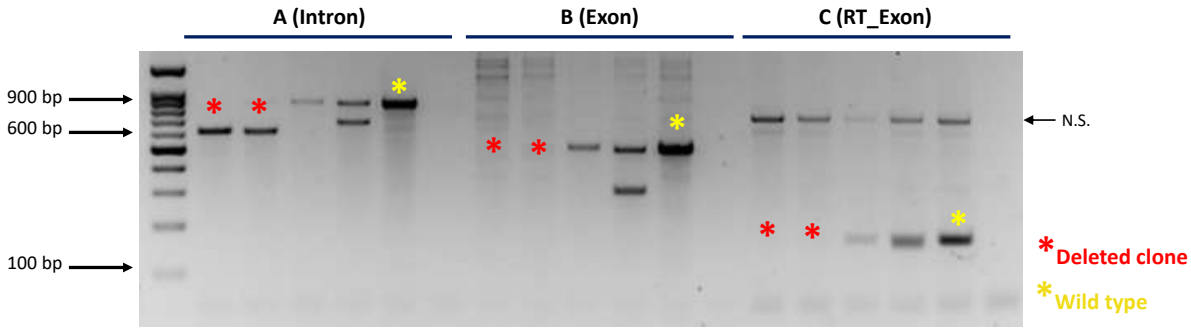


Figure 10. Example of PCR products run into a 1,2% agarose gel. The bands corresponding to wild type cells and positive deleted clones are marked (red: NF-YAl-KO clones; yellow: wild type cells).

Two isolated homozygous clones for each cell line (clones #63 and #149 for mESCs and clones #83 and #117 for C2C12 cells) were obtained. By sequencing the amplification product derived from primers A, the precise genome deletion of about 300 pb was confirmed (Fig. 11).

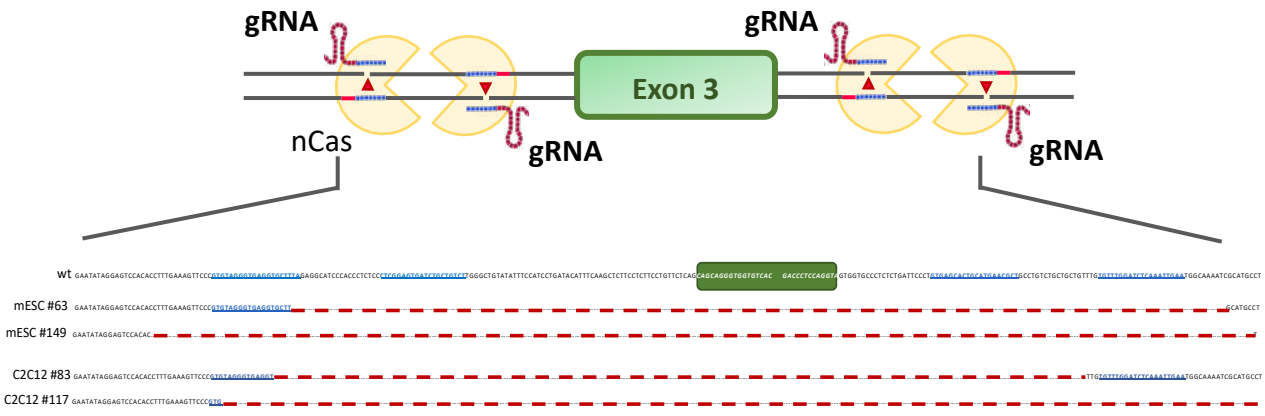


Figure 11. Sequencing of the amplification products derived from primers A. For each NF-YAl-KO clone, the dashed line shows the deleted sequence that includes the exon 3 of NF-YA.

4.2 Mouse Embryonic Stem cells

Mouse ESCs are pluripotent cells with the capability to differentiate into cells of three germ layers. The short NF-YA protein represents the preferential expressed isoform, but after differentiation by Embryoid Bodies formation, the short NF-YA protein levels decrease while the long increase, suggesting different roles of the two NF-YA isoforms in the regulation of stemness and differentiation pathways.

The generation of a mESC line with the genomic deletion of the exon3 can help to understand the specific pathways controlled by the two NF-YA isoforms.

The effective exclusive presence of the short isoform in NF-YAl-KO clones was analyzed by Western Blot which showed the unique expression of the short NF-YA protein in NF-YAl-KO clones while the expression of both protein isoforms in wt cells (Fig. 11A). These data were confirmed by qRT-PCR which showed low levels of the NF-YA long mRNA only in wild type cells while the absence in the two clones (Fig. 12B).

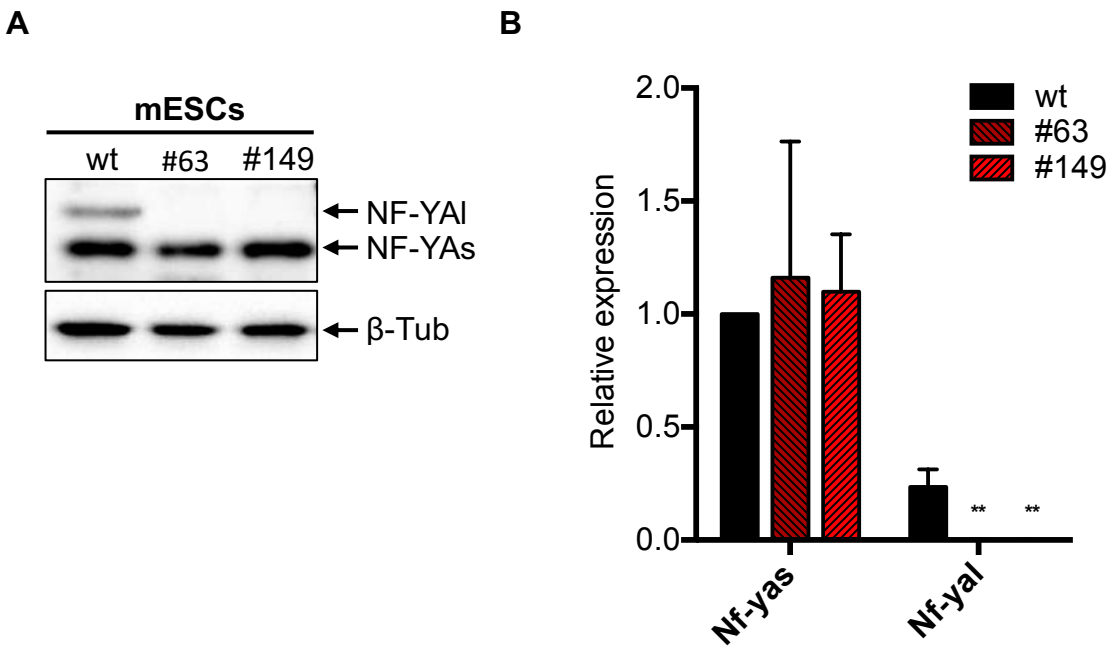


Figure 12. (A): Western blot analysis of NF-YA protein isoforms in wt mESCs and the two NF-YAl-KO clones (#63 and #149). **(B):** Gene expression analysis of NF-YA short and long levels in wt and NF-YAl-KO clones. Error bars represent the SD of three independent experiments. P-values were calculated using the one-way ANOVA test followed by Dunnet's post-test using wt cells as control column.

4.2.1 NF-YAl-KO mESC clones maintain stemness identity

The NF-YA short isoform is considered the "stemness isoform" of the NF-Y complex, with higher expression level in mouse ESCs, while after differentiation the NF-YA long levels increase becoming the predominant isoform.

Based on these data, the generation of a mouse ESC line which express only the short isoform of NF-YA could results in mESCs that, under external stimuli, would be unable to differentiate or would show a different pattern.

Once obtained NF-YAl-KO mESCs clones, the first step was to compare wt cells and the two deleted clones (clone #63 and #149) checking for stemness maintenance.

Phase-contrast analysis of NF-YAl-KO mESC clones

Undifferentiated mouse ES cells, when observed under light microscopy, grew in typically spherical and smooth colonies of variable size. Individual cells are difficult to distinguish within the ES cell colonies, except for the nuclei comprising normally one or more nucleoli.

Morphological analysis highlighted no differences between wild type (wt) cells and the two NF-YAl-KO clones, which maintain the classical stem cell morphology with round shape, large nucleoli and a scan cytoplasm. A more specific assay was performed to highlight the maintained stemness identity of NF-YAl-KO clones. The colorimetric Alkaline phosphatase (ALP) assay is a colorimetric assay that can highlight differences in the expression of this stemness marker by a very fast protocol. The results confirmed that the two NF-YAl-KO clones (#63 and #149) maintained high expression levels of ALP, comparable to the wt cells, determining their stemness identity.

To further confirm these data, immunostaining for the stemness transcription factor NANOG was performed. In normal stem cell conditions (+LIF), wt cells and the NF-YAl-KO clones showed the same expression level of NANOG, corroborating the results obtained by morphological analysis and ALP assay (Fig. 13).

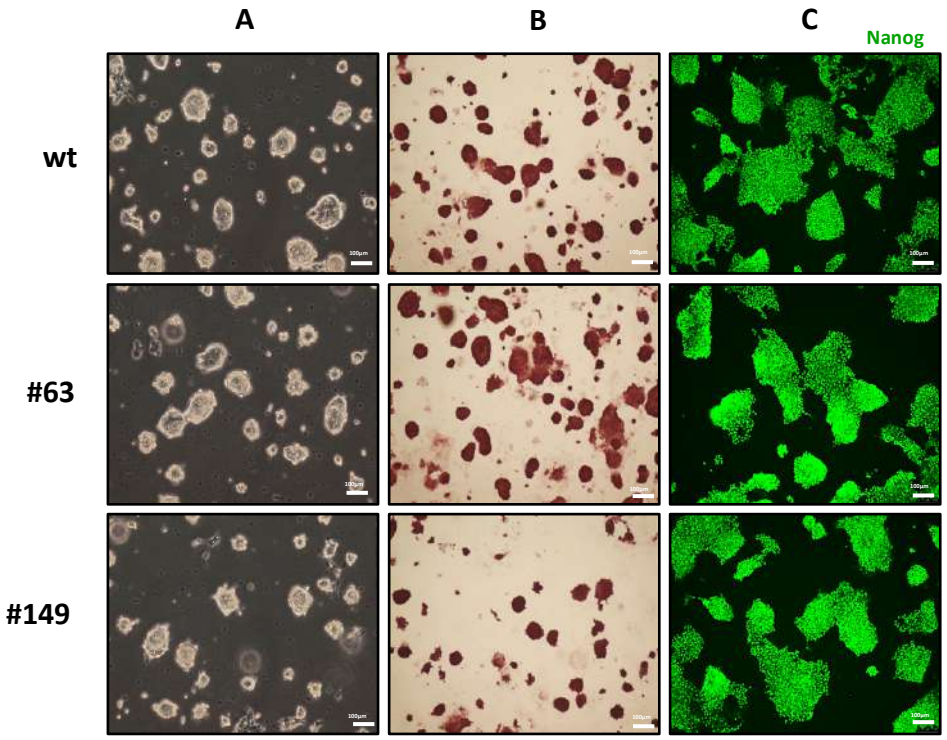


Figure 13. Phase contrast analysis of mESCs wt and the two NF-YAl-KO clones (#63 and #149). Morphological analysis (A), ALP staining (B) and Immunostaining for Nanog (C) are shown.

Gene expression analysis of stemness markers

The Stem cells transcriptome is peculiar of these cells, with genes specifically expressed only in Embryonic Stem cells. Nanog is one of the key regulators of stemness maintenance and its expression rapidly decreases after differentiation stimuli. To further characterize the obtained clones, gene expression analysis was performed on key stemness regulators like Nanog, Sox2, Klf4, Oct4, Napsa and ALP. Compared to wild type cells, NF-YAl-KO clones showed similar expression levels for each analyzed gene (Fig. 14 A). Western Blot analysis confirmed these data showing comparable expression levels of NANOG protein. Moreover, the NF-YB protein levels were valuated as further internal control (Fig. 14 B).

These results indicated that the exon 3 deletion, and the consequent lost expression of NF-YA long protein isoform, didn't affect stemness identity, which is instead maintained by the exclusive expression of the NF-YA short protein.

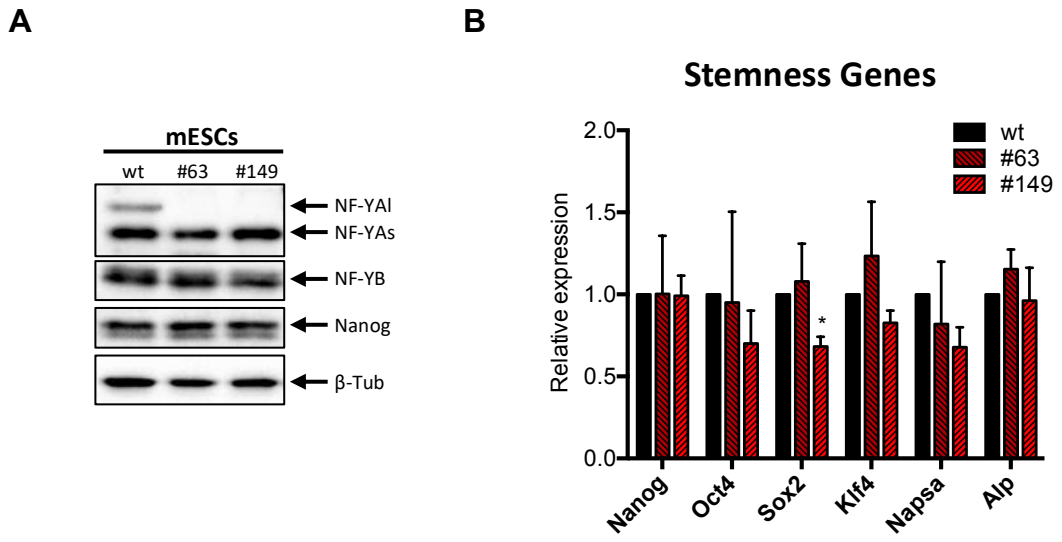


Figure 14. (A): Western blot analysis of protein expression levels in mESCs wt and the NF-YAl-KO clones. β-Tubulin is taken as internal control. **(B):** Gene expression analysis of key stemness genes in wt and NF-YAl-KO cells. Error bars represent the SD of three independent experiments. P-values were calculated using the one-way ANOVA test followed by Dunnet's post-test using wt cells as control column.

4.2.2 NF-YAl-KO mESC clones maintain stemness identity under differentiation stimuli

Mouse ESCs preferentially express the short NF-YA isoform, but after differentiation by Retinoic Acid treatment and/or Embryoid Bodies formation the expression of NF-YA protein completely switches to the long isoform, suggesting that NF-YA long is important for the differentiation while the short isoform could be the "stemness isoform" necessary for pluripotency maintenance.

In order to check the differentiation potential of NF-YAl-KO clones, Stem cells were grown in several differentiation conditions. At first, it was tested the capability of clones to maintain stemness identity under LIF withdrawal (necessary for stemness maintenance). 48 hours after withdrawal of LIF, classical stemness colonies disappeared in wt cells becoming less dense and more flattened with an increasing number of individual separately growing cells, while both NF-YAl-KO clones continued growing in compact and dense colonies, like Embryonic Stem cells. Similar results were obtained by using the Retinoic Acid (RA), a vitamin A metabolite essential for early Embryonic development which promotes Stem cell neural lineage specification. In addition to LIF withdrawal, the RA into the Stem cell medium gave similar results to those obtained by simple LIF withdrawal: the two clones maintained a stemness morphology, differently from wt cells which instead started to differentiate (Fig. 15 A).

NANOG protein is one of the key stemness regulators and its gene expression levels rapidly decrease after differentiation induction. To further validate our data, an immunostaining for NANOG were performed 48 hours after LIF withdrawal. In wt cells a low signal was detected by the anti-NANOG antibody confirming low expression levels after differentiation. On the contrary, in NF-YAl-KO clones a strong fluorescent signal from NANOG protein was detected, confirming the maintenance of high levels of this key stemness regulator and the consequent maintained stemness identity (Fig. 15 A).

Analyses on NF-YA protein levels revealed that 48 hours after LIF withdrawal and Retinoic Acid treatment the NF-YA short and long protein levels are the same that in LIF growing condition. These data suggest that the long isoform of NF-YA, although expressed at low level, is important for cells transition to a differentiated state while the exclusive presence of the short isoform arrests cells to the stemness state (Fig. 15 B).

To test if the absence of the long isoform of NF-YA affects the first stages of embryonal development, the Epiblast like cells (EpiLCs) differentiation protocol was tested. EpliLCs are transient cells that form from naïve Stem Cells after 48 hours of differentiation induction into Epiblast Stem Cells (EpiSCs), showing a transcriptome similar to the pre-implantation Embryo. In order to obtain more homogeneous Stem colonies, Stem cells were grown in 2i medium [65]. Alkaline Phosphatase assay after EpiLCs differentiation revealed a low expression levels of this stemness marker in wt cells, as expected, while high levels in both NF-YAl-KO clones (Fig. 15 C). The analysis of gene expression levels of differentiation markers like Fgf15, Fgf5 and Otx2 revealed a lower induction in the two NF-YAl-KO clones compared to wt cells (Fig. 15 D). These data confirm that in Stem cells which express only the short isoform of NF-YA the differentiation process seems to be inhibited despite external stimuli.

Western Blot analysis revealed the absence of NF-YA protein switch from the short to the long isoform even after EpiLCs differentiation (Fig. 15 C), confirming that also during the first steps of development the presence of the long protein isoform of NF-YA, although at low levels, is crucial to overcome the stemness stage while the exclusive presence of the short protein inhibit the differentiation process.

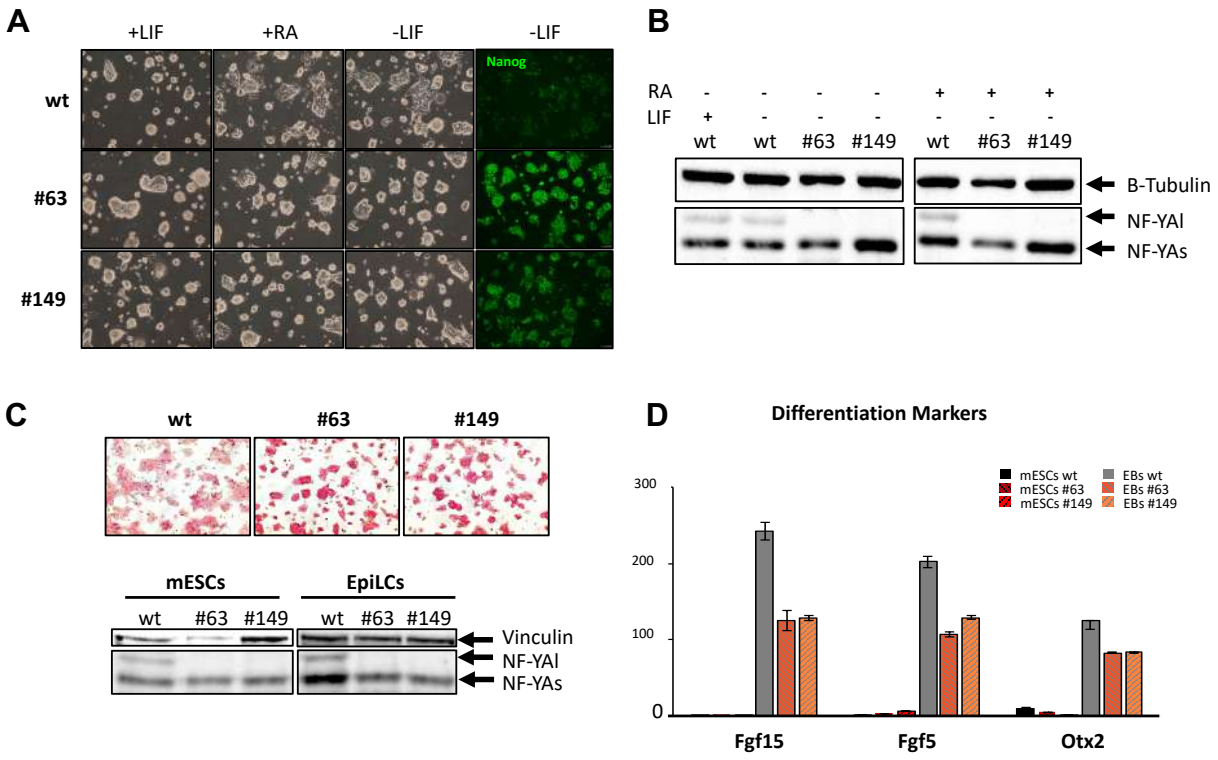


Figure 15. (A): Morphological analysis and Immunostaining for Nanog of wt cells and NF-YAl-KO clones in the

presence of Leukemia Inhibitory Factor (LIF) and 48 hours after LIF withdrawal and Retinoic Acid (RA) treatment. **(B):** Western Blot analysis of NF-YA isoform proteins 48 hours after LIF withdrawal and RA treatment. **(C):** Alkaline Phosphatase staining (upper panel) and Western Blot analysis of NF-YA isoform proteins (lower panel) on wt cells and NF-YAl-KO clones 48 hours after differentiation into EpiLCs. **(D**): qRT-PCR analysis of mRNA levels of differentiation markers in wt cells and NF-YAl-KO clones before and after 48 hours of differentiation into EpiLCs.

4.2.3 NF-YA protein switch is necessary for cardiomyocytic differentiation

The most robust method for generating differentiated cell types is through the formation of Embryoid bodies (EBs) in which ES cells spontaneously develop into 3-dimensional aggregates similar to Embryos at the egg-cylinder stage, consisting of three tissues: the epiblast, the extraembryonic ectoderm, and the visceral endoderm. These aggregates mimic the early steps of pre-implantation development, including the formation of an outer layer of Primitive Endoderm and an inner Epiblast that will cavitate and form a columnar epithelium [101].

In order to investigate the differentiation potential of mES NF-YAl-KO clones, wt and NF-YAl-KO clones were used to form floating EBs. The morphological analysis revealed that, starting from 4 days after EBs formation, the EBs wt started to cavitate while the EBs NF-YAl-KO maintain a round shape and a compact conformation until the 6^{th} day (Fig. 16 A). More interesting, wt and NF-YAl-KO cells were dissociated from 6-days-old EBs and re-cultured in "2i medium" to induce Stem cell colonies re-formation and then stained for ALP activity. 4 days after plating, the ALP assay highlighted the presence of a great number of Stem cell colonies only in the NF-YAl-KO clones (Fig. 16 A). Counting the total amount of pluripotent colonies, the wt cells formed 23 colonies, the NF-YAl-KO clones (#63 and #149) formed 238 and 223 colonies respectively with an average of ten times more stemness colonies in the NF-YAl-KO clones (Fig. 16 B), suggesting that a huge number of NF-YAl-KO cells maintain the stemness identity even after differentiation stimulus.

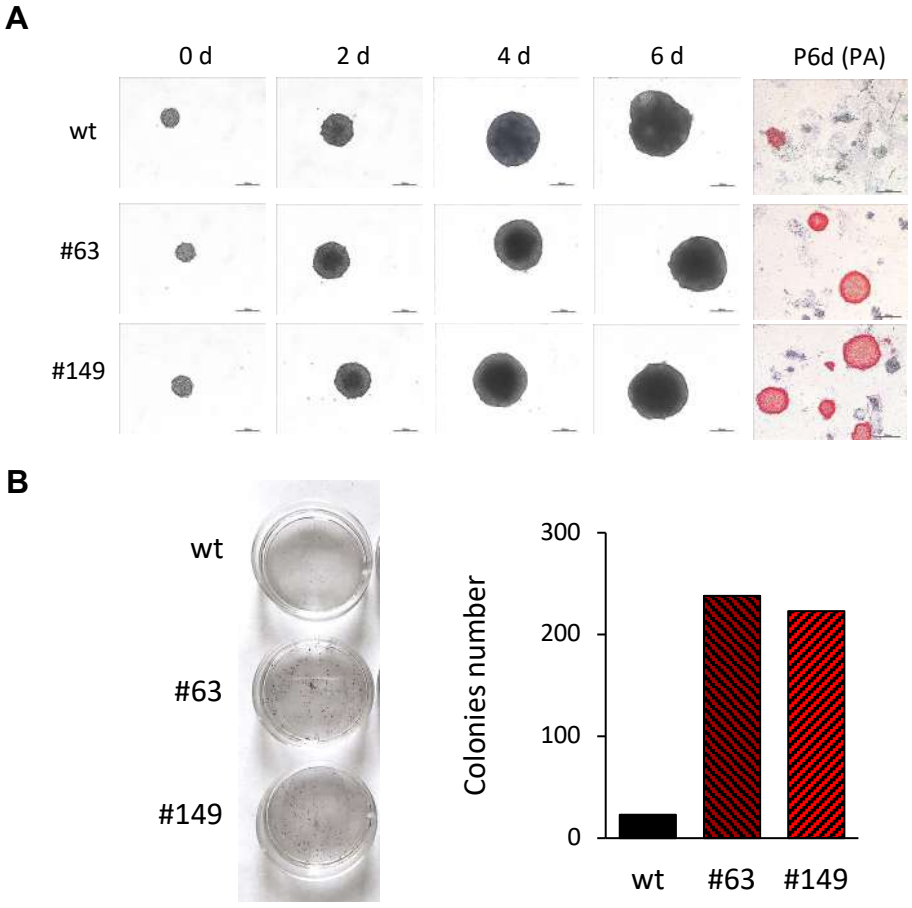


Figure 16. (A): Phase-contrast analysis of floating EBs morphology starting from mESCs wt and the two NF-YAl-KO clones, at different time-points. ALP assay performed 4 days after plating of cells dissociated from 6-days-old EBs and re-cultured in "2i medium". **(B):** Quantification of stem cell colonies number in wt and NF-YAl-KO clones after EBs plating in "2i medium".

In order to confirm that the absence of the long protein isoform affects the differentiation process and the short isoform maintain cells in the stemness state, wt cells and the two KO-NF-YAl clones were differentiated into cardiomyocytes. In standard condition, EBs spontaneously differentiate into cardiomyocytes, comparable to fetal cardiomyocytes, after plating [102]. 7 days after plating, adherent cardiomyocytes were collected for protein and RNA analysis. Analyzing the NF-YA protein level, wt cells shown a complete switch from the short isoform to the long one (as previously showed by *Dolfini D. et al.* [39]), while the two KO-NF-YAl clones expressed only the short protein isoform, as expected. NF-YB protein expression levels remain stable before and after differentiation in both wt and KO-NF-YAl

clones. Checking for stemness maintenance, NANOG expression levels were analyzed: in mESCs protein levels were comparable, but after differentiation wt cells completely lost the expression of the key stemness regulator while the two KO-NF-YAl clones maintained the expression of NANOG, although at lower level compared to mESCs. Moreover, the analysis of NESTIN protein levels as differentiation marker revelated that it was expressed only in wt cells after differentiation into cardiomyocytes (Fig. 17).

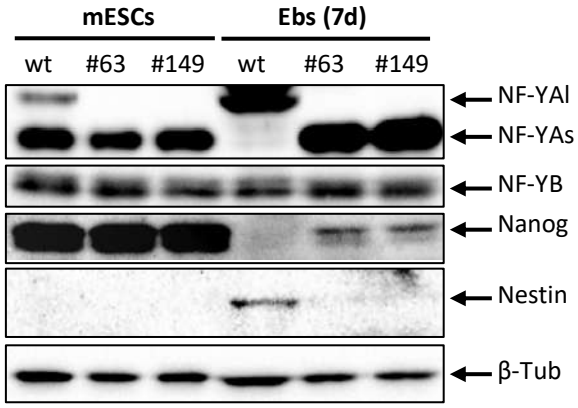


Figure 17. Western Blot analysis of specific protein levels in wt cells and the two NF-YAl-KO clones (#63 and #148) before and 7 days after differentiation into cardiomyocytes.

In order to confirm these data and to find the specific pathways up and down-regulated in NF-YAl-KO clones before and after differentiation, mESCs and EBs of three different experiments were used to perform the RNA-seq experiments.

The Principal component analysis (PCA) is a dimension-reduction tool used to reduce a large set of variables to a small set that still contains most of the information in the large set, making data easy to explore and visualize. Using the data from the RNA-seq, PCA analysis revealed a high similarity between wt and NF-YAl-KO clone in the stemness state, while after differentiation, wt cells are completely different from mESCs, as expected. Analyzing the NF-YAl-KO EBs, the two clones shown high similarity to each other but a great difference from wt EBs (Fig. 18 A).

Gene expression analysis highlight similar levels of key stemness markers in mESCs wt and NF-YAl-KO clones, while after differentiation, NF-YAl-KO clones maintain higher expression levels of stemness genes and a lower expression of differentiation markers compared to wt EBs (Fig.

18 B). qRT-PCR was performed to validate the gene expression of stemness and differentiation markers found to be different between wt and NF-YAl-KO EBs (Fig. 18 C).

Finally, Gene Ontology (GO) analysis confirmed the main up-regulated pathways of cell-cycle and Embryo development while the main down-regulated genes were related to vesicle-mediated transport and regulation of cell migration.

Although the data derived from the RNA-seq experiments are preliminary analysis, taken together these results highlight the importance of NF-YA protein switch from the short to the long isoform during the differentiation process, confirming the pivotal role of the short NF-YA protein in stemness regulation maintaining high expression levels of key stemness regulators even under differentiation stimuli, and moreover, the absence of the long isoform of NF-YA delay the expression of differentiation markers.

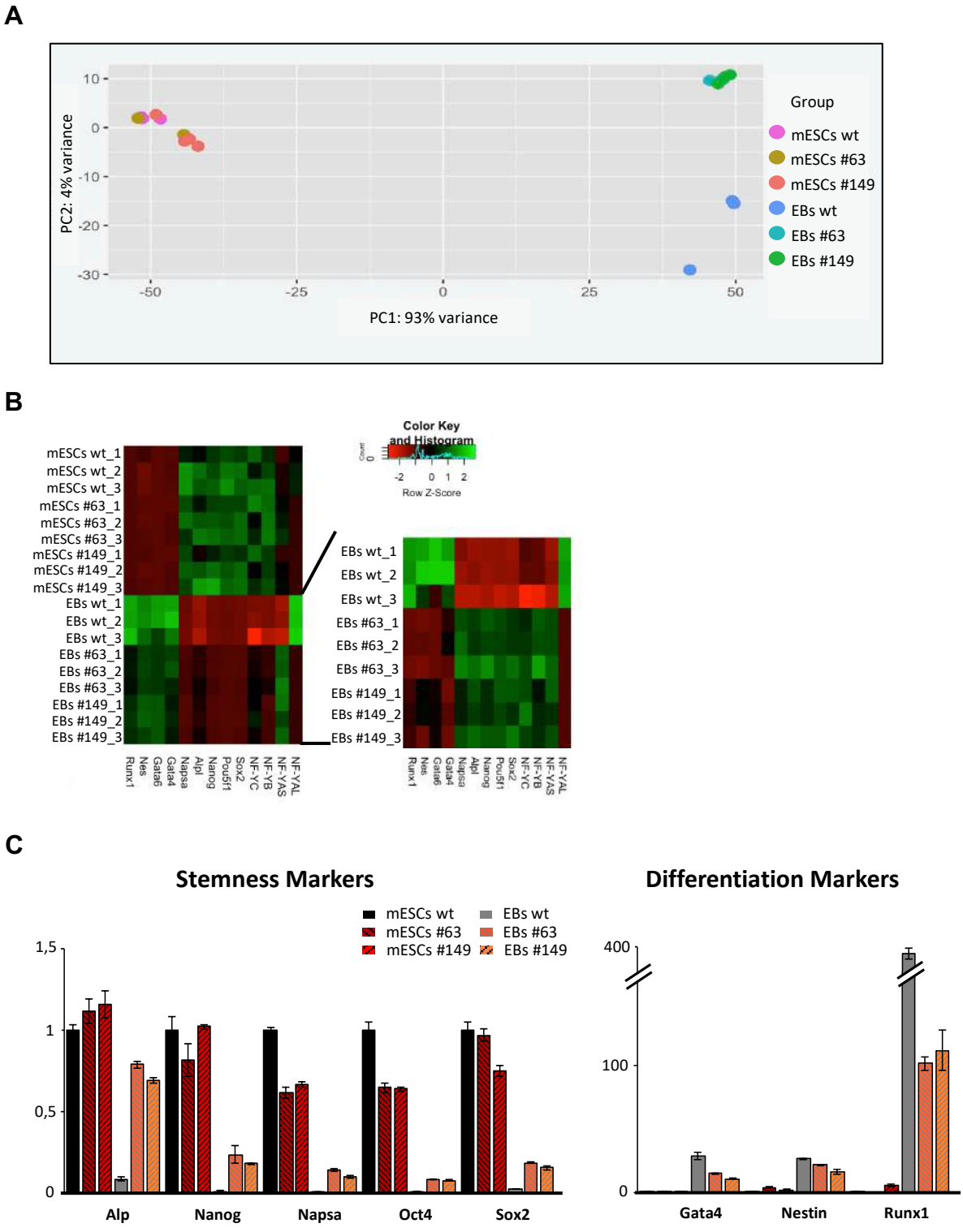


Figure 18. (A): Schematic representation of Principal component analysis (PCA) made on expression profiling data in indicated samples. **(B):** Heatmap representing expression levels of stemness and differentiation related genes in indicated samples. **(C):** Validation of RNA-seq analysis by qRT-PCR of stemness and differentiation markers in wt cells and the two NF-YAl-KO clones before and after 7 days of differentiation by EBs formation.

4.3 Myoblast C2C12 cells

4.3.1 KO-NF-YAl C2C12 clones maintain myoblast morphology

C2C12 cells are mouse myoblast cells with the intrinsic capability to differentiate into myotubes 72 hours after serum starvation. The Myogenic Regulatory Factors (MRFs), which are composed by MyoD1, Myf5, Myogenin and Mrf4, are the E-box binding Transcription Factors that interact with the myocyte enhancer factor-2 (Mef2) activating the expression of muscle-specific genes [73], [79], [80].

Analyzing the NF-YA protein expression levels, C2C12 cells express only the long isoform and Chip analysis revealed that NF-Y directly binds the promoter of Mef2D gene [103], meaning that NF-Y is important to activate the differentiation process. After differentiation the NF-YAl protein levels decreased until undetectable levels in skeletal muscle and heart with the consequent loss of NF-Y binding to the CCAAT boxes [44]. Moreover, stable over-expressing NF-YAs C2C12 cells revealed high expression levels of proliferation markers with the inhibition of differentiation, suggesting different roles of the two NF-YA protein isoforms [103].

In order to highlight the function of NF-YA short and long protein in muscle cells pathways, a C2C12 cell line with the genomic deletion of the exon 3 was generated by using the CRISPR/Cas9n system to obtain C2C12 cells expressing only the short NF-YA protein.

Based on previous published data obtained in C2C12 cells [103] and data described in this thesis obtained in NF-YAl-KO mESCs and in C2C12 cells transfected with the two NF-YA protein isoforms [98], myoblasts with the exclusive expression of NF-YA short isoform would show a delay in the expression of muscle-specific markers and the consequent delay in the differentiation into myotubes.

NF-YAl-KO C2C12 clones maintain muscle identity

Muscle C2C12 cells express only the long NF-YA isoform. Previous studies have shown that the stable overexpression of NF-YAl enhances the differentiation into myotubes, while the short protein isoform inhibits the differentiation maintaining high expression levels of cell-cycle genes [103], suggesting that the long NF-YA isoform represents the muscle-isoform by activating lineage-specific genes.

In order to check if the exclusive presence of the short NF-YA isoform interfere with the muscle-identity of C2C12 cells, the morphology of NF-YAl-KO C2C12 obtained clones was

analyzed in growth medium (GM) condition. Phase-contrast analysis revealed similar morphology of the two obtained clones compared to wt cells, suggesting that the absence of the long NF-YA isoform didn't affect cells morphology (Fig. 19 A). The analysis of NF-YA protein levels confirmed the exclusive presence of the long NF-YA protein in wt cells and the NF-YA short protein expression in NF-YAl-KO clones, while the NF-YB protein was express at similar levels in both wt and NF-YAl-KO clones. Western Blot analysis showed the same protein levels of cyclin B1 (Fig. 19 B) and proliferation assay showed that growth curves are similar to the wt (Fig. 19 D), confirming high similarity of these cell lines in growing condition. Analyzing the expression level of NF-YA long and short isoforms, qRT-PCR confirmed the presence of the long NF-YA mRNA only in wt cells and the expression of the short NF-YA only in NF-YAl-KO clones (Fig. 19 C). Moreover, Cyclin B1/2 and Pcna showed similar expression levels in wt and NF-YAl-KO clones (Fig. 19 E).

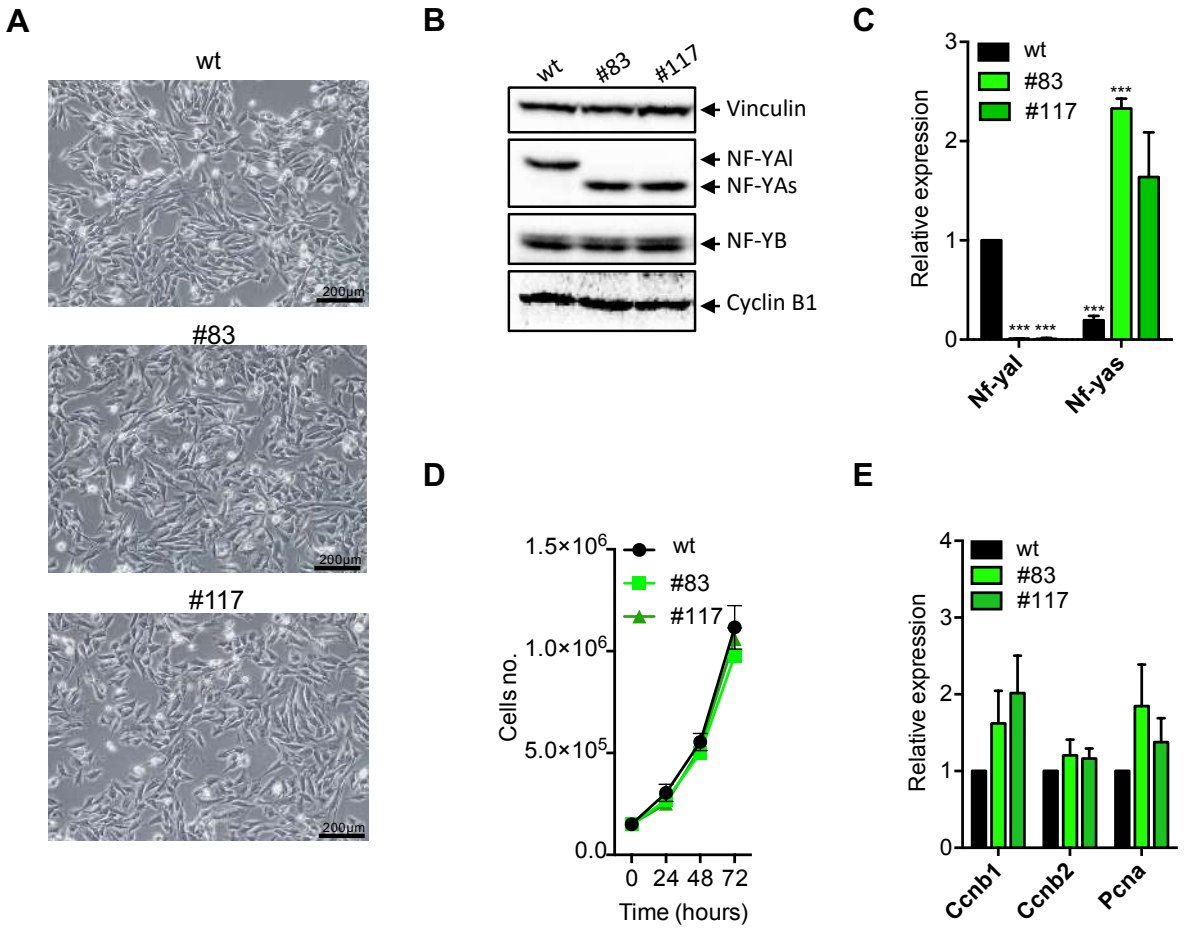


Figure 19. (A): Morphological analysis of myoblast wt and the two NF-YAl-KO clones by phase-contrast imaging. **(B):** Western Blot performed on total protein extracts of wt cells and the two NF-YAl-KO clones. The specific

used antibodies are indicated. **(C):** Gene expression analysis of NF-YA short and long levels in ctr and C2-YAl-KO clones (#83 and #117) in GM condition. Error bars represent the SD of three different experiments. P-values were calculated using the one-sample t-test. **(D):** Proliferation assay performed in GM condition counting every 24 hours until 72 hours using the Trypan Blue dye exclusion test. Error bars represent the SD of three different experiments. P-values were calculated using the one-way ANOVA test. **(E):** Gene expression analysis of key cell-cycle regulators in ctr and C2-YAl-KO clones (#83 and #117) in GM condition. Error bars represent the SD of three different q-PCR replicates. P-values were calculated using the one-sample t-test.

These results showed that clones exclusively expressing NF-YAs have an apparently normal morphology and growing curves.

4.3.2 NF-YAl is crucial for myotubes formation

Myoblast C2C12 cells are widely used cells with the intrinsic capability to recapitulate the process of adult myogenesis. After differentiation induction by serum starvation (differentiation medium, DM), C2C12 cells start to differentiate into myotubes showing a drop in the expression levels of Myf5 and an increase of MyoD1, followed by Myogenin and Myosin [103].

In order to check if the long NF-YA isoform, which is the exclusive expressed isoform in C2C12 cells, is crucial for muscle differentiation, wt and NF-YAl-KO clones were grown in DM. Morphological analysis was performed until 72 hours after differentiation induction. Starting from 48 hours of differentiation, wt cells shown the classical myotubes formation with a low number of non-differentiated cells, while the two NF-YAl-KO clones completely lost the capability to fuse and differentiate into myotubes showing the exclusive presence of single myoblast cells, like in GM condition (Fig. 20).

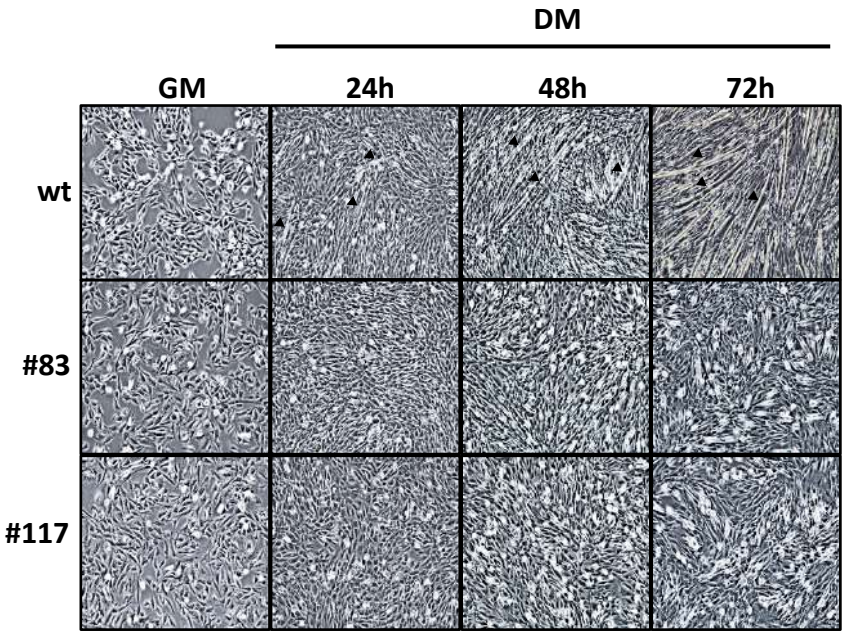


Figure 20. Phase contrast analysis of myoblasts (wt and the NF-YAl-KO clones) before and after differentiation induction. Myoblast formed starting from 24 hours after differentiation induction are indicated (arrow).

In order to validate the morphological data, a time course of the expression levels of key muscle-specific markers were evaluated. The Six5 and HtrA1 genes, known to be important for muscle development [104], [105], showed similar expression pattern in wt cells and the two NF-YAl-KO clones in GM condition and after differentiation induction, while specific muscle-differentiation markers, like Myogenin and Mef2C, shown a delay in the expression levels in the NF-YAl-KO clones (Fig. 21 A). Moreover, 72 hours after differentiation induction NF-YA protein levels decreased in both wt and NF-YAl-KO clone while only wt cells expressed high levels of Myosin Heavy Chains (MyHCs) proteins confirming the differentiation into myotubes of wt cells but not NF-YAl-KO clones (Fig. 21 B). Consistent with these data, cyclin B1 and B2 revealed higher expression levels in NF-YAl-KO clones while wt cells substantially lost the expression of these cell cycle genes (Fig. 21 A).

Interestingly, Wisp2, Bmp4, Kitl and Figf genes showed significant increase in the expression levels in NF-YAl-KO clones compared to wt cells (Fig. 21 C). In *Basile V. et al.* [103] the stable overexpression of the short isoform of NF-YA in C2C12 cells gave similar results: after 72 hours of differentiation induction they observed high expression levels of these genes in C2C12 cells that stable overexpress the short isoform of NF-YA but not in NF-YA long overexpressing cells.

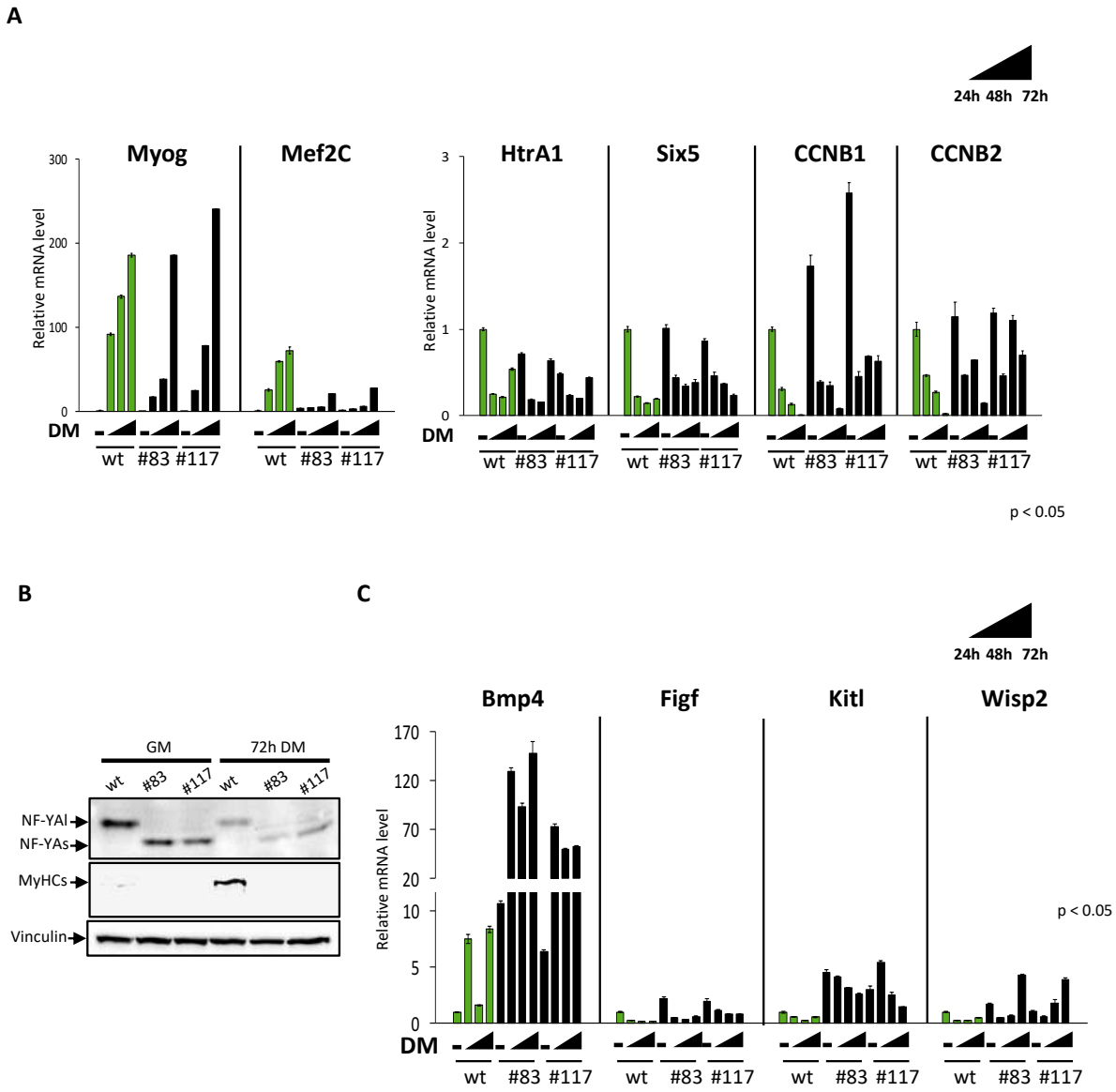


Figure 21. (A) Gene expression analysis of key muscle differentiation (left panel), muscle development and cell-cycle related genes (right panel), before and after differentiation at different time points, of wt cells and the two NF-YAl-KO clones. **(B)** Western Blot analysis of NF-YA isoforms and MyHCs protein levels in wt and the NF-YAl-KO clones, before and after 72 hours of differentiation induction. **(C)** Gene expression analysis of known up-regulated genes in C2C12 cells overexpressing the NF-YA short protein isoform (*Basile V et al.*, 2016), performed in wt cells and the NF-YAl-KO clones before and after differentiation induction.

The up-regulation of these genes in NF-YAl-KO clones suggests a possible specific regulation of these targets by the short NF-YA protein isoform. The two protein isoforms differ only for 28 amino acids located at the N-terminus into the Q-rich transactivation domain, while the

DNA-binding domain is localized at the C-terminus giving the same CCAAT-binding affinity to the trimers composed by NF-YAs or NF-YAl. This means that the two isoforms cooperate with specific interactors to activate different genes: while the short isoform could regulate cell-cycle genes like Cyclin B1 and B2, the long isoform could activate the transcription of muscle differentiation markers.

This idea is confirmed by the similarity of NF-YAl-KO clones to wt cells in GM condition and the concomitant loss of the capability to differentiate into myotubes showing a delay in the expression levels of specific muscle-differentiation markers, sustaining high levels of cell-cycle genes.

4.3.3 Expression of MRFs, Myomaker and Myomixer are impaired in NF-YAl-KO clones

NF-YAl-KO clones completely lost the capability to fuse and to originate myotubes. There are two genes, Myomaker (Mymk) and Myomixer (Mymx) that are transcriptionally induced during muscle differentiation and codified for myoblast-specific proteins essential for myoblast fusion. These two proteins works independently mediating distinct steps in the fusion pathway: Mymk controls the membrane hemifusion while Mymx is required for fusion pore formation [106], [107], [108].

Analyzing the expression levels of Mymk and Mymx, we observed a strong lower expression in NF-YAl-KO clones compared to the wt cells already in GM condition (Fig. 22 A). Moreover, Myogenin mRNA levels are strongly reduced in NF-YAl-KO clones in GM condition as 24 hours after differentiation induction, data confirmed by the protein levels (Fig. 22 B). MyoD1 and Mef2C mRNA levels are substantially normal in GM, but induction is reduced upon differentiation, compared to control cells (Fig. 22 A). Interestingly MyoD1 protein levels are substantially reduced NF-YAl-KO clones both in GM and DM conditions, compared to wt cells (Fig. 22 B), differently from the mRNA results (especially in GM conditions), which calls for post-transcriptional regulation of this gene.

Myf5 expression is basally similar in the edited clones, and higher after differentiation while Mef2D expression is not affected in the two NF-YAl-KO clones compare to wt cells (Fig. 22 A). All these data confirm that the two NF-YAl-KO clones completely lost the capability to differentiate.

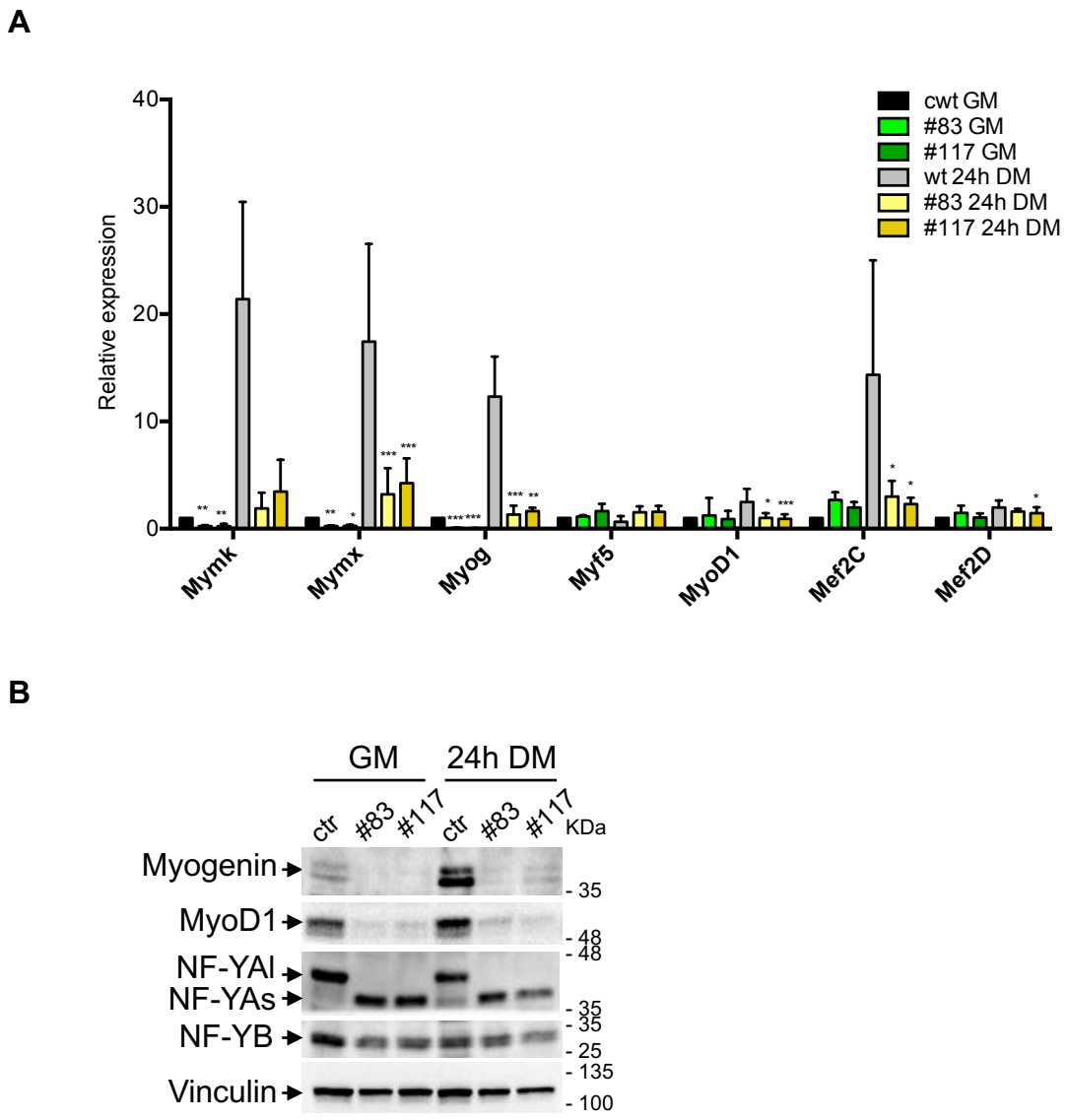


Figure 22. (A): Gene expression analysis of key muscle differentiation markers in wt cells and NF-YAl-KO clones before in GM and 24 hours after differentiation induction (24h DM). Error bars represent the SD of three different experiments. P-values were calculated using the one-sample t-test. **(B):** Western blot analysis of Myogenin, MyoD1, NF-YA and NF-YB protein levels in wt cells and NF-YAl-KO clones before in GM and 24 DM. Vinculin was used as loading control.

4.3.4 NF-YAl in not sufficient to rescue muscle differentiation

In order to check if the maintenance of myoblast identity under differentiation stimuli is sustained by the NF-YA short isoform or is due to the absence of the long protein isoform, C2C12 cells were transfected with the recombinant NF-YAl protein [98] and the mutant NF-

YAm23-33 protein, which is mutated in the two domains responsible for NF-YB/NF-YC interaction and DNA-binding. Protein transfection efficiency was checked by Western Blot analysis collecting samples 30 minutes after protein transfections, confirming the exclusively entry of the wild type NF-YAl protein into cells (Fig. 23 B). 72 hours after differentiation induction, recombinant NF-YAl protein levels were still high (Fig. 23 C), as expected, and morphological analysis revealed that C2C12 wt cells normally differentiated into myotubes like in not transfected (NT) cells, while the NF-YAl-KO clones remained single cell myoblasts with the absence of fused myotubes (Fig. 23 A). Gene expression analysis performed on Myogenin, MyHCs and Mef2C genes confirmed the lower expression levels of these muscle-differentiation markers in NF-YAl-KO clones compared to wt cells even in cells transfected with the long protein of NF-YA (Fig. 23 D). These data suggest that the transfection of the long isoform in C2C12 cells expressing the short NF-YA protein is not sufficient to rescue the phenotype in differentiation conditions. Further experiments to rescue the differentiation capability can be done by silencing NF-YA in NF-YAl-KO clones and transfecting the long protein isoform. In case of rescued phenotype, this experiment could answer to the question of the specific role of NF-YA long in enhancing muscle differentiation.

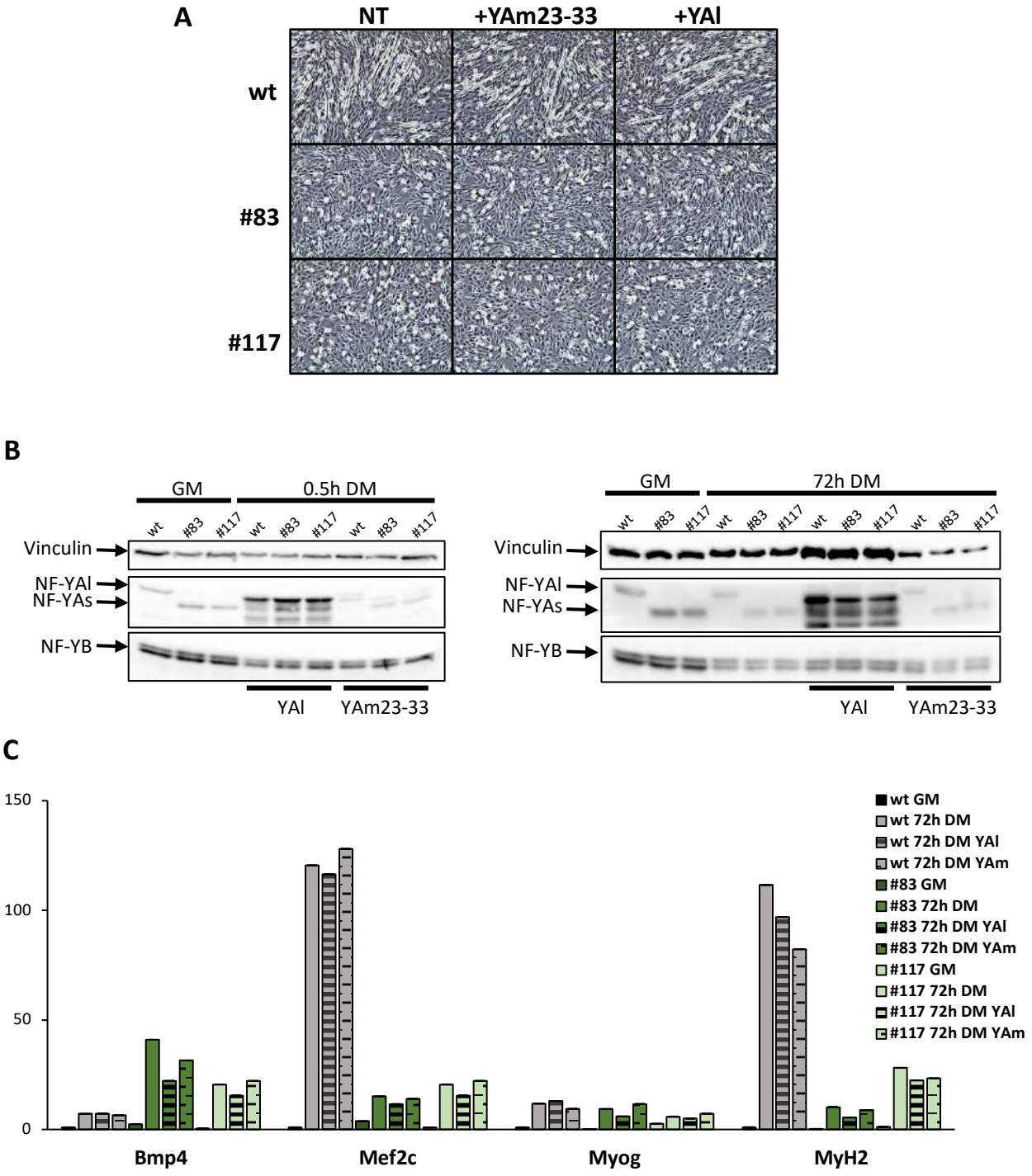


Figure 23. (A): Phase-contrast analysis of C2C12 wt and NF-YAl-KO clones 72 hours after differentiation induction and the concomitant transfection of NF-YA long wt and mutant proteins. **(B):** Western Blot analysis of recombinant NF-YA protein levels after C2C12 cells transfection. **(C):** Gene expression analysis of key muscle differentiation regulators in C2C12 wt and the two NF-YAl-KO clones, in GM, 72h DM and 72h DM plus recombinant NF-YA long (wt and mutant) protein transfections.

4.4 NF-YA enters cells through cell penetrating peptides

Cell penetrating peptides (CPPs) were large studied in the last years in order to avoid problems of drugs delivery through cell membrane, which represent an obstacle in the treatment of many diseases.

As mentioned in previous paragraphs, CPPs are short peptides with the ability to overpass cell membranes and transport fused molecules into cells. The main characteristics of this class of proteins is the presence of amino acid sequences rich in basic residues and the organization of their secondary structure in α-helices. One of the most used and characterized CPP is the HIV TAT peptide which was demonstrated to be able to efficiently transduce exogenous proteins into mammalian cells [84]. In particular*, Domashenko et al.* [72] demonstrated that the GST-TAT-NF-YA fusion protein can transduce into primary hematopoietic stem cells (HSCs) enhancing cell proliferation, suggesting a possible use of this fused protein as a tool for HSC/progenitors ex vivo expansion and therapy.

The NF-YA subunit represent the regulatory elements of the NF-Y trimer, composed by NF-YA, NF-YB and NF-YC, all necessary to bind the DNA. In particular, the NF-YA protein is found in two different isoforms, the long (NF-YAl) and the short (NF-YAs), derived by the alternative splicing of the exon 3.

Studies on the evolutionary conserved part of NF-YA structure have highlighted the presence of two features typical of Cell-Penetrating Peptides (CPPs): the presence of two α-helices (A1 and A2) and the abundance of Arginine [2]. Based on this evidence, this study [98] had the objective to test whether recombinant NF-YA would enter cells in the absence of any fused CPP.

The recombinant proteins NF-YAs and NF-YAl were produced in *E. Coli*, renatured from Inclusion Bodies and controlled by Coomassie staining. The effective capability of NF-YA to transduce cells without any fused CPPs was checked by adding recombinant proteins in serum-free condition for two 30’ intervals into the medium. Human colon carcinoma HCT116 cells, which express only the short isoform of NF-YA, were used to transfect NF-YA recombinant proteins in order to readily check the effective presence of the recombinant proteins into nuclei. Western blot analysis on nuclear extracts after 2 hours of protein transfection revealed that NF-YA enters cells at nanomolar concentrations, as it was previously shown for GST-TAT-NF-YA [72], following a dose-response addition.

Different cell lines were used to perform a time course of NF-YA long transfected protein: maximum protein levels were observed 30' after transfection and still high at 2 hours, while after 6 hours the protein amounts were considerably decreased and not detected after 24 hours.

Interestingly, the transfection of HCT116 and HeLa cells with an His-tagged NF-YAl recombinant protein fused to Thioredoxin (TRX-NF-YAl), produced and purified from the bacterial soluble fraction through NTA affinity chromatography, revealed an efficient entry of the TRX-NF-YAl protein comparable to the one shown for NF-YAl, validating that NF-YA has an intrinsic capacity to transduce into nuclei, regardless of method of production and purification, and of the presence of additional tags.

The localization of NF-YA into cells is exclusively in the nucleus but after mutation of both the K276-R277-R278 sequence in the A1 helix and the R310-K311-R312 residues in helix A2, the nuclear localization of NF-YA, in human cells *in vivo* and Xenopus cells *in vitro*, is completely lost [109]. Based on previous studies showing the possible overlap between NLSs and CPPs and based on the biochemical studies on NF-YA structure, the two helices A1 and A2 were tested for CPP properties.

In order to identify the specific amino acid sequence necessary for NF-YA transduction, NF-YAl proteins with mutations in the two helices responsible for the Nuclear Localization Signal (NLS), which were thought to be also important for cell membrane transduction, were produced.

To verify their functionality, all mutants were tested for DNA-binding capability in Electrophoretic Mobility Shift Assays (EMSAs) on the high affinity CCAAT box of the HSP70 promoter, while to test whether these mutants maintain the capability to penetrate cell membrane and to localize into the nuclei, HCT116 cells transfections were performed. The results highlighted the presence of at least two separate CPPs in NF-YA, one in the A1 helix, centered on the K276-R277-R278 sequence, and a second in the DNA-binding subdomain of the A2 helix, encompassing the basic R310-K311-R312 stretch. These data arouse grate interest because there are previous studies demonstrating the overlap between NLSs and CPPs. It is the case of IL-1α cytokine that has an NLS/CPP peptide comparable in efficiency to the TAT peptide [110]. More important, numerous transcriptional regulators, including TFs NF-κB, OCT6 and TCF1, were shown to have NLS peptides that could behave as carriers for cell

entry while the transcription factor OCT4 have been demonstrated to harbor a CPP in the third helix of the homeodomain [111], [112].

The capability of TFs to behave as CPPs suggests that they might be the signaling molecules used by cells for local paracrine circuits, like in the case of Monocyte cells that undergo differentiation to macrophages by activating a set of genes, many of which are CCAAT-dependent so transcriptionally regulated by NF-Y [113].

In order to check if the absence of the NF-YA mutant proteins in nuclear extracts is due to the mutation of the NLS with the retention of protein in the cellular cytoplasm, rather than a defect in CPP activity, cytoplasmic extracts of HCT116 transfected cells were analyzed. Western blot confirmed the small amounts of NF-YAl and mutated proteins in cytoplasmic extracts confirming that mutations in the NLS affect the NF-YA CPP activity.

Previous studies have shown the capability of GST-TAT-NF-YAs protein to promote the ex vivo proliferation and to increase the stem cells pools in HSCs preparations, which can be used for BMT (Bone Marrow Transplantation). BMT is the specific therapy used for many cancer types to filter healthy stem cells and to transfuse back to the same patient of to another people. For this reason, the presence of internal CPP(s) in NF-YA is thus potentially relevant, being possible to transfect cells and to increase the stem cells pool in HSCs.

To test the NF-YA maintained capability to interact with NF-YB/NF-YC dimer and to specifically bind the CCAAT box after cells transfection and nuclear uptake, nuclear extracts were used to perform Electrophoretic Mobility Shift Assays (EMSAs). HeLa cells, which express the short NF-YA isoform, were transfected with NF-YA short and long proteins. In order to distinguish the transfected NF-YA protein isoforms to the endogenous one, the NF-YB/YB minimal domain (NF-YB/YCmd) [34] was used. Interestingly, by adding the NF-YB/YCmd, nuclear extracts of HeLa cells transfected with the long isoform of NF-YA shown two separate bands: the faster band which correspond to the NF-YAs endogenous protein bound to NF-YB/YCmd and DNA, and the slower and more intense band which indeed correspond to the recombinant NF-YAl transfected protein bound to NF-YB/YCmd and DNA. After the addition of NF-YB/YCmd, Nuclear extracts of cells transfected with the short NF-YA isoform shown a single intense band corresponding to the sum of NF-YAs endogenous and recombinant proteins bound to NF-YB/YCmd and DNA.

To further substantiate these results and to assess for a biological function of NF-YA recombinant transfected proteins, C2C12 cells were used to perform protein transfections.

Based on published data, the NF-YAl protein expression, which is the major isoform present in proliferating C2C12, is switched off upon terminal differentiation and absent in myotubes. Moreover, the NF-YAs protein overexpression inhibits muscle differentiation [103].

C2C12 cells were transfected with both NF-YA recombinant protein every 24 hours and transfection efficiency was checked by western blot analysis, collecting cells 2 hours after treatment. Interestingly, 72 hours after differentiation induction cells transfected with NF-YAs, and to a slightly lesser extent NF-YAl, impaired formation of organized myotubes. Western blot analysis confirmed a reduction of the expression of the myotubes marker MyHCs in NF-YA recombinant protein transfected cells, while similar expression levels of Myogenin, which is indeed expressed in proliferating cells and throughout the process. Analysis on the growth rates of C2C12 transfected cells shown no significantly differences upon NF-YAs or NF-YAl uptake.

Taken together, these data indicate that recombinant NF-YA protein efficiently transduce cells penetrating into the nucleus and maintaining the specific capability to interact with NF-YB/YC dimer and bind the CCAAT boxes to regulate specific pathways like the process of C2C12 myocytes differentiation.

These findings highlighted the possible use of NF-YA recombinant proteins to direct transfect cells like Hematopoietic Stem Cells to increase the stem cells pool, using them in many cancer types therapies. However, the quick degradation of NF-YA protein requires further studies to prolong proteins stability. Interestingly*, Isabella M. et al.* [45] shown that the levels of NF-YA proteins are regulated post-translationally by ubiquitylation and acetylation: the mutation of four Lysines in the conserved domain into Arginines increased the NF-YA half-life in DNA transfections. Since these Lysines are not involved in the two identified CPP motifs, it's possible that the mutation of the four Lysines involved in poly-ubiquitination and degradation can prolong the NF-YA half-life, while not affecting CPP activity, nuclear retention and functionality of transfected NF-YA recombinant proteins.

5. Conclusions

The work performed during my PhD showed new important properties of the Nuclear Factor Y. The protein structure analysis highlighted the possibility of NF-YA to behave as CPP. Thanks to the work described in this thesis, NF-YA can now be catalogued into the CPPs family. Indeed, specific basic amino acids residues, located in the evolutionary conserved domain of NF-YA, were demonstrated to be necessary to confer to the protein the capability to penetrate the cell membrane and to translocate into the nucleus. The evidence of NF-YA to carry protein transduction domain and more important the maintained functionality after translocation into the nucleus are the basis to think about possible therapeutic applications. In bone marrow transplantation, the GST-TAT-NF-YA fusion protein transduction is used to stimulate HSCs growth enhancing cell proliferation. Thanks to the NF-YA properties, the fused protein can be replaced by the use of the recombinant NF-YA, avoiding tags or epitopes. Moreover, we demonstrated that NF-YA can transduce additional fused tags maintaining its intrinsic capacity to enter mammalian cells, suggesting a possible use in clinical trials to transduce specific tags into cells. It will be a possible future perspective to validate whether the identified CPPs of NF-YA can be used alone to transduce other molecules. However, the instability of the NF-YA transduced protein prevents the possibility to achieve a prolonged effect. Previous studies have shown that the mutations of Lysine residues of the conserved domain, which are involved in poly-ubiquitination and degradation, into Arginine residues increased the half-life of NF-YA in DNA transfections [45]. By using these mutations, it can be possible to increase the half-life of transfected protein without affecting transfection efficiency and protein functionality.

The other aspect of NF-YA studied in this thesis concerns the specific role of the short and long isoforms in the regulation of stemness and differentiation. The analysis of the NF-YA null mice revealed an early embryonic lethality, with no embryo found older than 8.5 days, suggesting a pivotal role of NF-YA in early mouse Embryonic development [70]. Interestingly, mESCs preferentially express the short NF-YA isoform but after differentiation by RA treatment and/or EBs formation, a complete switch from the short to the long isoform was observed [6]. The transfection of the TAT-NF-YA short protein into mESCs after LIF withdrawal can stimulate cell proliferation maintaining cells in the pluripotent state [39]. Moreover, in C2C12 cells, which express only the long isoform of NF-YA, the overexpression of the short isoform

inhibited the cell capacity to differentiate into myotubes, while NF-YAl-overexpressing cells showed a downregulation of cell cycle genes while an enhanced expression of muscle-specific genes [7]. Based on these evidences, my work focused on the study of the specific roles played by the two NF-YA isoforms in maintaining cells in the stemness state and in enhancing the differentiation process. By using the CRISPR-Cas9 system two cell lines, the mESCs and the C2C12 cells, were generated with the genomic deletion of the exon 3, which codified for the NF-YA long protein. The results presented in this thesis confirmed the pivotal role of the NF-YA short protein in maintaining high expression levels of key stemness regulators and the concomitant inhibition of differentiation markers. Starting from LIF withdrawal to cardiomyocytes differentiation by EBs formation, the exclusive presence of the NF-YA short protein could maintain high expression levels of key stemness markers like Nanog, Oct4 and Sox2. Because the only region that differs between the short and the long isoform is part of the Q-rich transactivation domain, while the two alpha helices responsible for NF-YB/YC interaction and DNA-binding are the same, it's easy to think that the reason of a higher expression levels of stemness markers in NF-YAl-KO mESCs, after differentiation induction, is due to specific interactors of the two NF-YA isoforms, leading to different regulatory pathways activation. The role of NF-YA as pioneer transcription factor enhancing chromatin accessibility and other TFs DNA-binding was previously described; in particular, it seems to enhance the OCT4/SOX2 DNA-binding [5]. Assuming that the short isoform is the responsible of stemness gene activation, the OCT4/SOX2 DNA-binding could be enhanced by NF-YA short and not by the long one. But if the short isoform is important to express stemness genes, the long one seems to be necessary to overcome the stemness state and to express differential markers by interacting with specific factors and activating the differentiation pathway. This is supported by the data obtained in C2C12 cells in which cells expressing only the short NF-YA protein isoform lost the capability to differentiate into myotubes. NF-YAl protein transfection didn't rescue the phenotype, meaning that the addition of the long protein isoform is not sufficient to undergo differentiation, which instead could require the downregulation or even the absence of the short NF-YA protein.

To answer to this question, NF-YAl protein transfection with the concomitant silencing of the short protein isoform could be performed in NF-YAl-KO cells; under differentiation stimuli, if NF-YAl stimulates the differentiation process while the presence of the short isoform arrests this pathway, a rescued phenotype will be observed.

It would be very interesting to study the specific NF-YA isoforms interactors: by performing a Co-Immunoprecipitation experiment in C2C12 wt and NF-YAl-KO cells it can be easy to distinguish between proteins which specifically interact with one isoform or the other. Moreover, Chip-seq experiment performed in mESCs wt and NF-YAl-KO, before and after differentiation into cardiomyocytes, could highlight the specific genes controlled by the two isoforms. All these experiments can give new and important knowledge about the roles played by NF-YA short and long isoforms during Embryonal development.

6. Materials and Methods

6.1 Cells culture and NF-YA proteins transfection

Mouse Embryonic Stem cells (mESCs, R1 ATCC) were grown in Glasgow minimal essential medium (GMEM) supplemented with 15% fetal bovine serum (FBS), 0.1 mM 2-mercaptoethanol, MEM Non-Essential Amino Acid (1X), 1,000 U/mL mouse leukemia inhibitory factor (LIF), 4 mM L-Glutamine, 100 units/ml penicillin and 100 µg/ml streptomycin, in a humidified 5% CO2 atmosphere at 37°C, plated on dishes pre-treated with 0,1 % gelatin. Mouse myoblast cells (C2C12, ATCC) were cultured in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% fetal bovine serum (FBS), 4 mM L-Glutamine, 100 units/ml penicillin and 100 µg/ml streptomycin, in a humidified 5% CO2 atmosphere at 37°C. Mouse ESCs differentiation was induced by adding 1 µM Retinoic Acid (RA) into the medium (GMEM medium without LIF) or by inducing the EpiLCs differentiation: briefly, cells were plated on dishes pre-treated with 5µg/ml Fibronectin in N2B27 based medium supplemented with 12µg/ml bFGF and 20ng/ml Activin A. C2C12 cells differentiation was induced plating 1×10^5 cells in 35 mm petri dishes in DMEM with 2% horse serum, 4 mM L-Glutamine, 100 units/ml penicillin and 100 µg/ml streptomycin. NF-YA transfections were performed by adding 20 nM of protein into the medium every 24 hours during the differentiation process.

6.2 Plasmids construction and cells transfection

To delete the exon 3 of NF-YA gene in mES and C2C12 cells, four guide RNAs (gRNAs) were designed into the two flanking Introns by using the online tool https://zlab.bio/guide-design-resources and possible off-target sites were assessed by the online tool https://crispr.cos.uni-heidelberg.de. gRNAs without common off-target sites were used to construct the two final plasmids (pX330A_ac, pX330A_bd) following the Multiplex CRISPR/Cas9 Assembly System Kit protocol (Yamamoto lab) [100]. To test transfection efficiency, 3 µg of a GFP plasmid were used to transfect cells by electroporation and Lipofectamine 3000. Finally, $1x10^6$ mES and C2C12 cells were transfected with 3 µg of the two CRISPR plasmids by electroporation and plated at low density in order to have single plated cells. Then single colonies were picked, plated into a 12 multi-well dish and screened by PCR.

6.3 DNA extraction and clones screening

For DNA extraction, cells were washed with PBS, collected by scraping and lysed in 100 µl ice-cold Lysis Buffer (40 mM Tris-HCl, 2mM EDTA, 0.08% SDS, 80 mM NaCl) and incubated overnight at 37°C. 100 µl of ice-cold 2-propanol and 0,3 M NaAc were added, sample were centrifuged and the supernatant was transferred in a new Eppendorf tube. 150 µl of 70% Et-OH were added, sample were centrifuged, pellet was let to dry and resuspended with 30 µl filtered H2O milliQ. Clones were screened by PCR.

6.4 Embryoid Bodies formation

For Embryoid Bodies (EBs) formation, mES cells were used to form 500-cells-drop in EB Differentiation medium (DMEM supplemented with 20% FBS, 0.1 mM 2-mercaptoethanol, MEM Non-Essential Amino Acid (1X), 4 mM L-Glutamine, 100 units/ml penicillin and 100 µg/ml streptomycin) to a concentration of $2x10^4$ cells/mL. 2 days after hanging drop formation, EBs were collected and placed in a new non-adherent bacterial dish and leave to growth for 5 days. EBs were plated in an adherent tissue culture dish coated with 0.1% gelatin and collected after 7 days for analysis.

6.5 Protein extraction and Western Blot Analysis

For total Protein extraction cell pellet was resuspended in ice-cold RPA buffer (10 mM Tris HCl pH 8.0, 1mM EDTA, 0.5 mM EGTA, 0.1% SDS, 1% Triton X-100, 0.1% Deoxycolic acid, 140 mM NaCl, PIC, 1mM PMSF) and incubated for 30 minutes on ice, shaking vigorously occasionally. Samples were centrifuged and the supernatant containing total protein extract was transferred to a new Eppendorf tube. Cell extracts quantification was performed by using the Bradford protein assay.

Western blots of total Protein extracts of mES and C2C12 cells were performed according to standard procedures, with the primary antibodies and a peroxidase-conjugated secondary antibody. Primary antibodies used: anti-NF-YA (G2, Santa Cruz), anti-Tubulin (Sigma), anti-Nanog (Bethyl), anti-NF-YB (GeneSpin), anti-Vinculin (H10, Santa Cruz), anti-Cyclin B1 (Abcam), anti-MyHCs (DHSB MF20), anti-Myog (DHSB, IF5D) and anti-MyoD1 (C-20, Santa Cruz).

6.6 Reverse transcriptase PCR and Real-time PCR

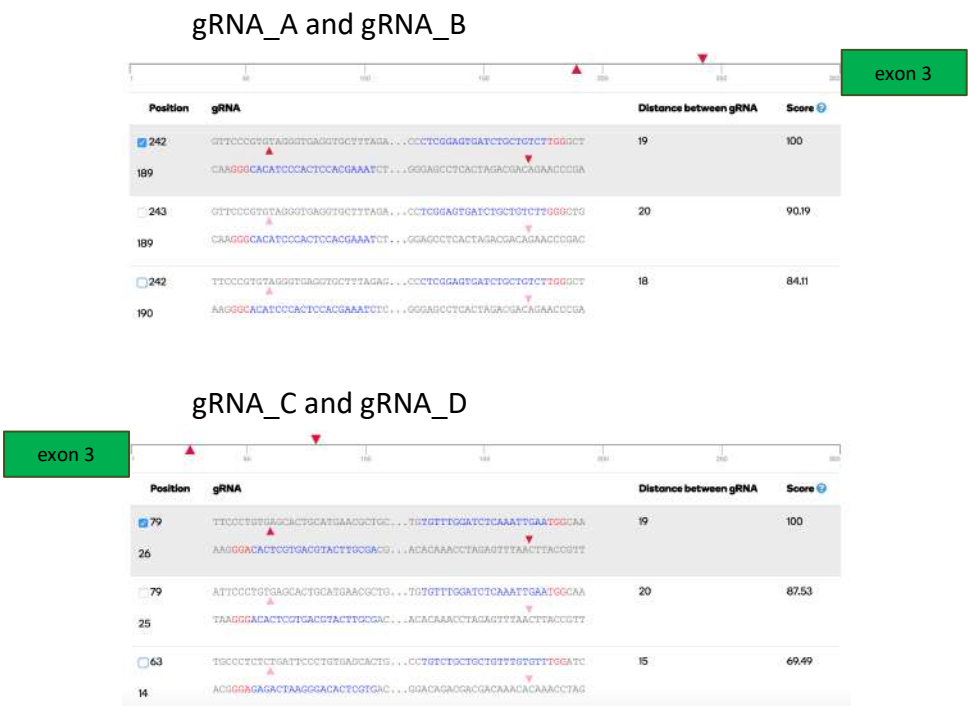
RNA was isolated by the Tri-Reagent (Sigma) protocol according to the manufacturer's instruction. The cDNA was produced starting from 1 µg RNA using the MMLV Reverse Transcription Mix (GeneSpin) and used for Real-Time PCR (SYBR® Green Master Mix, Biorad) analysis.

6.7 Immunofluorescence and Alkaline-phosphatase colouring assay

For immunofluorescence analyses, cells were washed three times with PBS and fixed 10 minutes with 4% paraformaldehyde, washed with PBS and permeabilized 5 minutes with 0,1% Triton X-100. Cells were washed with PBS and incubated 30 minutes with the blocking solution (1% Bovine Serum Albumine, BSA). After two washes with PBS, cells were incubated overnight with the primary antibody at 4°C. Cells were washed three times with PBS and incubated for 2 hours with secondary antibody at room temperature light protected. The acquisition was performed by using the microscope Leicadmi 6000 B.

To perform Alkaline-phosphatase (AP) colouring assay, cells were fixed with 4% paraformaldehyde (PFA) for 1 min at room temperature (RT) and stained for 15 minutes with the Fast Violet B (Sigma) according to the manufacturer's instructions. Cells were then examined under phase contrast microscope.

7. Supplementary Figures



Supplementary Figure 1. gRNAs obtained by using the online tool https://www.atum.bio/eCommerce/cas9/input . The paired gRNAs with higher score were chosen (highlighted with the grey line) to make the NF-YA-exon3 excision.

Off-targets gRNA_A

CCTop - CRISPR/Cas9 target online predictor

CCTop
Help
Supported species
CCTop Q/A forum
CCTop standalone

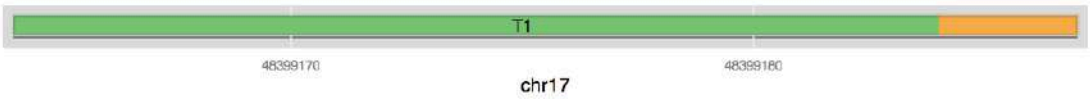
Results for unnamed

Download full results file here.
Download sgRNAs target sites as fasta file here.
Visualize sgRNA target sites in the UCSC genome browser.

Detailed results

Species: Mouse (Mus musculus GRCm38/mm10)
Input: TAAAGCACCTCACCCTACACGGG

PAM:	NGG
Target site length:	20
Target site 5' limitation:	NN
Target site 3' limitation:	NN
Core length:	12
Core MM:	2
Total MM:	4



Legend for off-target site positon: E = exonic; I = intronic; - = intergenic

T1 out of 1
<Previous Next>
Sequence: TAAAGCACCTCACCCTACACGGG
Oligo pair with 5' extension fwd: CACCgTAAAGCACCTCACCCTACAC rev: AAACGTGTAGGGTGAGGTGCTTTAC
Oligo pair with 5' substitution fwd: CACCgAAAGCACCTCACCCTACAC rev: AAACGTGTAGGGTGAGGTGCTTTC
Top 20 offtarget sites out of 64 (including on target; for full list see xls file)

Coordinates	strand	MM	target_seq	PAM	distance		gene name	gene id
chr17:48399164-48399186	+	0	TAAAGCAC[CTCACCCTACAC]	GGG	0	E	Nfya	ENSMUSG00000023994
chr1:114958772-114958794	-	4	TATTGCTC[TTCACCCTACAC]	TGG	NA	-	NA	NA
chr14:85582373-85582395	+	2	TACAGCAC[CTCACACTACAC]	AGG	NA	-	NA	NA
chr18:38693594-38693616	+	4	TGCAGCAG[CTGACCCTACAC]	TGG	14370	I	Arhgap26	ENSMUSG00000036452
chr19:36225855-36225877	-	4	AAAAGCAG[TCCACCCTACAC]	CGG	NA	-	NA	NA
chr16:94262194-94262216	+	4	TAAAGGTC[AACACCCTACAC]	TGG	376	I	Hlcs	ENSMUSG00000040820
chr13:59170724-59170746	-	2	TAAAGCAC[CTTAGCCTACAC]	TGG	36754	-	Ntrk2	ENSMUSG00000055254
chr19:54416335-54416357	+	4	TTAAGAAC[CACTCCCTACAC]	AGG	NA	-	NA	NA
chr15:85568665-85568687	-	4	TCAACCAC[CCCAACCTACAC]	AGG	8821	I	Wnt7b	ENSMUSG00000022382
chr1:134836076-134836098	-	4	TCAAGACC[CTCACTCTACAC]	TGG	0	E	Ppp1r12b	ENSMUSG00000073557
chr1:63812603-63812625	+	4	CAAAACAA[CTCACCTTACAC]	TGG	32202	-	Gm13751	ENSMUSG00000080817
chr15:34930825-34930847	+	3	TCAAGCAA[CTCACCCAACAC]	TGG	14159	I	Stk3	ENSMUSG00000022329
chr3:93395141-93395163	+	4	TTAAGTAC[CTAACTCTACAC]	TGG	25	I	Rptn	ENSMUSG00000041984
chr2:69159581-69159603	-	4	CAAAGCAA[CTCCTCCTACAC]	AGG	750	I	Nostrin	ENSMUSG00000034738
chr13:102586342-102586364	-	4	AAAGGCAT[CTCACCCAACAC]	CGG	NA	-	NA	NA
chr7:57378047-57378069	+	4	AGTAGCAC[CTCACCCTTCAC]	AGG	3520	I	Gabrg3	ENSMUSG00000055026
chr2:127515982-127516004	+	4	TACAGTTC[CTCACCCAACAC]	AGG	4988	I	Kcnip3	ENSMUSG00000079056
chr16:9859944-9859966	+	4	TAAGGCAA[GTCACCATACAC]	TGG	62611	I	Grin2a	ENSMUSG00000059003
chr10:121375091-121375113	-	4	TGAGGCAC[CACACCCGACAC]	AGG	527	I	Gns	ENSMUSG00000034707
chr11:121208779-121208801	-	4	TAAGGCTG[CTCACCCCACAC]	GGG	1004	I	Hexdc	ENSMUSG00000039307

Off-targets gRNA_B

CCTop - CRISPR/Cas9 target online predictor

Centre for Organismal Studies Heidelberg

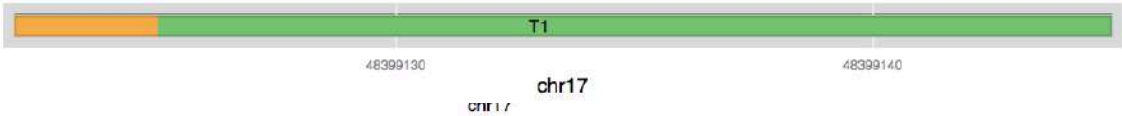
CCTop
Help
Supported species
CCTop Q/A forum
CCTop standalone

Results for unnamed

Download full results file here.
Download sgRNAs target sites as fasta file here.
Visualize sgRNA target sites in the UCSC genome browser.

Detailed results

Species: Mouse (Mus musculus GRCm38/mm10) PAM: NGG
Input: CCAAGACAGCAGATCACTCCGAG Target site length: 20
Target site 5' limitation: NN
Target site 3' limitation: NN
Core length: 12
Core MM: 2
Total MM: 4



Legend for off-target site positon: E = exonic; I = intronic; - = intergenic

T1 out of 1
<Previous Next>
Sequence: CTCGGAGTGATCTGCTGTCTTGG
Oligo pair with 5' extension fwd: CACCgCTCGGAGTGATCTGCTGTCT rev: AAACAGACAGCAGATCACTCCGAGC
Oligo pair with 5' substitution fwd: CACCgTCGGAGTGATCTGCTGTCT rev: AAACAGACAGCAGATCACTCCGAC
Top 20 offtarget sites out of 60 (including on target; for full list see xls file)

Coordinates	strand	MM	target_seq	PAM	distance		gene name	gene id
chr17:48399122-48399144	-	0	CTCGGAGT[GATCTGCTGTCT]	TGG	0	E	Nfya	ENSMUSG00000023994
chr11:63054622-63054644	-	4	CAGTGTGT[GATCTGCTGTCT]	AGG	7010	-	Tekt3	ENSMUSG00000042189
chr13:92903642-92903664	+	4	CCAAGAAT[GATCTGCTGTCT]	GGG	45412	-	Mtx3	ENSMUSG00000021704
chr18:35369451-35369473	-	3	CTGAGAGT[GAGCTGCTGTCT]	TGG	20691	I	Sil1	ENSMUSG00000024357
chr5:70323486-70323508	-	3	CTCTGAGA[AATCTGCTGTCT]	TGG	NA	-	NA	NA
chr6:90455521-90455543	-	4	ACAGGAGT[GAACTGCTGTCT]	GGG	4168	-	RP23-370A2.6	ENSMUSG00000107605
chr15:81820302-81820324	+	4	CTCAGCAA[GATCTGCTGTCT]	TGG	2441	-	Tef	ENSMUSG00000022389
chr11:75672256-75672278	+	4	GTGGGAGG[GACCTGCTGTCT]	TGG	5	I	Myo1c	ENSMUSG00000017774
chr3:37530656-37530678	+	4	GAGGGAGT[GATCAGCTGTCT]	TGG	637	I	Spata5	ENSMUSG00000027722
chr14:24643214-24643236	-	4	CTGGGTCT[GAGCTGCTGTCT]	GGG	14014	-	4930428N03Rik	ENSMUSG00000089887
chr9:14519451-14519473	-	4	CAATGAGT[GATCAGCTGTCT]	GGG	8874	-	Cwc15	ENSMUSG00000004096
chr6:67915628-67915650	-	4	GTCGGATT[ATTCTGCTGTCT]	AGG	1576	-	Igkv11-125	ENSMUSG00000095737
chr16:88144466-88144488	+	3	CACGGAGT[GCTGTGCTGTCT]	TGG	88290	-	Grik1	ENSMUSG00000022935
chr9:98945648-98945670	-	4	CTAAGAGG[GATATGCTGTCT]	TGG	4111	-	Foxl2os	ENSMUSG00000097072
chr18:74621923-74621945	+	4	CTGGGAGG[ATTCTGCTGTCT]	GGG	3617	I	Myo5b	ENSMUSG00000025885
chr2:52767634-52767656	-	4	CTGGGAGC[TCTCTGCTGTCT]	TGG	25353	-	Stam2	ENSMUSG00000055371
chr15:85532178-85532200	+	4	CCAGGAGG[GATCAGCTGTCT]	AGG	3239	-	Wnt7b	ENSMUSG00000022382
chr12:99689239-99689261	+	4	TTCGGAGG[GCACTGCTGTCT]	GGG	28270	-	Efcab11	ENSMUSG00000021176
chr12:103088566-103088588	+	4	CTTGGAAT[TATTTGCTGTCT]	TGG	35	I	Unc79	ENSMUSG00000021198
chr1:113306774-113306796	+	4	CTAGGAGG[AATTTGCTGTCT]	AGG	16409	-	Gm28189	ENSMUSG00000101453

Off-targets gRNA_C

CCTop - CRISPR/Cas9 target online predictor

Centre for Organismal Studies Heidelberg

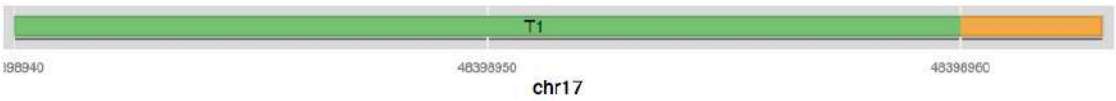
CCTop
Help
Supported species
CCTop Q/A forum
CCTop standalone

Results for unnamed

Download full results file here.
Download sgRNAs target sites as fasta file here.
Visualize sgRNA target sites in the UCSC genome browser.

Detailed results

Species: Mouse (Mus musculus GRCm38/mm10) PAM: NGG
Input: AGCGTTCATGCAGTGCTCACAGG Target site length: 20
Target site 5' limitation: NN
Target site 3' limitation: NN
Core length: 12
Core MM: 2
Total MM: 4

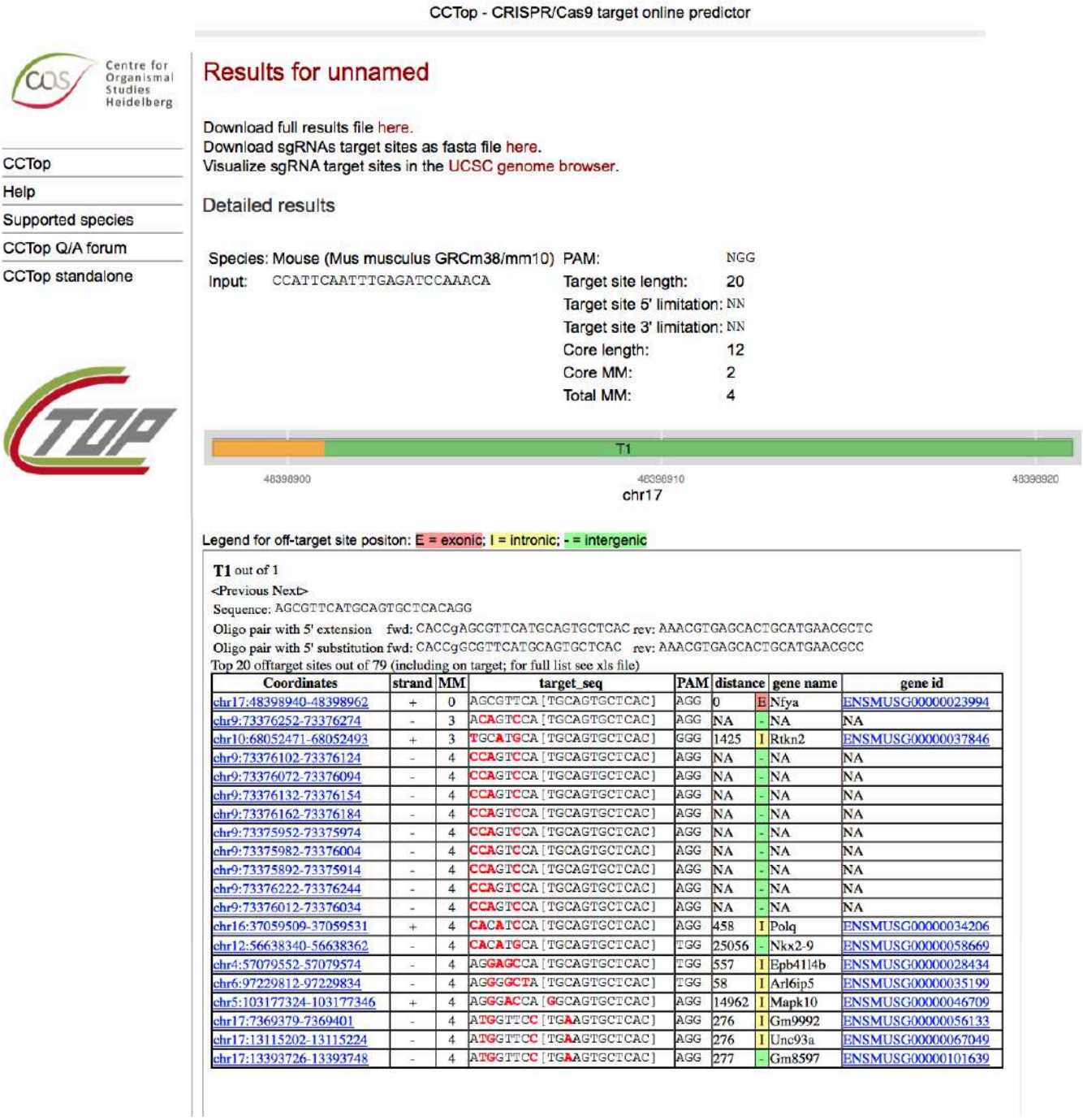


Legend for off-target site positon: E = exonic; I = intronic; - = intergenic

T1 out of 1
<Previous Next>
Sequence: AGCGTTCATGCAGTGCTCACAGG
Oligo pair with 5' extension fwd: CACCgAGCGTTCATGCAGTGCTCAC rev: AAACGTGAGCACTGCATGAACGCTC
Oligo pair with 5' substitution fwd: CACCgGCGTTCATGCAGTGCTCAC rev: AAACGTGAGCACTGCATGAACGCC
Top 20 offtarget sites out of 79 (including on target; for full list see xls file)

Coordinates	strand	MM	target_seq	PAM	distance		gene name	gene id
chr17:48398940-48398962	+	0	AGCGTTCA[TGCAGTGCTCAC]	AGG	0	E	Nfya	ENSMUSG00000023994
chr9:73376252-73376274	-	3	ACAGTCCA[TGCAGTGCTCAC]	AGG	NA	-	NA	NA
chr10:68052471-68052493	+	3	TGCATGCA[TGCAGTGCTCAC]	GGG	1425	I	Rtkn2	ENSMUSG00000037846
chr9:73376102-73376124	-	4	CCAGTCCA[TGCAGTGCTCAC]	AGG	NA	-	NA	NA
chr9:73376072-73376094	-	4	CCAGTCCA[TGCAGTGCTCAC]	AGG	NA	-	NA	NA
chr9:73376132-73376154	-	4	CCAGTCCA[TGCAGTGCTCAC]	AGG	NA	-	NA	NA
chr9:73376162-73376184	-	4	CCAGTCCA[TGCAGTGCTCAC]	AGG	NA	-	NA	NA
chr9:73375952-73375974	-	4	CCAGTCCA[TGCAGTGCTCAC]	AGG	NA	-	NA	NA
chr9:73375982-73376004	-	4	CCAGTCCA[TGCAGTGCTCAC]	AGG	NA	-	NA	NA
chr9:73375892-73375914	-	4	CCAGTCCA[TGCAGTGCTCAC]	AGG	NA	-	NA	NA
chr9:73376222-73376244	-	4	CCAGTCCA[TGCAGTGCTCAC]	AGG	NA	-	NA	NA
chr9:73376012-73376034	-	4	CCAGTCCA[TGCAGTGCTCAC]	AGG	NA	-	NA	NA
chr16:37059509-37059531	+	4	CACATCCA[TGCAGTGCTCAC]	AGG	458	I	Polq	ENSMUSG00000034206
chr12:56638340-56638362	-	4	CACATGCA[TGCAGTGCTCAC]	TGG	25056	-	Nkx2-9	ENSMUSG00000058669
chr4:57079552-57079574	-	4	AGGAGCCA[TGCAGTGCTCAC]	TGG	557	I	Epb41l4b	ENSMUSG00000028434
chr6:97229812-97229834	-	4	AGGGGCTA[TGCAGTGCTCAC]	TGG	58	I	Arl6ip5	ENSMUSG00000035199
chr5:103177324-103177346	+	4	AGGGACCA[GGCAGTGCTCAC]	AGG	14962	I	Mapk10	ENSMUSG00000046709
chr17:7369379-7369401	-	4	ATGGTTCC[TGAAGTGCTCAC]	AGG	276	I	Gm9992	ENSMUSG00000056133
chr17:13115202-13115224	-	4	ATGGTTCC[TGAAGTGCTCAC]	AGG	276	I	Unc93a	ENSMUSG00000067049
chr17:13393726-13393748	-	4	ATGGTTCC[TGAAGTGCTCAC]	AGG	277	-	Gm8597	ENSMUSG00000101639

Off-targets gRNA_D



Coordinates	strand	MM	target_seq	PAM	distance		gene name	gene id
chr17:48398940-48398962	+	0	AGCGTTCA[TGCAGTGCTCAC]	AGG	0	E	Nfya	ENSMUSG00000023994
chr9:73376252-73376274	-	3	ACAGTCCA[TGCAGTGCTCAC]	AGG	NA	-	NA	NA
chr10:68052471-68052493	+	3	TGCATGCA[TGCAGTGCTCAC]	GGG	1425	I	Rtkn2	ENSMUSG00000037846
chr9:73376102-73376124	-	4	CCAGTCCA[TGCAGTGCTCAC]	AGG	NA	-	NA	NA
chr9:73376072-73376094	-	4	CCAGTCCA[TGCAGTGCTCAC]	AGG	NA	-	NA	NA
chr9:73376132-73376154	-	4	CCAGTCCA[TGCAGTGCTCAC]	AGG	NA	-	NA	NA
chr9:73376162-73376184	-	4	CCAGTCCA[TGCAGTGCTCAC]	AGG	NA	-	NA	NA
chr9:73375952-73375974	-	4	CCAGTCCA[TGCAGTGCTCAC]	AGG	NA	-	NA	NA
chr9:73375982-73376004	-	4	CCAGTCCA[TGCAGTGCTCAC]	AGG	NA	-	NA	NA
chr9:73375892-73375914	-	4	CCAGTCCA[TGCAGTGCTCAC]	AGG	NA	-	NA	NA
chr9:73376222-73376244	-	4	CCAGTCCA[TGCAGTGCTCAC]	AGG	NA	-	NA	NA
chr9:73376012-73376034	-	4	CCAGTCCA[TGCAGTGCTCAC]	AGG	NA	-	NA	NA
chr16:37059509-37059531	+	4	CACATCCA[TGCAGTGCTCAC]	AGG	458	I	Polq	ENSMUSG00000034206
chr12:56638340-56638362	-	4	CACATGCA[TGCAGTGCTCAC]	TGG	25056	-	Nkx2-9	ENSMUSG00000058669
chr4:57079552-57079574	-	4	AGGAGCCA[TGCAGTGCTCAC]	TGG	557	I	Epb41l4b	ENSMUSG00000028434
chr6:97229812-97229834	-	4	AGGGGCTA[TGCAGTGCTCAC]	TGG	58	I	Arl6ip5	ENSMUSG00000035199
chr5:103177324-103177346	+	4	AGGGACCA[GGCAGTGCTCAC]	AGG	14962	I	Mapk10	ENSMUSG00000046709
chr17:7369379-7369401	-	4	ATGGTTCC[TGAAGTGCTCAC]	AGG	276	I	Gm9992	ENSMUSG00000056133
chr17:13115202-13115224	-	4	ATGGTTCC[TGAAGTGCTCAC]	AGG	276	I	Unc93a	ENSMUSG00000067049
chr17:13393726-13393748	-	4	ATGGTTCC[TGAAGTGCTCAC]	AGG	277	-	Gm8597	ENSMUSG00000101639

Supplementary figure 2. List of possible off-targets for each single gRNA. The possible exonic/intronic/intergenic region and the number of mismatches is shown.

8. Bibliography

[1] K. Kurrikoff, M. Gestin, and Ü. Langel, 'Recent in vivo advances in cell-penetrating peptide-assisted drug delivery.', *Expert Opin. Drug Deliv.*, vol. 13, no. 3, pp. 373–87, 2016.

[2] V. Nardone, A. Chaves-Sanjuan, and M. Nardini, 'Structural determinants for NF-Y/DNA interaction at the CCAAT box.', *Biochim. Biophys. acta. Gene Regul. Mech.*, vol. 1860, no. 5, pp. 571–580, May 2017.

[3] F. Lu, Y. Liu, A. Inoue, T. Suzuki, K. Zhao, and Y. Zhang, 'Establishing Chromatin Regulatory Landscape during Mouse Preimplantation Development.', *Cell*, vol. 165, no. 6, pp. 1375–1388, Jun. 2016.

[4] J. Zhu, Y. Zhang, G. J. Joe, R. Pompetti, and S. G. Emerson, 'NF-Ya activates multiple hematopoietic stem cell (HSC) regulatory genes and promotes HSC self-renewal.', *Proc. Natl. Acad. Sci. U. S. A.*, vol. 102, no. 33, pp. 11728–33, Aug. 2005.

[5] A. J. Oldfield *et al.*, 'Histone-fold domain protein NF-Y promotes chromatin accessibility for cell type-specific master transcription factors.', *Mol. Cell*, vol. 55, no. 5, pp. 708–22, Sep. 2014.

[6] M. Grskovic, C. Chaivorapol, A. Gaspar-Maia, H. Li, and M. Ramalho-Santos, 'Systematic identification of cis-regulatory sequences active in mouse and human embryonic stem cells.', *PLoS Genet.*, vol. 3, no. 8, p. e145, Aug. 2007.

[7] V. Basile *et al.*, 'NF-YA splice variants have different roles on muscle differentiation', *Biochim. Biophys. Acta - Gene Regul. Mech.*, vol. 1859, no. 4, pp. 627–638, Apr. 2016.

[8] A. L. Olins and D. E. Olins, 'Spheroid chromatin units (v bodies).', *Science*, vol. 183, no. 4122, pp. 330–2, Jan. 1974.

[9] T. Tsukiyama and C. Wu, 'Purification and properties of an ATP-dependent nucleosome remodeling factor.', *Cell*, vol. 83, no. 6, pp. 1011–20, Dec. 1995.

[10] I. Versteege *et al.*, 'Truncating mutations of hSNF5/INI1 in aggressive paediatric cancer.', *Nature*, vol. 394, no. 6689, pp. 203–6, Jul. 1998.

[11] M. Rodríguez-Paredes and M. Esteller, 'Cancer epigenetics reaches mainstream oncology.', *Nat. Med.*, vol. 17, no. 3, pp. 330–9, Mar. 2011.

[12] G. Orphanides, T. Lagrange, and D. Reinberg, 'The general transcription factors of RNA polymerase II.', *Genes Dev.*, vol. 10, no. 21, pp. 2657–83, Nov. 1996.

[13] S. K. Burley and R. G. Roeder, 'BIOCHEMISTRY AND STRUCTURAL BIOLOGY OF TRANSCRIPTION FACTOR IID (TFIID)', *Annu. Rev. Biochem.*, vol. 65, no. 1, pp. 769–799, Jun. 1996.

[14] R. Tjian and T. Maniatis, 'Transcriptional activation: a complex puzzle with few easy pieces.', *Cell*, vol. 77, no. 1, pp. 5–8, Apr. 1994.

[15] S. Malik and R. G. Roeder, 'Transcriptional regulation through Mediator-like coactivators in yeast and metazoan cells.', *Trends Biochem. Sci.*, vol. 25, no. 6, pp. 277–83, Jun. 2000.

[16] A. M. Näär, B. D. Lemon, and R. Tjian, 'Transcriptional coactivator complexes.', *Annu. Rev. Biochem.*, vol. 70, pp. 475–501, 2001.

[17] T. Jenuwein and C. D. Allis, 'Translating the histone code.', *Science*, vol. 293, no. 5532, pp. 1074–80, Aug. 2001.

[18] K. A. Jones and J. T. Kadonaga, 'Exploring the transcription-chromatin interface.', *Genes Dev.*, vol. 14, no. 16, pp. 1992–6, Aug. 2000.

[19] A. H. Brivanlou and J. E. Darnell, 'Signal transduction and the control of gene expression.', *Science*, vol. 295, no. 5556, pp. 813–8, Feb. 2002.

[20] L. A. Chodosh, J. Olesen, S. Hahn, A. S. Baldwin, L. Guarente, and P. A. Sharp, 'A yeast and a human CCAAT-binding protein have heterologous subunits that are functionally interchangeable.', *Cell*, vol. 53, no. 1, pp. 25–35, Apr. 1988.

[21] P. Bucher and E. N. Trifonov, 'CCAAT box revisited: bidirectionality, location and context.', *J. Biomol. Struct. Dyn.*, vol. 5, no. 6, pp. 1231–6, Jun. 1988.

[22] C. Benoist, K. O'Hare, R. Breathnach, and P. Chambon, 'The ovalbumin gene-sequence of putative control regions.', *Nucleic Acids Res.*, vol. 8, no. 1, pp. 127–42, Jan. 1980.

[23] P. Bucher, 'Weight matrix descriptions of four eukaryotic RNA polymerase II promoter elements derived from 502 unrelated promoter sequences.', *J. Mol. Biol.*, vol. 212, no. 4, pp. 563–78, Apr. 1990.

[24] R. Mantovani, 'The molecular biology of the CCAAT-binding factor NF-Y.', *Gene*, vol. 239, no. 1, pp. 15–27, Oct. 1999.

[25] A. Testa *et al.*, 'Chromatin immunoprecipitation (ChIP) on chip experiments uncover a widespread distribution of NF-Y binding CCAAT sites outside of core promoters.', *J. Biol. Chem.*, vol. 280, no. 14, pp. 13606–15, Apr. 2005.

[26] R. Mantovani, 'A survey of 178 NF-Y binding CCAAT boxes.', *Nucleic Acids Res.*, vol. 26,

no. 5, pp. 1135–43, Mar. 1998.

[27] Y.-C. Lin, Y.-N. Chen, K.-F. Lin, F.-F. Wang, T.-Y. Chou, and M.-Y. Chen, 'Association of p21 with NF-YA suppresses the expression of Polo-like kinase 1 and prevents mitotic death in response to DNA damage.', *Cell Death Dis.*, vol. 5, p. e987, Jan. 2014.

[28] F. Bolognese *et al.*, 'The cyclin B2 promoter depends on NF-Y, a trimer whose CCAAT-binding activity is cell-cycle regulated.', *Oncogene*, vol. 18, no. 10, pp. 1845–53, Mar. 1999.

[29] S. Osada, H. Yamamoto, T. Nishihara, and M. Imagawa, 'DNA binding specificity of the CCAAT/enhancer-binding protein transcription factor family.', *J. Biol. Chem.*, vol. 271, no. 7, pp. 3891–6, Feb. 1996.

[30] H. Zorbas, T. Rein, A. Krause, K. Hoffmann, and E. L. Winnacker, 'Nuclear factor I (NF I) binds to an NF I-type site but not to the CCAAT site in the human alpha-globin gene promoter.', *J. Biol. Chem.*, vol. 267, no. 12, pp. 8478–84, Apr. 1992.

[31] A. Dorn, J. Bollekens, A. Staub, C. Benoist, and D. Mathis, 'A multiplicity of CCAAT box-binding proteins.', *Cell*, vol. 50, no. 6, pp. 863–72, Sep. 1987.

[32] X. Y. Li, R. Mantovani, R. Hooft van Huijsduijnen, I. Andre, C. Benoist, and D. Mathis, 'Evolutionary variation of the CCAAT-binding transcription factor NF-Y.', *Nucleic Acids Res.*, vol. 20, no. 5, pp. 1087–91, Mar. 1992.

[33] F. Coustry, S. N. Maity, and B. de Crombrugghe, 'Studies on transcription activation by the multimeric CCAAT-binding factor CBF.', *J. Biol. Chem.*, vol. 270, no. 1, pp. 468–75, Jan. 1995.

[34] M. Nardini *et al.*, 'Sequence-specific transcription factor NF-Y displays histone-like DNA binding and H2B-like ubiquitination.', *Cell*, vol. 152, no. 1–2, pp. 132–43, Jan. 2013.

[35] J. Zwicker *et al.*, 'Cell cycle regulation of cdc25C transcription is mediated by the periodic repression of the glutamine-rich activators NF-Y and Sp1.', *Nucleic Acids Res.*, vol. 23, no. 19, pp. 3822–30, Oct. 1995.

[36] S. N. Maity and B. de Crombrugghe, 'Role of the CCAAT-binding protein CBF/NF-Y in transcription.', *Trends Biochem. Sci.*, vol. 23, no. 5, pp. 174–8, May 1998.

[37] M. Frontini *et al.*, 'NF-Y recruitment of TFIID, multiple interactions with histone fold TAF(II)s.', *J. Biol. Chem.*, vol. 277, no. 8, pp. 5841–8, Feb. 2002.

[38] H. Izumi, C. Molander, L. Z. Penn, A. Ishisaki, K. Kohno, and K. Funa, 'Mechanism for

the transcriptional repression by c-Myc on PDGF beta-receptor.', *J. Cell Sci.*, vol. 114, no. Pt 8, pp. 1533–44, Apr. 2001.

[39] D. Dolfini, M. Minuzzo, G. Pavesi, and R. Mantovani, 'The short isoform of NF-YA belongs to the embryonic stem cell transcription factor circuitry.', *Stem Cells*, vol. 30, no. 11, pp. 2450–9, Nov. 2012.

[40] M. Ceribelli, P. Benatti, C. Imbriano, and R. Mantovani, 'NF-YC complexity is generated by dual promoters and alternative splicing.', *J. Biol. Chem.*, vol. 284, no. 49, pp. 34189–200, Dec. 2009.

[41] Z. F. Chang and C. J. Liu, 'Human thymidine kinase CCAAT-binding protein is NF-Y, whose A subunit expression is serum-dependent in human IMR-90 diploid fibroblasts.', *J. Biol. Chem.*, vol. 269, no. 27, pp. 17893–8, Jul. 1994.

[42] M.-R. Lee *et al.*, 'Transcription factors NF-YA regulate the induction of human OGG1 following DNA-alkylating agent methylmethane sulfonate (MMS) treatment.', *J. Biol. Chem.*, vol. 279, no. 11, pp. 9857–66, Mar. 2004.

[43] A. Farina *et al.*, 'Down-regulation of cyclin B1 gene transcription in terminally differentiated skeletal muscle cells is associated with loss of functional CCAAT-binding NF-Y complex.', *Oncogene*, vol. 18, no. 18, pp. 2818–27, May 1999.

[44] A. Gurtner *et al.*, 'Requirement for down-regulation of the CCAAT-binding activity of the NF-Y transcription factor during skeletal muscle differentiation.', *Mol. Biol. Cell*, vol. 14, no. 7, pp. 2706–15, Jul. 2003.

[45] I. Manni *et al.*, 'Posttranslational regulation of NF-YA modulates NF-Y transcriptional activity.', *Mol. Biol. Cell*, vol. 19, no. 12, pp. 5203–13, Dec. 2008.

[46] A. Carr and M. D. Biggin, 'A comparison of in vivo and in vitro DNA-binding specificities suggests a new model for homeoprotein DNA binding in Drosophila embryos.', *EMBO J.*, vol. 18, no. 6, pp. 1598–608, Mar. 1999.

[47] R. Joseph *et al.*, 'Integrative model of genomic factors for determining binding site selection by estrogen receptor-α', *Mol. Syst. Biol.*, vol. 6, no. 1, p. 456, Jan. 2010.

[48] C. C. Adams and J. L. Workman, 'Binding of disparate transcriptional activators to nucleosomal DNA is inherently cooperative.', *Mol. Cell. Biol.*, vol. 15, no. 3, pp. 1405–21, Mar. 1995.

[49] N. L. V. Máté Pálfy, Gunnar Schulze, Eivind Valen, 'Chromatin accessibility established by Pou5f3, Sox19b and Nanog primes genes for activity during zebrafish genome

activation', *Biorxiv*, 2019.

[50] R. I. Sherwood *et al.*, 'Discovery of directional and nondirectional pioneer transcription factors by modeling DNase profile magnitude and shape.', *Nat. Biotechnol.*, vol. 32, no. 2, pp. 171–178, Feb. 2014.

[51] A. G. Smith *et al.*, 'Inhibition of pluripotential embryonic stem cell differentiation by purified polypeptides.', *Nature*, vol. 336, no. 6200, pp. 688–90, Dec. 1988.

[52] D. V. Do *et al.*, 'A genetic and developmental pathway from STAT3 to the OCT4-NANOG circuit is essential for maintenance of ICM lineages in vivo.', *Genes Dev.*, vol. 27, no. 12, pp. 1378–90, Jun. 2013.

[53] I. Chambers *et al.*, 'Functional expression cloning of Nanog, a pluripotency sustaining factor in embryonic stem cells.', *Cell*, vol. 113, no. 5, pp. 643–55, May 2003.

[54] I. Chambers and A. Smith, 'Self-renewal of teratocarcinoma and embryonic stem cells', *Oncogene*, vol. 23, no. 43, pp. 7150–7160, Sep. 2004.

[55] D. C. Ambrosetti, H. R. Schöler, L. Dailey, and C. Basilico, 'Modulation of the activity of multiple transcriptional activation domains by the DNA binding domains mediates the synergistic action of Sox2 and Oct-3 on the fibroblast growth factor-4 enhancer.', *J. Biol. Chem.*, vol. 275, no. 30, pp. 23387–97, Jul. 2000.

[56] S. Masui *et al.*, 'Pluripotency governed by Sox2 via regulation of Oct3/4 expression in mouse embryonic stem cells.', *Nat. Cell Biol.*, vol. 9, no. 6, pp. 625–35, Jun. 2007.

[57] S. Yamanaka and H. M. Blau, 'Nuclear reprogramming to a pluripotent state by three approaches.', *Nature*, vol. 465, no. 7299, pp. 704–12, Jun. 2010.

[58] I. Chambers *et al.*, 'Nanog safeguards pluripotency and mediates germline development.', *Nature*, vol. 450, no. 7173, pp. 1230–4, Dec. 2007.

[59] A. Marson *et al.*, 'Connecting microRNA genes to the core transcriptional regulatory circuitry of embryonic stem cells.', *Cell*, vol. 134, no. 3, pp. 521–33, Aug. 2008.

[60] L. A. Boyer *et al.*, 'Core transcriptional regulatory circuitry in human embryonic stem cells.', *Cell*, vol. 122, no. 6, pp. 947–56, Sep. 2005.

[61] M. H. Kagey *et al.*, 'Mediator and cohesin connect gene expression and chromatin architecture.', *Nature*, vol. 467, no. 7314, pp. 430–5, Sep. 2010.

[62] P. B. Rahl *et al.*, 'c-Myc regulates transcriptional pause release.', *Cell*, vol. 141, no. 3, pp. 432–45, Apr. 2010.

[63] Y. Huang, R. Osorno, A. Tsakiridis, and V. Wilson, 'In Vivo differentiation potential of

epiblast stem cells revealed by chimeric embryo formation.', *Cell Rep.*, vol. 2, no. 6, pp. 1571–8, Dec. 2012.

[64] P. J. Tesar *et al.*, 'New cell lines from mouse epiblast share defining features with human embryonic stem cells.', *Nature*, vol. 448, no. 7150, pp. 196–9, Jul. 2007.

[65] Q.-L. Ying *et al.*, 'The ground state of embryonic stem cell self-renewal.', *Nature*, vol. 453, no. 7194, pp. 519–23, May 2008.

[66] J. Betschinger, J. Nichols, S. Dietmann, P. D. Corrin, P. J. Paddison, and A. Smith, 'Exit from pluripotency is gated by intracellular redistribution of the bHLH transcription factor Tfe3.', *Cell*, vol. 153, no. 2, pp. 335–47, Apr. 2013.

[67] G. Guo *et al.*, 'Klf4 reverts developmentally programmed restriction of ground state pluripotency.', *Development*, vol. 136, no. 7, pp. 1063–9, Apr. 2009.

[68] C. Buecker *et al.*, 'Reorganization of enhancer patterns in transition from naive to primed pluripotency.', *Cell Stem Cell*, vol. 14, no. 6, pp. 838–53, Jun. 2014.

[69] S.-H. Yang *et al.*, 'Otx2 and Oct4 drive early enhancer activation during embryonic stem cell transition from naive pluripotency.', *Cell Rep.*, vol. 7, no. 6, pp. 1968–81, Jun. 2014.

[70] S. N. Maity, 'NF-Y (CBF) regulation in specific cell types and mouse models.', *Biochim. Biophys. acta. Gene Regul. Mech.*, vol. 1860, no. 5, pp. 598–603, May 2017.

[71] A. Stathopoulos and M. Levine, 'Dorsal gradient networks in the Drosophila embryo.', *Dev. Biol.*, vol. 246, no. 1, pp. 57–67, Jun. 2002.

[72] A. D. Domashenko, G. Danet-Desnoyers, A. Aron, M. P. Carroll, and S. G. Emerson, 'TAT-mediated transduction of NF-Ya peptide induces the ex vivo proliferation and engraftment potential of human hematopoietic progenitor cells.', *Blood*, vol. 116, no. 15, pp. 2676–83, Oct. 2010.

[73] S. J. Tapscott, 'The circuitry of a master switch: Myod and the regulation of skeletal muscle gene transcription.', *Development*, vol. 132, no. 12, pp. 2685–95, Jun. 2005.

[74] M. E. Massari and C. Murre, 'Helix-loop-helix proteins: regulators of transcription in eucaryotic organisms.', *Mol. Cell. Biol.*, vol. 20, no. 2, pp. 429–40, Jan. 2000.

[75] T. H. Smith, N. E. Block, S. J. Rhodes, S. F. Konieczny, and J. B. Miller, 'A unique pattern of expression of the four muscle regulatory factor proteins distinguishes somitic from embryonic, fetal and newborn mouse myogenic cells.', *Development*, vol. 117, no. 3, pp. 1125–33, Mar. 1993.

[76] T. Braun, M. A. Rudnicki, H. H. Arnold, and R. Jaenisch, 'Targeted inactivation of the muscle regulatory gene Myf-5 results in abnormal rib development and perinatal death.', *Cell*, vol. 71, no. 3, pp. 369–82, Oct. 1992.

[77] M. A. Rudnicki, T. Braun, S. Hinuma, and R. Jaenisch, 'Inactivation of MyoD in mice leads to up-regulation of the myogenic HLH gene Myf-5 and results in apparently normal muscle development.', *Cell*, vol. 71, no. 3, pp. 383–90, Oct. 1992.

[78] P. Hasty *et al.*, 'Muscle deficiency and neonatal death in mice with a targeted mutation in the myogenin gene.', *Nature*, vol. 364, no. 6437, pp. 501–6, Aug. 1993.

[79] J. D. Molkentin and E. N. Olson, 'Combinatorial control of muscle development by basic helix-loop-helix and MADS-box transcription factors.', *Proc. Natl. Acad. Sci. U. S. A.*, vol. 93, no. 18, pp. 9366–73, Sep. 1996.

[80] Y. Liu, A. Chu, I. Chakroun, U. Islam, and A. Blais, 'Cooperation between myogenic regulatory factors and SIX family transcription factors is important for myoblast differentiation.', *Nucleic Acids Res.*, vol. 38, no. 20, pp. 6857–71, Nov. 2010.

[81] J. D. Molkentin, B. L. Black, J. F. Martin, and E. N. Olson, 'Cooperative activation of muscle gene expression by MEF2 and myogenic bHLH proteins.', *Cell*, vol. 83, no. 7, pp. 1125–36, Dec. 1995.

[82] A. D. Frankel and C. O. Pabo, 'Cellular uptake of the tat protein from human immunodeficiency virus.', *Cell*, vol. 55, no. 6, pp. 1189–93, Dec. 1988.

[83] V. E. Ott M, Dorr A, Hetzer-Egger C, Kaehlcke K, Schnolzer M, Henklein P, Cole P, Zhou MM, 'Tat acetylation: a regulatory switch between early and late phases in HIV transcription elongation', *Novartis Found Symp*, 2004.

[84] J. Park *et al.*, 'Mutational analysis of a human immunodeficiency virus type 1 Tat protein transduction domain which is required for delivery of an exogenous protein into mammalian cells.', *J. Gen. Virol.*, vol. 83, no. Pt 5, pp. 1173–81, May 2002.

[85] C. Bechara and S. Sagan, 'Cell-penetrating peptides: 20 years later, where do we stand?', *FEBS Lett.*, vol. 587, no. 12, pp. 1693–702, Jun. 2013.

[86] A. Bolhassani, B. S. Jafarzade, and G. Mardani, 'In vitro and in vivo delivery of therapeutic proteins using cell penetrating peptides.', *Peptides*, vol. 87, pp. 50–63, 2017.

[87] F. Milletti, 'Cell-penetrating peptides: classes, origin, and current landscape.', *Drug Discov. Today*, vol. 17, no. 15–16, pp. 850–60, Aug. 2012.

[88] H. D. Herce *et al.*, 'Arginine-rich peptides destabilize the plasma membrane, consistent with a pore formation translocation mechanism of cell-penetrating peptides.', *Biophys. J.*, vol. 97, no. 7, pp. 1917–25, Oct. 2009.

[89] J. S. Wadia, R. V Stan, and S. F. Dowdy, 'Transducible TAT-HA fusogenic peptide enhances escape of TAT-fusion proteins after lipid raft macropinocytosis.', *Nat. Med.*, vol. 10, no. 3, pp. 310–5, Mar. 2004.

[90] D. Derossi, G. Chassaing, and A. Prochiantz, 'Trojan peptides: the penetratin system for intracellular delivery.', *Trends Cell Biol.*, vol. 8, no. 2, pp. 84–7, Feb. 1998.

[91] A. Elmquist, M. Hansen, and U. Langel, 'Structure-activity relationship study of the cell-penetrating peptide pVEC.', *Biochim. Biophys. Acta*, vol. 1758, no. 6, pp. 721–9, Jun. 2006.

[92] K. H. Vousden and X. Lu, 'Live or let die: the cell's response to p53.', *Nat. Rev. Cancer*, vol. 2, no. 8, pp. 594–604, Aug. 2002.

[93] E. L. Snyder, B. R. Meade, C. C. Saenz, and S. F. Dowdy, 'Treatment of terminal peritoneal carcinomatosis by a transducible p53-activating peptide.', *PLoS Biol.*, vol. 2, no. 2, p. E36, Feb. 2004.

[94] G. E. Roeder, J. L. Parish, P. L. Stern, and K. Gaston, 'Herpes simplex virus VP22-human papillomavirus E2 fusion proteins produced in mammalian or bacterial cells enter mammalian cells and induce apoptotic cell death.', *Biotechnol. Appl. Biochem.*, vol. 40, no. Pt 2, pp. 157–65, Oct. 2004.

[95] B. Suresh, S. Ramakrishna, and H. Kim, 'Cell-Penetrating Peptide-Mediated Delivery of Cas9 Protein and Guide RNA for Genome Editing.', *Methods Mol. Biol.*, vol. 1507, pp. 81–94, 2017.

[96] J. S. Orange and M. J. May, 'Cell penetrating peptide inhibitors of nuclear factor-kappa B.', *Cell. Mol. Life Sci.*, vol. 65, no. 22, pp. 3564–91, Nov. 2008.

[97] B. Fang, L. Jiang, M. Zhang, and F. Z. Ren, 'A novel cell-penetrating peptide TAT-A1 delivers siRNA into tumor cells selectively.', *Biochimie*, vol. 95, no. 2, pp. 251–7, Feb. 2013.

[98] D. Libetti *et al.*, 'NF-YA enters cells through cell penetrating peptides.', *Biochim. Biophys. acta. Mol. cell Res.*, vol. 1866, no. 3, pp. 430–440, 2019.

[99] F. A. Ran, P. D. Hsu, J. Wright, V. Agarwala, D. A. Scott, and F. Zhang, 'Genome engineering using the CRISPR-Cas9 system.', *Nat. Protoc.*, vol. 8, no. 11, pp. 2281–

2308, Nov. 2013.

[100] T. Sakuma, A. Nishikawa, S. Kume, K. Chayama, and T. Yamamoto, 'Multiplex genome engineering in human cells using all-in-one CRISPR/Cas9 vector system.', *Sci. Rep.*, vol. 4, p. 5400, Jun. 2014.

[101] J. M. Brickman and P. Serup, 'Properties of embryoid bodies.', *Wiley Interdiscip. Rev. Dev. Biol.*, vol. 6, no. 2, 2017.

[102] P. A. Doevendans, S. W. Kubalak, R. H. An, D. K. Becker, K. R. Chien, and R. S. Kass, 'Differentiation of cardiomyocytes in floating embryoid bodies is comparable to fetal cardiomyocytes.', *J. Mol. Cell. Cardiol.*, vol. 32, no. 5, pp. 839–51, May 2000.

[103] V. Basile *et al.*, 'NF-YA splice variants have different roles on muscle differentiation.', *Biochim. Biophys. Acta*, vol. 1859, no. 4, pp. 627–38, Apr. 2016.

[104] S. Sato, 'Identification of transcriptional targets for Six5: implication for the pathogenesis of myotonic dystrophy type 1', *Hum. Mol. Genet.*, vol. 11, no. 9, pp. 1045–1058, May 2002.

[105] A. N. Tiaden and P. J. Richards, 'The emerging roles of HTRA1 in musculoskeletal disease.', *Am. J. Pathol.*, vol. 182, no. 5, pp. 1482–8, May 2013.

[106] V. Salizzato *et al.*, 'Protein kinase CK2 subunits exert specific and coordinated functions in skeletal muscle differentiation and fusogenic activity.', *FASEB J.*, vol. 33, no. 10, pp. 10648–10667, Oct. 2019.

[107] D. P. Millay *et al.*, 'Structure-function analysis of myomaker domains required for myoblast fusion.', *Proc. Natl. Acad. Sci. U. S. A.*, vol. 113, no. 8, pp. 2116–21, Feb. 2016.

[108] M. J. Petrany and D. P. Millay, 'Cell Fusion: Merging Membranes and Making Muscle.', *Trends Cell Biol.*, vol. 29, no. 12, pp. 964–973, Dec. 2019.

[109] J. Kahle, M. Baake, D. Doenecke, and W. Albig, 'Subunits of the heterotrimeric transcription factor NF-Y are imported into the nucleus by distinct pathways involving importin beta and importin 13.', *Mol. Cell. Biol.*, vol. 25, no. 13, pp. 5339–54, Jul. 2005.

[110] C. El Chartouni, C. Benner, M. Eigner, M. Lichtinger, and M. Rehli, 'Transcriptional effects of colony-stimulating factor-1 in mouse macrophages.', *Immunobiology*, vol. 215, no. 6, pp. 466–74, Jun. 2010.

[111] J.-H. Koo, H. Yoon, W.-J. Kim, S. Lim, H.-J. Park, and J.-M. Choi, 'Cell membrane

penetrating function of the nuclear localization sequence in human cytokine IL-1α.', *Mol. Biol. Rep.*, vol. 41, no. 12, pp. 8117–26, Dec. 2014.

[112] A. D. Ragin, R. A. Morgan, and J. Chmielewski, 'Cellular import mediated by nuclear localization signal Peptide sequences.', *Chem. Biol.*, vol. 9, no. 8, pp. 943–8, Aug. 2002.

[113] E. Harreither *et al.*, 'Characterization of a novel cell penetrating peptide derived from human Oct4.', *Cell Regen. (London, England)*, vol. 3, no. 1, p. 2, 2014.

PART II

“NF-YA enters cells through cell penetrating peptides”

Contribution to the published article

The article “NF-YA enters cells through cell penetrating peptides” published in the BBA - Molecular Cell Research Journal and attached above is related to the work performed during my PhD thesis. My contribution to this study started from proteins production using NF-YA wt and mutated DNA sequences already present in the laboratory. I produced and renatured from Inclusion Bodies all proteins used in this work, except for the TRX-NF-YAl protein which was produced by dr. Andrea Bernardini. I used the recombinant proteins to transfect cells analyzing the half-life of wt and mutated proteins. I checked the capability to transduce nuclear membrane of wt and mutated proteins identifying the CPPs responsible for protein transduction. Then, I checked for protein functionality after transfection by performing EMSAs using nuclear extracts of HeLa cells transfected with NF-YAs, NF-YAl or the mutant control. Once analyzed the effective capability of transfected wt proteins to interact with the NF-YB/YC dimer and bind the CCAAT-box, I used C2C12 cells to check if the transfected proteins affect myotubes differentiation. Based on previous published data that showed the inhibition of muscle differentiation in C2C12 overexpressing the short isoform of NF-YA [103], I checked for differentiation efficiency in C2C12 cells transfected with NF-YAs, NF-YAl recombinant proteins and the mutant control. As expected, C2C12 cells transfected with the short protein of NF-YA completely lost the capability to differentiate, confirming the functionality of recombinant proteins after cells transfection and nucleus penetration.

The data produced during my PhD thesis and showed in this article confirmed NF-YA to behave as a Cell Penetrating Peptide.

BBA - Molecular Cell Research 1866 (2019) 430–440

Contents lists available at ScienceDirect

BBA - Molecular Cell Research

journal homepage: www.elsevier.com/locate/bbamcr

NF-YA enters cells through cell penetrating peptides

Debora Libetti, Andrea Bernardini, Maria Luisa Chiaramonte, Mario Minuzzo, Nerina Gnesutta, Graziella Messina, Diletta Dolfini*, Roberto Mantovani*

Dipartimento di Bioscienze, Università degli Studi di Milano, Via Celoria 26, 20133 Milano, Italy

ARTICLE INFO

Keywords:
CCAAT
NF-Y
Cell penetrating peptide
Transcription factor

ABSTRACT

Cell Penetrating Peptides -CPPs- are short aminoacidic stretches present in proteins that have the ability to translocate the plasma membrane and facilitate delivery of various molecules. They are usually rich in basic residues, and organized as alpha helices. NF-Y is a transcription factor heterotrimer formed by two Histone Fold Domain –HFD- subunits and the sequence-specific NF-YA. NF-YA possesses two α-helices rich in basic residues. We show that it efficiently enters cells at nanomolar concentrations in the absence of carrier peptides. Mutagenesis identified at least two separate CPPs in the A1 and A2, which overlap with previously identified nuclear localization signals (NLS). The half-life of the transduced protein is short in human cancer cells, longer in mouse C2C12 myoblasts. The internalized NF-YA is capable of trimerization with the HFD subunits and binding to the target CCAAT box. Functionality is further suggested by protein transfection in C2C12 cells, leading to inhibition of differentiation to myotubes. In conclusion, NF-YA contains CPPs, hinting at novel -and unexpected- properties of this subunit.

1. Introduction

Cell Penetrating Peptides -CPPs- are short aminoacidic stretches with the capacity to mediate the internalization of proteins, also functioning as carriers of additional molecular cargoes, including nucleic acids and drugs [1]. Currently, there is a long list of such peptides [2]. They were originally discovered within sequences of transcriptional activators, the HIV TAT [3,4] and the Drosophila *Antennapedia* [5–8]. These peptides were extensively characterized by mutagenesis and, when fused to unrelated proteins, they introduce them into cells. Such protein transduction system appears to be efficient for a number of proteins, with a wealth of possible biotechnological and therapeutic implications being currently explored [9–11]. Two common structural features of CPPs are the presence of amphipathic α-helices and the abundance of arginine residues.

The TAT CPP was used with the NF-YA subunit of the heterotrimeric transcription factor NF-Y, demonstrating that the fusion protein could be efficiently introduced into the nuclei of hematopoietic cells by protein transfection; exposure to this fusion protein affected transcription of NF-Y target genes [12]. Importantly, treatment of hematopoietic stem cells -HSCs- led to an expansion of the stem precursors pool *ex vivo*, with a subsequent increase in the efficiency of bone marrow engraftment in irradiated nude mice. The functionality of recombinant GST-TAT-NF-YA was also documented in mouse ES cells [13] and human embryonal carcinoma NT2 cells [14].

NF-Y is considered a “pioneer” Transcription Factor (TF), responsible to set up a favorable chromatin environment for other TFs to bind DNA and induce mRNA elongation by the RNA Pol II machinery [15–18]. The DNA sequence bound is the CCAAT box, a conserved and important element found in promoters and enhancers [15]. The trimer comprises two Histone Fold Domain -HFD- subunits, NF-YB and NF-YC, similar to histones H2B and H2A [19], and the NF-YA regulatory subunit, which confers sequence selectivity. Two structures of the evolutionarily conserved domain bound to the CCAAT box illustrate the modalities of trimerization and DNA recognition [20,21], the last exerted through minor groove contacts by the A2 helix and the RGXGGRF motif of NF-YA. The HFD-interacting part is located in α-helix A1. This 56 amino acids domain, rich in basic residues, provides two additional important functions: (i) lysine poly-ubiquitination and degradation, responsible for its relatively rapid turnover [22]. (ii) Nuclear localization, with two identified nuclear localization signals -NLS- in A1 and A2, respectively, driving import through Importin β [23,24]. Finally, NF-YA is present in two major isoforms due to alternative splicing, resulting in a “short” -NF-YAs- and a “long” -NF-YAl- isoform, the latter containing 28 additional amino acids within the N-term activation domain, coded by Exon 3 [25]. The two isoforms were shown to have distinct roles in at least two cellular systems, mESCs and myoblasts [13,26].

* Corresponding authors.
E-mail addresses: diletta.dolfini@unimi.it (D. Dolfini), mantor@unimi.it (R. Mantovani).

https://doi.org/10.1016/j.bbamcr.2018.10.004
Received 1 June 2018; Received in revised form 10 September 2018; Accepted 3 October 2018
Available online 06 October 2018

Inspection of the 3D crystallographic structure of NF-YA suggests the presence of features typical of CPPs, notably two α-helices -A1 and A2- rich in arginine residues. For this reason, we wondered whether NF-YA has the intrinsic potential to penetrate cellular membranes.

2. Materials and methods

2.1. Protein production and purification

The full length long and short NF-YA, as well as NF-YAm29 and m23 cloned in pET3b plasmids were described [27,28]. The mutants YAm31, 32, 33 and 34 of NF-YA long were constructed by PCR, using appropriately mismatched oligonucleotides; the fragments were inserted using *Kpn*I-*Bam*HI sites in the pET3b/NF-YAl backbone. pET3b/NF-YAm23 and pET3b/NF-YAm29 were used as templates to create mutants m23–33, m23–34, m31–29 and m32–29 with the same strategy. The thioredoxin (TRX) tagged NF-YA construct was obtained by subcloning the full length NF-YA coding sequence (long isoform) from the pSG5 construct [32] into the pET32b expression vector (Novagen), using *Eco*RI-*Sac*I sites. All resulting plasmids were checked by DNA sequencing.

NF-YA recombinant proteins were produced using *Escherichia coli* BL21(DE3) pLysS: logarithmically growing cells were induced for 3 h with 0.1 mM IPTG at 37 °C. The cell pellets were lysed in 300 mM NaCl, 20 mM Tris-HCl pH 7.5, 0.1 mM DTT, 0.05% NP40 and protease inhibitors (Sigma), briefly sonicated and centrifuged [27,28]. Inclusion bodies were solubilized in 8 M urea, 150 mM NaCl, 20 mM Tris-HCl pH 7.5, 0.1 mM DTT, 1 mM PMSF, renatured by step-wise dialysis in urea-free buffer (150 mM NaCl, 20 mM Tris-HCl pH 7.5, 0.1 mM DTT, 1 mM PMSF, 10% glycerol), centrifuged and stored at −80 °C. At least three batches of each recombinant NF-YA (wt and mutants) were prepared. For the expression of TRX-NF-YAl recombinant fusion protein, logarithmically growing BL21(DE3) cells were induced for 3 h with 0.5 mM IPTG. Cell pellet was resuspended in buffer A1 (150 mM NaCl, 20 mM Tris-HCl pH 8.0, 2 mM $MgCl_2$, 10 mM imidazole) and lysed by sonication. TRX-NF-YAl was purified from the soluble fraction through nickel-affinity chromatography (His-Select® Nickel Affinity Gel, Sigma-Aldrich) exploiting the 6xHis-tag present in the fusion construct. The protein was eluted in buffer A1 containing 100 mM imidazole and dialyzed against buffer B1 (150 mM NaCl, 20 mM Tris-HCl pH 8.0, 1 mM DTT, 10% glycerol). LPS was removed from protein preparations using Triton X114, as described in [29], except that one step of precipitation was performed. Proteins were analyzed by SDS-polyacrylamide gel electrophoresis and visualized by staining with Coomassie Brilliant Blue R-250 solution (0.25% Coomassie Brilliant Blue R-250, 50% ethanol, 10% acetic acid) for 40 min.

NF-YAl, NF-YAm 23-33 and m23-34 recombinant proteins were further purified by size-exclusion chromatography (SEC) in urea-free buffer (150 mM NaCl, 20 mM Tris-HCl pH 7.5, 0.1 mM DTT). 10% glycerol was added after the chromatography. SEC was performed on a Superdex 200 Increase 10/300 GL column (GE Healthcare) connected to an Äkta Pure chromatography system (GE Healthcare).

Recombinant NF-YB and NF-YC (37 kD isoform) were produced separately as described [27,30], except that the lysis solution was 300 mM KCl, 20 mM Tris-HCl pH 7.8, 0.1 mM EDTA pH 8, 0.05% NP40, 5 mM β-mercaptoethanol, protease inhibitors. Inclusion bodies were solubilized in 6 M GnCl, 100 mM KCl, 20 mM Tris-HCl pH 7.8, 5 mM β-mercaptoethanol. NF-YB and NF-YC were mixed in equimolar amounts and renaturated by dialysis in 500 volumes of 10 mM Tris-HCl pH 8, 300 mM KCl, 10% glycerol, 1 mM DTT and 3.5 M urea for 2 h at room temperature and then in same buffer without urea and stored at −80 °C. After renaturation, the NF-YB/NF-YC dimer was purified using His-select® Nickel Affinity Gel, exploiting the His-Tag of NF-YC. The dimer was eluted in 250 mM imidazole pH 8, 10 mM Tris-HCl pH 8, 300 mM KCl, 10% glycerol, 1 mM DTT, dialyzed over-night in the same buffer without imidazole, quick-frozen in liquid nitrogen and stored at −80 °C. For production and purification of the NF-YB/NF-YCmd used in the EMSAs of Fig. 4, we followed the protocol used in [31].

2.2. Cell culture and NF-YA protein transfection

Human colorectal carcinoma cells (HCT-116), human cervical adenocarcinoma cells (HeLa), human glioblastoma cells (T98G), human osteosarcoma cells (U2OS), human immortalized epithelial keratinocytes (HaCaT) and mouse myoblast cells (C2C12, ATCC) were cultured in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% fetal bovine serum (FBS), 4 mM L-Glutamine, 100 units/ml penicillin and 100 μg/ml streptomycin, in a humidified 5% CO_2 atmosphere at 37 °C.

NF-YA protein transfections were performed in 6 wells plates plating 3×10^5 cells per well. Exponentially growing cells were washed with PBS, treated with the indicated amounts of recombinant NF-YA in serum-free medium for 30′ at 37 °C in 5% CO_2. After 30′, cells were washed with PBS and incubated with recombinant NF-YA for further 30′. Finally, cells were washed with PBS, cultured in complete DMEM, and collected at the indicated time points.

C2C12 cells differentiation was induced plating 1×10^5 cells in 35 mm petri dishes in DMEM with 2% horse serum, 4 mM L-Glutamine, 100 units/ml penicillin and 100 μg/ml streptomycin. NF-YA transfections were performed by adding 20 nM of protein into the medium every 24 h during the differentiation process. Each transfection was performed in triplicate.

2.3. Cell extracts and Western blot analysis

For cytosolic and nuclear extracts, cells were washed with PBS, collected by scraping and lysed in ice-cold Buffer A (10 mM KCl, 10 mM Hepes, pH 7.8, 0.1 mM EDTA, 0.1 mM EGTA, protease inhibitors) by incubation on ice for 30 min. 0.65% NP-40 was added and cells were briefly vortexed and centrifuged at 2500 rpm for 5 min. Supernatant, corresponding to cytosolic extracts, was transferred in a new Eppendorf tube. Pelleted nuclei were resuspended in ice-cold Buffer C (0.75 M NaCl, 20 mM Hepes pH 7.8, 1 mM EDTA, 1 mM EGTA, 1 mM dithiothreitol, protease inhibitors). For total protein extracts, cells were washed with PBS, collected by scraping, lysed in ice-cold RIPA buffer (10 mM Tris-HCl, pH 8.0, 1 mM EDTA, 0.5 mM EGTA, 0.1% SDS, 0.1% Deoxycholic acid, 140 mM NaCl, 1% Triton X-100, 1 mM PMSF, protease inhibitors) and incubated for 30 min on ice, shaking vigorously every 5 min. Samples were centrifuged at 13,000 rpm for 10 min at 4 °C and the supernatant containing total protein extract was transferred to a new Eppendorf tube. Protein concentrations were determined by Bradford Assay (Bio-Rad Protein Assay). 10 μg of cytosolic, nuclear and total extracts were loaded on a 7–15% SDS-polyacrylamide gel and analyzed by Western blot using primary antibodies and a peroxidase-conjugate secondary antibody (Sigma). Quantifications were performed by using the Image Lab™ software (Bio-Rad) and the standard deviations of at least three independent replicates calculated. The primary antibodies used were: anti-NF-YA (Santa Cruz Biotechnology, SC-10779 and Bethyl Laboratories, A302-104A), anti-Vinculin (Santa Cruz Biotechnology, V284), anti-Lamin B (Santa Cruz Biotechnology, SC-6217), anti-LPS (Hycult Biotech, WN1 222–5), anti-all sarcomeric MyHCs (DHSB MF20), anti-Myog (DHSB IF5D), anti-β-Tubulin (Sigma T5293), anti-H2B (Abcam, ab1790).

2.4. Electrophoretic mobility shift assay (EMSA)

The probe used was a 31 bp double-stranded oligonucleotide containing the Hsp70 promoter CCAAT-Box (5′-CTTCTGAGCCAATCACCG AGCTCGATGAGGC-3′) 5′-labelled with Cy5. Dose-response experiments were performed with different concentrations of NF-YA, 20 nM of Hsp70 probe and 60 nM of purified NF-YB/NF-YC dimer. Proteins were diluted in the dilution buffer (400 mM NaCl, 10 mM Tris-HCl, pH 8.0,

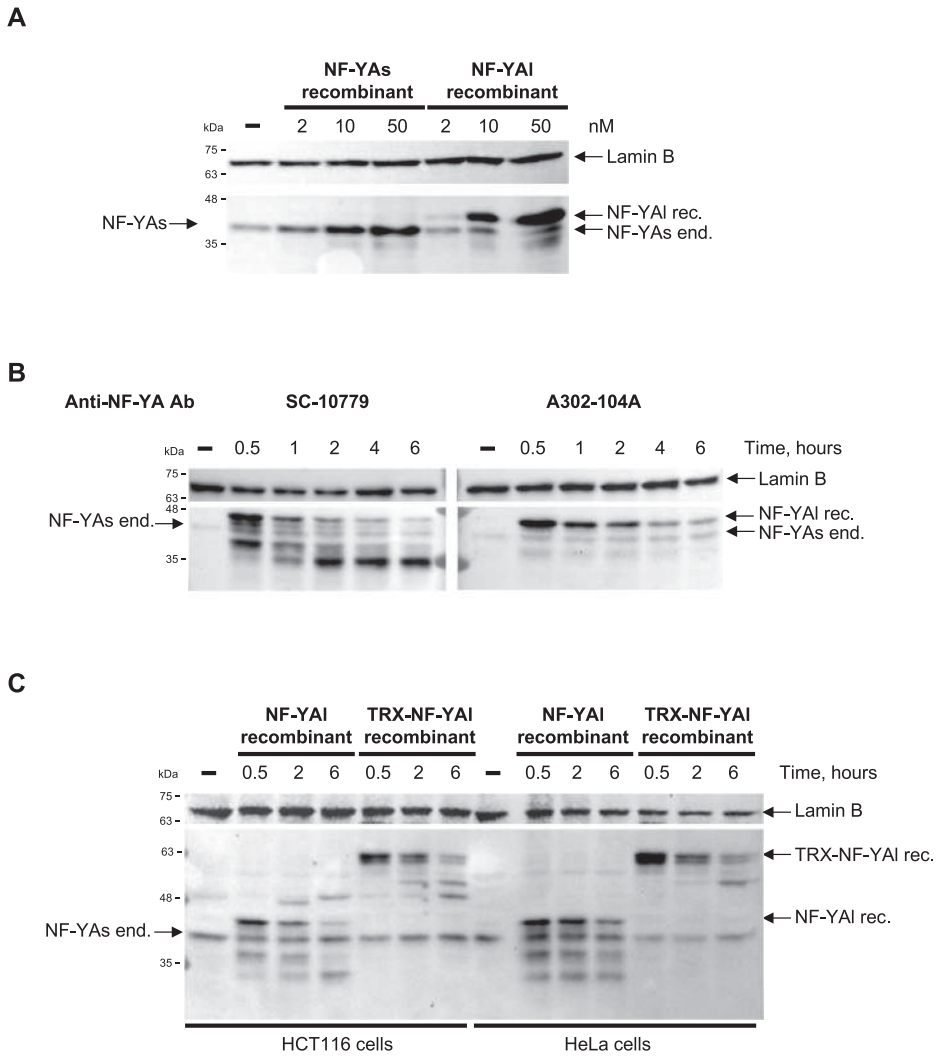


Fig. 1. Transfections of NF-YAl and TRX-NF-YAl recombinant proteins.
(A) Western blot analysis of NF-YA in nuclear extracts of HCT116 cells treated with increasing concentrations (2, 10, 50 nM) of NF-YAs and NF-YAl recombinant proteins. Cells were collected 2 h after treatments. The endogenous (end.) and transfected recombinant (rec.) proteins are indicated. Lamin B was used as loading control. (B) Kinetics analysis of NF-YAl protein transfections in HCT116 cells by using anti-NF-YA Santa Cruz Biotechnology (SC-10779) and anti-NF-YA Bethyl Laboratories (A302-104A). Cells were treated with 10 nM of recombinant NF-YAl and nuclear extracts prepared at the indicated times post-transfection. (C) Kinetics analyses of recombinant NF-YAl and TRX-NF-YAl −10 nM- in HCT116 and HeLa cells for the indicated times post-transfection.

10% glycerol, 0.1 mg/ml BSA). For incubation, binding buffer 5× (250 mM KCl, 25 mM $MgCl_2$, 60 mM Tris-HCl pH 8, 2.5 mM EDTA, 60% glycerol, 12.5 mM DTT, 0.5 mg/ml BSA) was added to samples. For EMSAs performed with nuclear extracts, 10 nM of Hsp70 probe, 2 μg of nuclear extracts were used with different concentration of NF-YB/NF-YC. Incubation was performed diluting NF-Y-binding buffer 5× (25 mM $MgCl_2$, 100 mM Tris-HCl pH 7.5, 2.5 mM EDTA, 25% Glycerol, 12.5 mM DTT, 0.5 mg/ml BSA) and adding 100 ng of poly-dIdC. Reactions were incubated at 30 °C for 30 min, protected from light, and loaded on the gel (TBE 0.25X, 4.5% Acrylamide/Bis-acrylamide (19:1 ratio), 2.5% glycerol, 0.1% APS, 0.01% TEMED), in a dark room at 4 °C; electrophoresis was carried out at constant 100 V for 1 h. Detection was performed with the ChemiDoc MP imaging system (Biorad). Each EMSA experiment was performed in triplicate.

2.5. BrdU incorporation

For BrdU incorporation, 1×10^5 C2C12 cells were plated in 35 mm dishes and transfected with 20 nM of NF-YA recombinant proteins. 24 h after transfection, cells were exposed for 1 h to 50 μM BrdU, then fixed with ice-cold 95% EtOH-5% acetic acid for 20′ at room temperature, incubated with PBS-1.5 M HCl for 10′ at room temperature, washed twice with PBS, and permeabilized with 0.25% Triton X-100 in PBS. Primary antibody (1:100, anti-Bromo-deoxyuridine, Amersham) was incubated for 1 h at 4 °C. After three washes, secondary FITC anti-mouse antibody (1:500) plus DAPI (2 μg/ml) were incubated for 40′ at room temperature in the dark. The cells were washed twice with PBS before mounting. The acquisition was performed by using the confocal microscope Leica DMI6000B.

2.6. Immunofluorescence

For immunofluorescence analyses, cells were washed three times with PBS and fixed with ice-cold Acetone-Methanol (1:1) for 10 min at room temperature. After three washes, cells were permeabilized with 0.25% Triton X-100 in PBS for 5 min. Primary antibody to sarcomeric MyHCs was incubated for 1 h at room temperature. Cells were washed three times, permeabilized with 0.25% Triton X-100 in PBS for 5 min, and incubated with secondary FITC anti-mouse antibody (1:500) plus DAPI (2 μg/ml) for 40 min at room temperature light protected. The acquisition was performed by using the confocal microscope Leica SP2.

3. Results

3.1. Recombinant NF-YA enters cells efficiently

We noticed that the evolutionarily conserved part of NF-YA contains two features typical of CPPs, the presence of α-helices -A1 and A2- and arginines in the 56 amino acids conserved domain [20]. We decided to test whether recombinant NF-YA could enter cells in the absence of any exogenous CPP peptide. We produced in *E. coli* the two recombinant NF-YA isoforms, "long" and "short" -NF-YAl and NF-YAs- [25]: as expected, the proteins were found in inclusion bodies, as checked by SDS-PAGE (Fig. S1A). The recombinant proteins were solubilized in 8 M urea and renatured by slowly removing the denaturing agent. Note the presence of additional bands, mostly of lower molecular weight. To get rid of contaminating LPS-like material present in the bacterial preparations, we used the Triton X-114 precipitation protocol [29]: this procedure largely eliminates LPS from preparations, as checked by Western blot analysis, with modest loss of recombinant proteins (Fig. S1B). The purified proteins were quantified and controlled by Western blot using two antibodies: the rabbit SC-10779 anti-YA polyclonal antibody raised against AA 139–347 of NF-YA and the A302-104A-polyclonal whose epitope is a peptide at the very C-terminus of the protein (AA 320–333). Bands of smaller molecular weight were visualized with the former, but not with the latter (Fig. S1C), suggesting that the bands of lower molecular weight in Coomassie are NF-YA resulting from degradation involving the C-terminal of the protein. We used the recombinant proteins in transfections of human colon carcinoma HCT116 cells. The rationale for using this cell line was that it almost exclusively expresses NF-YAs [32]: any addition of the "long" NF-YA would thus be readily discernible by Western blot. We set up a simplified protocol providing the cells twice with NF-YA in serum-free medium at 30′ intervals. Because NF-YA is localized exclusively in the nucleus [23,24], we prepared and assayed nuclear extracts in all further experiments. Fig. 1A shows Western blot analysis of nuclear extracts prepared 2 h after transfection following a dose-response addition of NF-YAl: NF-YAl enters cells at nanomolar concentrations, as it was previously shown for GST-TAT-NF-YA [12]. The levels of endogenous NF-YAs of HCT116 are largely unchanged, serving as a further internal control, in addition to the loading control Lamin B. Note that, at 2 h time points, transfections yielded amounts of nuclear NF-YAl that are many-folds higher than those of the endogenous NF-YAs. We also transfected the "short" NF-YA, which showed a similar behavior: in this case, the signal of the transfected isoform in Western blot is superimposed with that of the endogenous protein, hence we notice an increase in the NF-YAs band (Fig. 1A).

We next performed a time course of the transfected proteins: Fig. 1B shows that NF-YAl is observed at near-maximum levels in cells already after 30′ post-transfection, and still high at 2 h; by 6 h, the protein amounts are considerably decreased. Interestingly, the short half-life of the transduced protein is similar to the one previously calculated in overexpression experiments based on DNA-mediated gene transfections [22]. The blots in Fig. 1B (Left Panel), probed with the rabbit SC-10779 anti-NF-YA polyclonal antibody, reveal specific bands of lower molecular weight increasing as the time course progressed, which were often recorded in our experiments (See also Fig. 4, S2B, S3, S5). The use of the A302-104A antibody did not visualize these bands, indeed indicating that they are degradation products of NF-YA (Fig. 1B Right Panel).

We performed two additional types of experiments. The first was purification of recombinant NF-YAl (and selected mutants, See below) by Gel Filtration (Fig. S2A): we used these proteins in transfections: wt NF-YAl, but not mutants (see below) entered HCT116 cells, as shown in Western blot of nuclear extracts (Fig. S2B). The second important control was the production of a His-tagged NF-YAl recombinant protein fused to Thioredoxin -TRX-NF-YAl- purified from the bacterial soluble fraction through NTA affinity chromatography (Fig. S1A). The recombinant protein was > 90% pure (Fig. S1). This allows to investigate two additional issues of protein transfections: (i) whether it would work with soluble NF-YA, and (ii) whether the addition of an external moiety would prevent entry into nuclei. Transfections were performed in parallel in HCT116 and in HeLa cells and the results of these experiments are shown in Fig. 1C: in both cell lines, efficient entry of TRX-NF-YAl was recorded, similar to the one shown for NF-YAl produced and renatured from inclusion bodies. The kinetics of elimination was also similar, as well as the presence of degradation products. We conclude that irrespective of the method of production and purification, and of the presence of additional tags, NF-YA maintains an intrinsic capacity to enter mammalian cells. We also show rapid proteolytic degradation of TRX-NF-YAl, mostly occurring at the C-terminal end, as previously described for proteins generated by DNA-mediated transfections [22].

3.2. Stability of transfected NF-YA in different cell lines

Transfection experiments of untagged NF-YA were then extended to other cell lines: HeLa, U2OS, T98G and HaCaT. We employed a 10 nM dose of recombinant NF-YAl over a 24 h time course (Fig. S3): in all cell lines, the protein is efficiently internalized, with levels being maximal, or near maximal, at 30 min/1 h post-transfection. Thereafter, the protein amounts were reduced already at 2 h, further decreased at 6 h and no exogenous protein detected at 24 h. Quantification of these results is shown in Fig. 2. In summary, we conclude that the entry of recombinant NF-YA is a generalized event observable in cell lines of different origin, with somewhat comparable kinetics of proteolysis.

3.3. Identification of NF-YA cell penetrating peptide(s)

The approach above was used to identify the amino acids responsible for internalization of NF-YA. NF-YA is a very basic protein (pI of 8,89), and positively charged amino acids are primarily located in the evolutionarily conserved domain (AA 262–321): 7 lysines, 4 histidines and 11 arginines. Endogenous NF-YA is exclusively localized in the nucleus [23,24]. Two separate signals are required for efficient nuclear localization, in human cells *in vivo,* and *Xenopus* cells *in vitro*: one is the KRR (AA 276–278) stretch in the A1 helix, the other involves RKR (AA 310–312) residues in helix A2 [23]; note that NF-YA mutations in either A1 or A2 NLS retained substantial nuclear localization, whereas mutations of both signals essentially abolished it. Along a similar line of reasoning, we thought that similar motifs might be also required for CPP activity. We mutagenized four stretches of three basic residues into alanines: YAm31 (K276-R277-R278), YAm32 (R294-R295-K296), YAm33 (R302-H303-R304) and YAm34 (R310-K311-R312) (Fig. 3A). Furthermore, we previously described a triple alanine mutant in A1 -YAm23 (L275-K276-R277)- that is partially overlapping with YAm31 [28]: we used it as a backbone to construct triple alanine mutations in the RHR (YAm33) and RKR (YAm34) stretches of A2, obtaining the respective double mutants. A similar strategy was used with another mutant previously described, YAm29 (R312-G313-E314), preparing double mutants in the A1 and linker, respectively (Fig. 3A).

All these mutants were constructed in the NF-YAl backbone, and proteins were prepared from bacterial inclusion bodies, as outlined

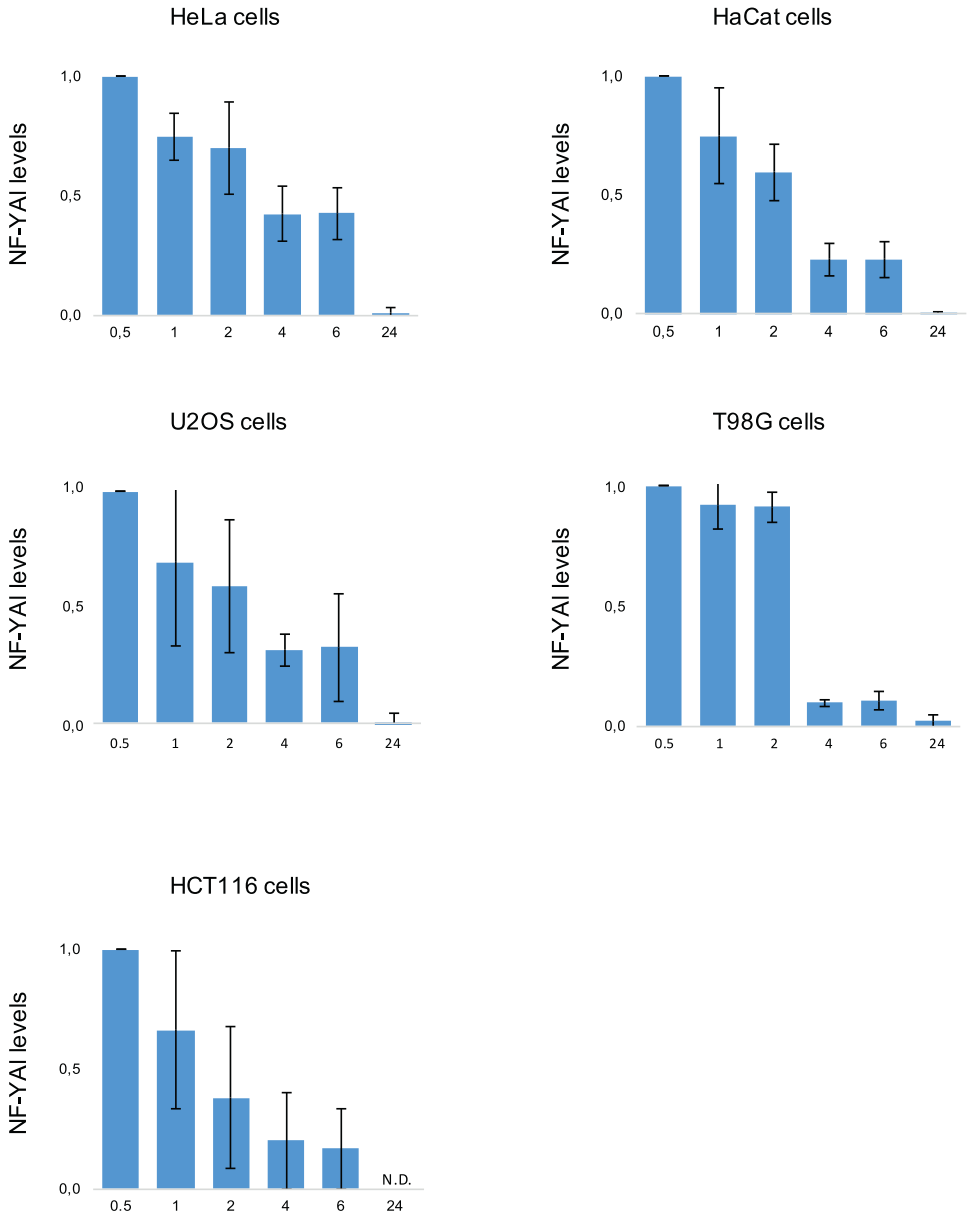


Fig. 2. Entry of NF-YAl in different cell lines.
Quantification of NF-YAl in nuclear extracts of different cell lines treated with 10 nM of NF-YAl recombinant protein. Cells were collected at different time points after transfection. In HCT116 cells the 24 h time point was not done.

above: Coomassie staining of SDS-PAGE of representative preparations are shown in Fig. S1A. To verify their functionality, we initially characterized these mutants for DNA-binding in EMSAs on the high affinity CCAAT box of the HSP70 promoter. Dose-response of wt NF-YAl on a fixed amount of recombinant NF-YB/NF-YC dimer generated a specific band, as expected. All mutants showed decreased binding, particularly YAm33 and YAm34, crippled in A2 and the RGXGGRF motif, and the double mutants YAm31–29, YAm32–29, YAm23–33 and YAm23–34 (Fig. S4A, S4B, S4C). These results are largely expected based on previous mutagenesis on the yeast and mammalian subunit [28,33–35] and with the 3D structures of the *fungi* and mammalian NF-Y/CCAAT quaternary complex [20,21].

Thereafter, we assayed these mutants in protein transfection assays: kinetic analysis at 10 nM concentrations was performed in HCT116 cells and NF-YA nuclear accumulation monitored by Western blots. Fig. 3B shows that YAm23, YAm31 and YAm34 entered cells at decreased levels compared to wt NF-YAl. YAm32 and YAm33 were only to some extent reduced in nuclear extracts with respect to wt NF-YA, but their overall levels did drop earlier, already at 2 h. YAm29 was internalized at similar rates and not consistently decreased. Note that in none of these mutants complete elimination of uptake was observed, suggesting the presence of more than one CPP. On the other hand, in the double

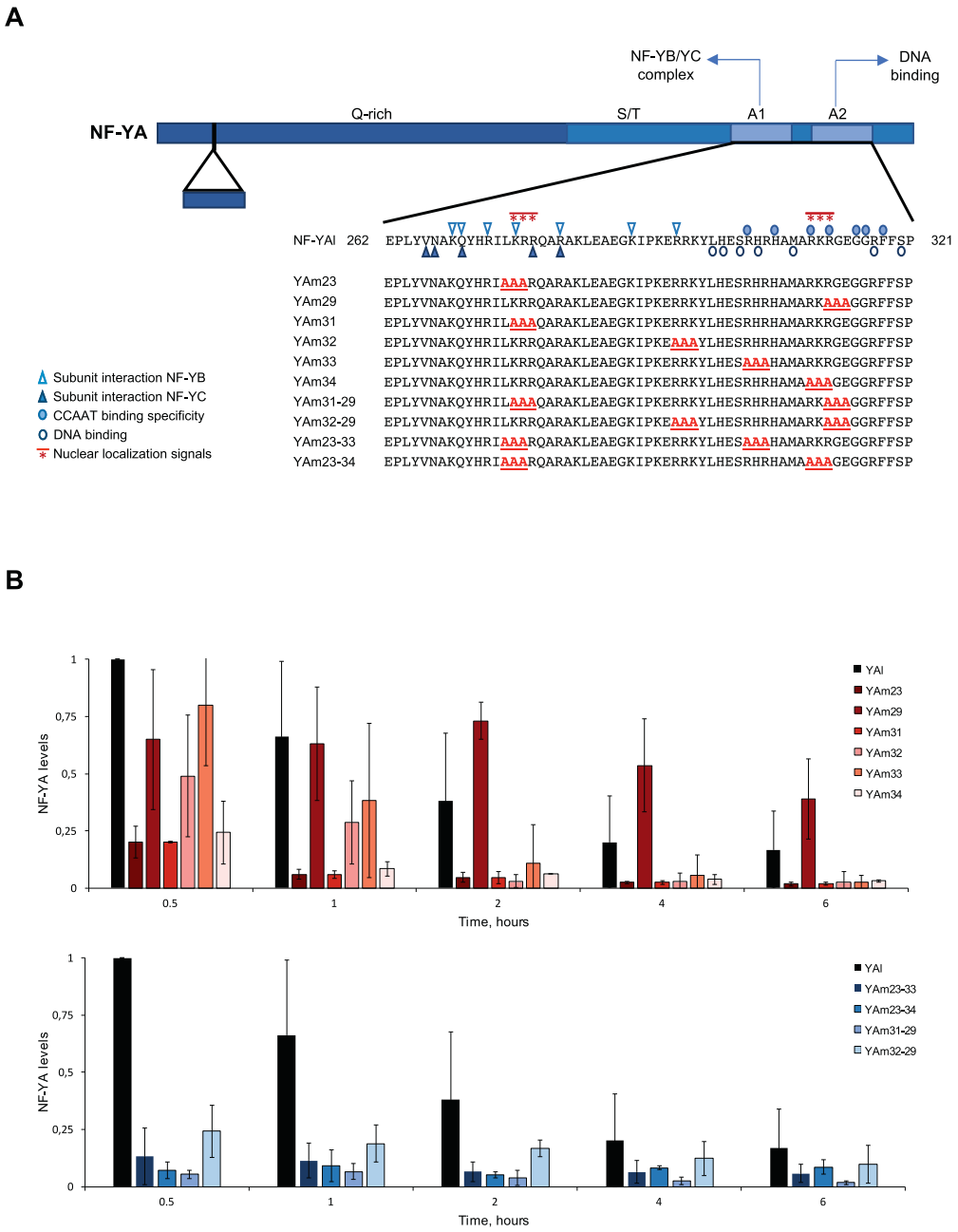


Fig. 3. Identification of NF-YAl cell penetrating peptide(s).
(A) The amino acid sequence of the evolutionarily conserved domain of NF-YA is shown. The sequence of the mutants employed in this study are underlined and highlighted in red. NF-YA amino acids involved in NF-YB/NF-YC interactions, CCAAT-specific binding, non sequence-specific DNA-binding and the nuclear localization signals are indicated. (B) Quantification of NF-YAl and mutants entry and stability in HCT116 cells, based on Western blots of nuclear extracts. Cells were treated with 10 nM of recombinant proteins and collected at different time points after protein transfections.

mutants, entry into cells was essentially abolished, with the exception of YAm32–29. Representative Western blots are shown in Fig. S5. Finally, we also checked the behaviour of the YAm23–33 and YAm23–34 mutants after purification by Gel Filtration (Fig. S2A): no significant entry into HCT116 cells was scored (Fig. S2B). Taken together, these results indicate that there are at least two separate CPPs in NF-YA, one in the A1, centered on the K276-R277-R278 sequence, and a second in the DNA-binding subdomain, encompassing the basic R310-K311-R312 stretch.

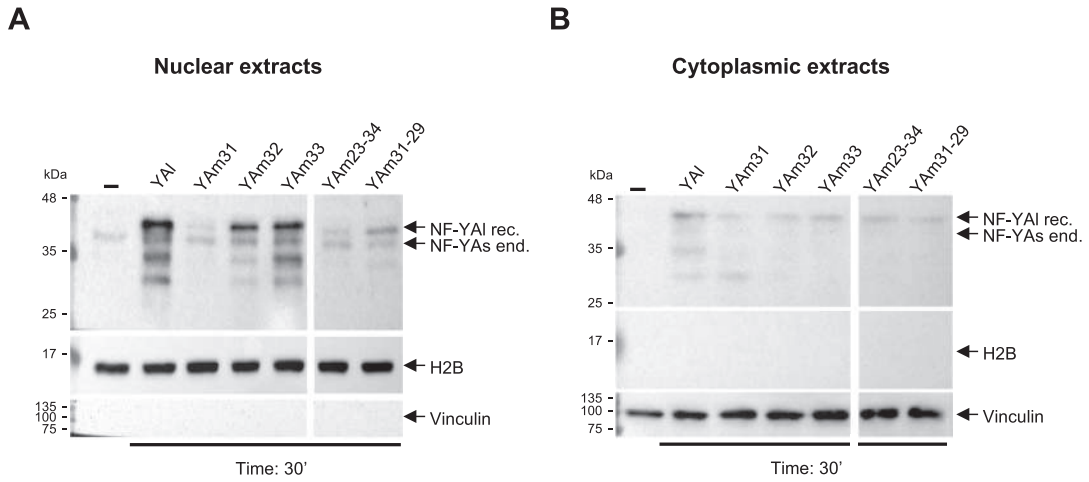


Fig. 4. NF-YA mutants are not retained in the cytoplasm.
HCT116 cells treated with 10 nM of recombinant NF-YAl, and the indicated mutants, were collected 30 min after transfection. Western blots of nuclear extracts (A) and cytoplasmic extracts (B) with anti-NF-YA antibody (Upper Panels), anti-H2B (Middle Panels, nuclear marker) and anti-Vinculin (Lower Panels, cytoplasmic marker). The endogenous (end.) and transfected recombinant (rec.) proteins are indicated.

3.4. NF-YA CCP mutants are not retained in the cytoplasm

In all the above experiments, we prepared and assayed nuclear extracts. The overlap between the NLS and CPP signals shown above raises the question of whether the absence of NF-YA mutants in nuclear extracts, specifically those in which the NLS is affected, might be due to retention in the cytoplasm, rather than a defect in CPP activity.

To verify this, we transfected wt NF-YAl and mutants YAm31, YAm32, YAm33, YAm23–34 and YAm31–29; we prepared, shortly after transfections, nuclear and cytoplasmic extracts. Fig. 4A and B shows Western blot analysis of such transfections: small amounts of NF-YAl, and even smaller of the different mutants, are present in the cytoplasm, but for the most part, the different NF-YA are found in the nucleus, as expected. No significant retention of any mutants due to mutation of the NLS was observed. We used cytoplasmic –Vinculin- and nuclear –histone H2B- markers to assess the effective separation of the two compartments in our extracts preparations. We conclude that the mutagenesis shown above indeed affects "*bona fide*" CPP activity.

3.5. DNA-binding of transfected NF-YA

To verify whether the internalized NF-YA was functional in terms of HFD-interactions and DNA-binding, we prepared nuclear extracts from HeLa cells transfected with NF-YAl and NF-YAs one-hour post-transfection, and performed EMSAs. To discriminate the complexes generated by the endogenous NF-Y trimer, we added a recombinant heterodimer composed of the HFD minimal domains (HFDmd) used for our crystallographic studies, termed NF-YB/NF-YCmd [31]: assembly into a DNA-binding trimer with NF-YA will generate a band of faster mobility in standard EMSAs. The results of adding two doses of HFDmd are shown in Fig. 5: in extracts of untransfected cells, the endogenous NF-Y band is shifted to a faster migrating band, as a result of association of endogenous NF-YA with recombinant HFDmd. In transfections with NF-YAs, an increase in intensity of the faster band is observed, as expected, since HCT116 cells express the "short" NF-YA and in extracts of NF-YAl-transfected cells two bands are indeed obtained, the lower corresponding to the endogenous NF-YAs interacting with the excess recombinant HFDmd added in the reaction, the upper corresponding to the recombinant NF-YAl associated with recombinant HFDmd. This latter is not seen with extracts of cells transfected with YAm23–33, used as a negative control. This indicates that the transfected NF-YA proteins are able to form complexes with the excess of HFDmd dimer and bind the HSP70 CCAAT box.

3.6. Transfected NF-YA affects C2C12 differentiation

The relatively short half-life of recombinant NF-YA in different cell lines makes it difficult to use it as a reagent to study function in growing cells. In the course of our tests with various cell types, we realized that a substantial amount of recombinant NF-YAl/NF-YAs were still visible at 24 h after transfection of exponentially growing C2C12 (Fig. S6A). Mouse C2C12 myoblasts are cells in which the levels of NF-YA isoforms are regulated [36–38]. Specifically, expression of NF-YAl, the major isoform present in proliferating C2C12, is switched off upon terminal differentiation and absent in myotubes [36]. This allowed us to test the phenotypic effect of NF-YA transfections on growth and differentiation: C2C12 cells were induced to differentiate by serum withdrawal and transfected at 24 h intervals with the recombinant proteins. Efficient transduction of NF-YA was checked by Western blot, collecting cells 2 h after treatment (Fig. 6A). After 72 h, cells were fixed and monitored by Immunofluorescence, and whole cell extracts were prepared for further biochemical analyses. Formation of well positioned multinucleated myotubes was observed, as expected, in mock transduced cells, as well as in cells transfected with the YAm23–33 mutant. Instead, NF-YAs, and to a slightly lesser extent NF-YAl, impaired formation of well organized myotubes (Fig. 6B). Evaluation of MyHC, a myotubes marker, by Western blot indeed indicate a reduction, whereas the levels of Myogenin, which is expressed in proliferating cells and throughout the process, is similar in the transfected and untransfected cells (Fig. 6C). We also evaluated the growth rates of C2C12, which did not change significantly upon NF-YAs or NF-YAl addition (Fig. S6B). Taken together, these data indicate that the transfected NF-YA is functional in terms of altering the process of C2C12 myocytes differentiation.

4. Discussion

CPPs were identified in several proteins, drawing considerable biotechnological and therapeutic interest. We show here that NF-YA adds to the list of TFs entering cells without exogenous tags *via* the presence of internal CPPs located in the evolutionarily conserved region.

Our work was motivated by (i) previous experiments with the GST-

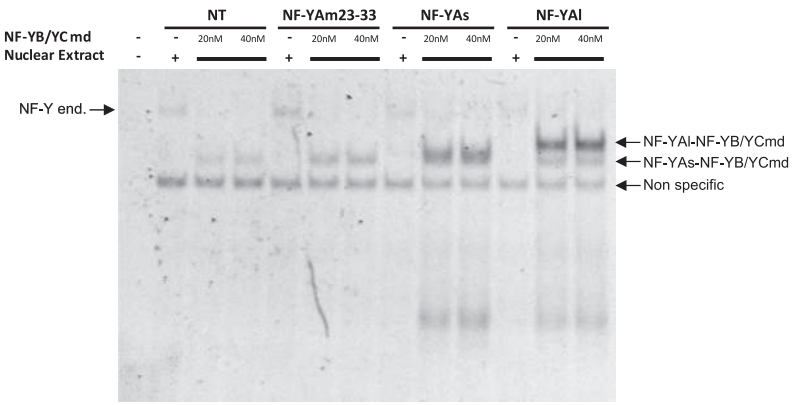


Fig. 5. CCAAT-binding of NF-YA transfected proteins. HeLa cells treated with 50 nM of recombinant NF-YA were collected 30 min after transfections. EMSAs were performed with 2 μg of nuclear extracts and an increasing amount of purified recombinant NF-YB/NF-YCmd dimer as indicated. CCAAT/NF-Y complexes are marked. The endogenous (end.) NF-Y is indicated.

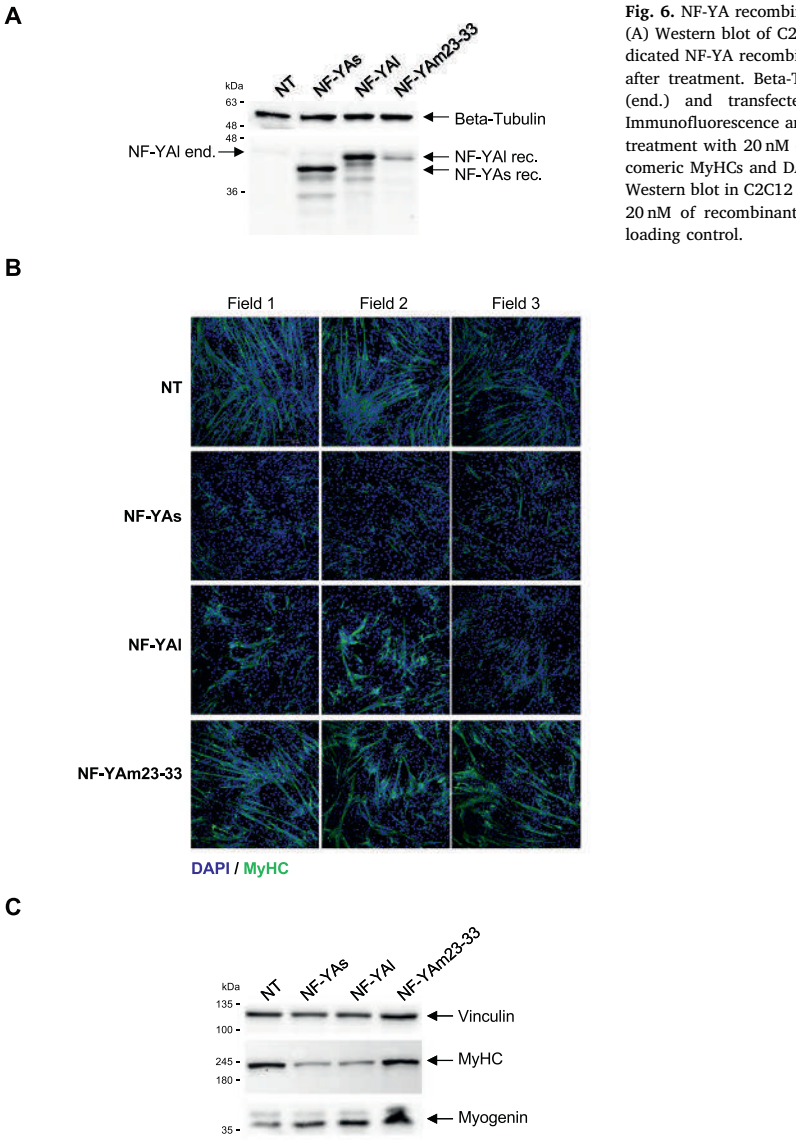


Fig. 6. NF-YA recombinant proteins affect C2C12 differentiation.
(A) Western blot of C2C12 total protein extracts treated with 20 nM of the indicated NF-YA recombinant proteins (wt and mutant). Cells were collected 2 h after treatment. Beta-Tubulin was used as loading control. The endogenous (end.) and transfected recombinant (rec.) proteins are indicated. (B) Immunofluorescence analyses of C2C12 cells after 72 h of differentiation, after treatment with 20 nM of recombinant proteins every 24 h; antibody to all sarcomeric MyHCs and DAPI were used. (C) Analysis of MyHC and Myogenin by Western blot in C2C12 cells after 72 h of differentiation; cells were treated with 20 nM of recombinant proteins every 24 h as above. Vinculin was used as loading control.

TAT-NF-YA fusion, showing efficient entry of the protein, and proficiency in transcriptional activation, including in primary HSCs: this established the notion that a transfected recombinant NF-YA would be functional *in vivo*. (ii) Inspection of the NF-Y/CCAAT 3D structure suggesting that CPPs could be present in A1 or in A2. The experiments confirm that this is the case. The protein concentrations used here are similar to previous studies, in the nanomolar range: thus, considering that some proteins are internalized in the low μM range, the efficiency of NF-YA CPPs is remarkable, yielding nuclear levels of NF-YA that are many-folds those of the endogenous protein.

Protein entry is matched by a short half-life of the protein, which could be viewed as indication of a physiological behaviour. In fact, NF-YA turnover is rapid, and subject to control by degradation. In HCT116 cells, as well as other cell lines, the half-life of the protein-transfected NF-YA appears to be very similar to the one measured in conventional DNA-mediated transfections [22], suggesting that the recombinant proteins behave as those produced by transient over-expression. Only in growing C2C12 cell line a prolonged presence of the exogenous NF-YA was observed. Functionally, overexpression of the either isoforms inhibited C2C12 differentiation, as expected based on previous transient expression experiments [37]. In fact, the levels of NF-YAl, the isoform predominantly expressed in C2C12, drop dramatically upon differentiation and myotubes formation [36]. This also emerged in C2C12 clones stably overexpressing NF-YAs, which showed reduced differentiation, but not with NF-YAl-overexpressing clones, in which differentiation was enhanced: as discussed previously, it is possible that adaptation to stable overexpression of these clones might account for the different behaviour [38]. Alternatively, it is possible that the precise timing and/or levels of NF-YAl is crucial: required at early stages, but detrimental in later stages.

The A1 and A2 helices mediate key functions of the protein, such as nuclear localization, HFD interactions and sequence-specific DNA-binding: the requirement for specificity in these tasks is enough to justify their remarkable evolutionary conservation in eukaryotes. Nuclear localization of the NF-Y trimer has been studied in detail. The HFD subunits enter nuclei as heterodimers, *via* interactions of the dimer, but not the single subunits, with the bi-directional receptor Importin 13 [23]. A very similar situation was also reported for related H2A/H2B-likes, NC2α/β, Pole3/Pole4 and Pole3/Chrac-17 [39,40]. On the other hand, NF-YA is imported through the Importin β system [23], specifically Imp-5 and Imp-8 [41]. Of particular relevance is the fact that mutagenesis of any stretch of three basic residues within the NF-YA homology region fails to abolish completely activity in human cells. Indeed, two NLSs were identified, and both the KRR (AA 276–278) and RRK (AA 294–296) -in the A1 and A2 respectively- or RKR (AA 310–312) had to be mutated for exclusion of the subunit from nuclei [23]. Of these stretches, the most important was clearly in the A1 (Fig. 7). Although the mutations introduced in the Kahle et al. study -K276 N-R277L-R278G, R294L-R295G-K296Q and R310L-K311A-R312G- were different from the Ala scan described here, the results are similar, in that the A1 mutation, which overlaps with YAm23 (L275A-K276A-R277A) and corresponds to YAm31 (K276A-R277A-R278A), has a larger impact with respect to the linker or A2 mutations. In the Kahle study, the two double mutants were defective, K276 N-R277L-R278G; R294L-R295G-K296Q and K276 N-R277L-R278G; R310L-K311A-R312G [23]: we do not have the equivalent of the former, but our double mutants YAm31–29 (K276-R277-R278; R312-G313-E314), YAm23–33 (L275-K276-R277; R302-H303-R304) and YAm23–34 (L275-K276-R277; R310-K311-R312) have only residual CPP activity. This suggests that the R302-H303-R304 stretch (not tested in the Kahle study) might help in CPP/NLS activity. Note that YAm32–29 (R294-R295-K296; R312-G313-E314), which has an intact A1, enters efficiently, further suggesting that this structure is of vital importance. Kahle et al. also reported that Importin 13 can compete with NF-YA for binding to the HFD dimer, which binds to A1, suggesting the importance of this helix both for nuclear localization and, as described here, for cell penetrating activity. In summary, we propose that CPP and NLS are composite signals relying on similar, if not identical, structures of NF-YA.

There are precedents of overlap between NLSs and CPPs. The IL-1α cytokine has an NLS/CPP peptide comparable in efficiency to the TAT peptide, which enables fused proteins to be targeted *in vivo*, that is in mice spleen, liver and intestine after peritoneal injection [42]. Relevant to our study, numerous transcriptional regulators, including TFs NF-κB, OCT6 and TCF1, were shown to have NLS peptides that could behave as carriers for cell entry [43]. OCT4 harbors a CPP, corresponding to the third helix of the homeodomain [44]. The fascinating question is whether the intrinsic cationic properties of some NLS will inevitably bring a CPP activity, or whether there might be a physiological role of the cell penetrating capacity. Concerning TFs, an hypothesis is that they might become signaling molecules in local paracrine circuits: for NF-YA this might be the case in macrophages. Monocytes undergo differentiation to macrophages by activation of sets of genes, many of which are CCAAT-dependent [45]. Among them, genes of the antigen presentation pathway, such as MHC Class I and II, are crucial for an effective immune response. These genes rely on the combination of NF-Y/RFX/CREB for activation [46]. Circulating monocytes contain NF-YB/NF-YC, but are devoid of NF-YA, which is upregulated/stabilized upon maturation and terminal differentiation [47,48]. An intriguing possibility is that NF-YA, released from the nuclei of cells infected by bacteria or viruses, might be taken up by circulating monocytes, form a complex with the constitutively expressed HFD subunits, contributing to the early activation of genes of the antigen presentation pathway and/or maturation of effector macrophages.

Finally, GST-TAT-NF-YAs was employed to increase the stem cells pools in HSCs preparations, thereafter used for BMT (Bone Marrow Transplantation) [12]. BMT is the treatment of choice in malignancies with bone marrow infiltration by cancer cells. The presence of internal CPP(s) in NF-YA is thus potentially relevant, and our study is encouraging to pursue experiments on primary cells *in vitro* with a TAT-less NF-YAs. However, there are hurdles, the major of which concerns the instability of the protein, as shown here, which makes a prolonged effect impossible to achieve. The lysines involved in poly-ubiquitination and degradation are all in the conserved domain, the process is regulated by acetylation, and Lys to Arg mutations increase NF-YA half-life in DNA transfections [22]. It is possible that use of such mutants will prolong the half-life of the protein, while not affecting CPP activity, nuclear retention and functionality of transfected NF-YA. A second issue concerns the role of NF-YA in cancer cells. Currently, there is a growing body of evidence linking NF-Y activity to cellular transformation, including reports of NF-YA being overexpressed in some tumors [49]; thus, it could boost proliferation of residual cancer cells of bone marrow, in addition to normal HSCs. On the other hand, overexpression of NF-YA was shown to have also pro-apoptotic functions, and this might depend upon the status -normal *versus* transformed- and/or the specific cell types. The reagents and assays developed here could be used to answer these questions.

Transparency document

The Transparency document associated with this article can be found, in online version.

Conflict of interest

The authors declare no conflicts of interest.

Acknowledgements

The work was supported by AFM Téléthon Trampoline Grant (n.16408) and AFM Téléthon Research Grant (n. 18364) and, partially, AIRC Grant (n.19050) to R.M.

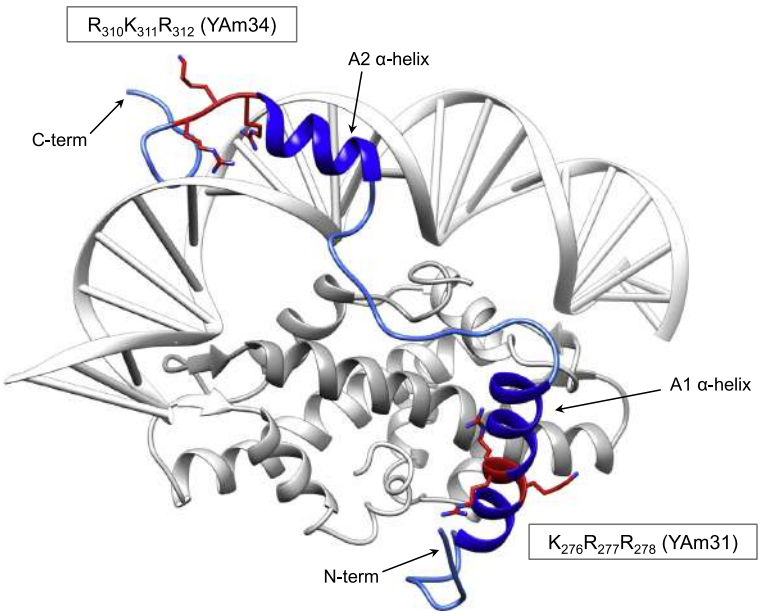


Fig. 7. Structural localization of NF-YA major CPP residues.
Ribbon representation of core NF-Y/DNA complex crystal structure (PDB ID: 4AWL). NF-YA subunit is depicted in blue. Amino acids mutated in YAm31 (K276R277R278) and YAm34 (R310K311R312) mutants presented in this work are highlighted in red. Image created with UCSF Chimera tool.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.bbamcr.2018.10.004.

References

[1] F. Milletti, Cell-penetrating peptides: classes, origin, and current landscape, Drug Discov. Today 17 (2012) 850–860.
[2] P. Agrawal, S. Bhalla, S.S. Usmani, S. Singh, K. Chaudhary, G.P. Raghava, A. Gautam, CPPsite 2.0: a repository of experimentally validated cell-penetrating peptides, Nucleic Acids Res. 44 (2016) D1098–D1103.
[3] A.D. Frankel, C.O. Pabo, Cellular uptake of the tat protein from human immunodeficiency virus, Cell 55 (1988) 1189–1193.
[4] M. Green, P.M. Loewenstein, Autonomous functional domains of chemically synthesized human immunodeficiency virus tat trans-activator protein, Cell 55 (1988) 1179–1188.
[5] A. Joliot, C. Pernelle, H. Deagostini-Bazin, A. Prochiantz, *Antennapedia* homeobox peptide regulates neural morphogenesis, Proc. Natl. Acad. Sci. U. S. A. 88 (1991) 1864–1868.
[6] F. Perez, A. Joliot, E. Bloch-Gallego, A. Zahraoui, A. Triller, A. Prochiantz, *Antennapedia* homeobox as a signal for the cellular internalization and nuclear addressing of a small exogenous peptide, J. Cell Sci. 102 (1992) 717–722.
[7] E. Vives, P. Brodin, B. Lebleu, D. Derossi, A.H. Joliot, G. Chassaing, A. Prochiantz, The third helix of the *Antennapedia* homeodomain translocates through biological membranes, J. Biol. Chem. 269 (1994) 10444–10450.
[8] D. Derossi, S. Calvet, A. Trembleau, A. Brunissen, G. Chassaing, A. Prochiantz, Cell internalization of the third helix of the *Antennapedia* homeodomain is receptor-independent, J. Biol. Chem. 271 (1996) 18188–18193.
[9] S. Futaki, T. Suzuki, W. Ohashi, T. Yagami, S. Tanaka, K. Ueda, Y. Sugiura, Arginine-rich peptides. An abundant source of membrane-permeable peptides having potential as carriers for intracellular protein delivery, J. Biol. Chem. 276 (2001) 5836–5840.
[10] F. Duchardt, M. Fotin-Mleczek, H. Schwarz, R. Fischer, R. Brock, A comprehensive model for the cellular uptake of cationic cell-penetrating peptides, Traffic 8 (2007) 848–866.
[11] W.B. Kauffman, T. Fuselier, J. He, W.C. Wimley, Mechanism matters: a taxonomy of cell penetrating peptides, Trends Biochem. Sci. 40 (2015) 749–764.
[12] A.D. Domashenko, G. Danet-Desnoyers, A. Aron, M.P. Carroll, S.G. Emerson, TAT-mediated transduction of NF-Ya peptide induces the *ex vivo* proliferation and engraftment potential of human hematopoietic progenitor cells, Blood 116 (2010) 2676–2683.
[13] D. Dolfini, M. Minuzzo, G. Pavesi, R. Mantovani, The short isoform of NF-YA belongs to the embryonic stem cell transcription factor circuitry, Stem Cells 30 (2012) 2450–2459.
[14] M. Mojsin, V. Topalovic, J. Marjanovic Vicentic, M. Stevanovic, Transcription factor NF-Y inhibits cell growth and decreases SOX2 expression in human embryonal carcinoma cell line NT2/D1, Biochemistry (Mosc) 80 (2015) 202–207.
[15] J.D. Fleming, G. Pavesi, P. Benatti, C. Imbriano, R. Mantovani, K. Struhl, NF-Y coassociates with FOS at promoters, enhancers, repetitive elements, and inactive chromatin regions, and is stereo-positioned with growth-controlling transcription factors, Genome Res. 23 (2013) 1195–1209.
[16] A.J. Oldfield, P. Yang, A.E. Conway, S. Cinghu, J.M. Freudenberg, S. Yellaboina, R. Jothi, Histone-fold domain protein NF-Y promotes chromatin accessibility for cell type-specific master transcription factors, Mol. Cell 55 (2014) 708–722.
[17] R.I. Sherwood, T. Hashimoto, C.W. O'Donnell, S. Lewis, A.A. Barkal, J.P. van Hoff, V. Karun, T. Jaakkola, D.K. Gifford, Discovery of directional and nondirectional pioneer transcription factors by modeling DNase profile magnitude and shape, Nat. Biotechnol. 32 (2014) 171–178.
[18] F. Lu, Y. Liu, A. Inoue, T. Suzuki, K. Zhao, Y. Zhang, Establishing chromatin regulatory landscape during mouse preimplantation development, Cell 165 (2016) 1375–1388.
[19] C. Romier, F. Cocchiarella, R. Mantovani, D. Moras, The NF-YB/NF-YC structure gives insight into DNA binding and transcription regulation by CCAAT factor NF-Y, J. Biol. Chem. 278 (2003) 1336–1345.
[20] V. Nardone, A. Chaves-Sanjuan, M. Nardini, Structural determinants for NF-Y/DNA interaction at the CCAAT box, Biochim. Biophys. Acta 1860 (5) (2017) 571–580.
[21] P. Hortschansky, H. Haas, E.M. Huber, M. Groll, A.A. Brakhage, The CCAAT-binding complex (CBC) in aspergillus species, Biochim. Biophys. Acta 1860 (5) (2017) 560–570.
[22] I. Manni, G. Caretti, S. Artuso, A. Gurtner, V. Emiliozzi, A. Sacchi, R. Mantovani, G. Piaggio, Posttranslational regulation of NF-YA modulates NF-Y transcriptional activity, Mol. Biol. Cell 19 (2008) 5203–5213.
[23] J. Kahle, M. Baake, D. Doenecke, W. Albig, Subunits of the heterotrimeric transcription factor NF-Y are imported into the nucleus by distinct pathways involving importin beta and importin 13, Mol. Cell. Biol. 25 (2005) 5339–5354.
[24] M. Frontini, C. Imbriano, I. Manni, R. Mantovani, Cell-cycle regulation of NF-YC nuclear localization, Cell Cycle 3 (2004) 217–222.
[25] X.Y. Li, R. Hooft van Huijsduijnen, R. Mantovani, C. Benoist, D. Mathis, Intron-exon organization of the NF-Y genes. Tissue-specific splicing modifies an activation domain, J. Biol. Chem. 267 (1992) 8984–8990.
[26] V. Basile, F. Baruffaldi, D. Dolfini, S. Belluti, P. Benatti, L. Ricci, V. Artusi, E. Tagliafico, R. Mantovani, S. Molinari, C. Imbriano, NF-YA splice variants have different roles on muscle differentiation, Biochim. Biophys. Acta 1859 (2016) 627–638.
[27] R. Mantovani, U. Pessara, F. Tronche, X.-Y. Li, A.M. Knapp, J.L. Pasquali, C. Benoist, D. Mathis, Monoclonal antibodies to NF-Y define its function in MHC class II and albumin gene transcription, EMBO J. 11 (1992) 3315–3322.
[28] R. Mantovani, X.-Y. Li, U. Pessara, R. van Huijsduijnen, C. Benoist, D. Mathis, Dominant negative analogs of NF-YA, J. Biol. Chem. 269 (1994) 20340–20346.
[29] S. Liu, R. Tobias, S. McClure, G. Styba, Q. Shi, G. Jackowski, Removal of endotoxin from recombinant protein preparations, Clin. Biochem. 30 (1997) 455–463.
[30] C. Liberati, A. di Silvio, S. Ottolenghi, R. Mantovani, NF-Y binding to twin CCAAT boxes: role of Q-rich domains and histone fold helices, J. Mol. Biol. 285 (4) (1999) 1441–1455.
[31] M. Nardini, N. Gnesutta, G. Donati, R. Gatta, C. Forni, A. Fossati, C. Vonrhein, D. Moras, C. Romier, M. Bolognesi, R. Mantovani, Sequence-specific transcription factor NF-Y displays histone-like DNA binding and H2B-like ubiquitination, Cell 152 (2013) 132–143.
[32] M. Ceribelli, P. Benatti, C. Imbriano, R. Mantovani, NF-YC complexity is generated

by dual promoters and alternative splicing, J. Biol. Chem. 284 (2009) 34189–34200.
[33] S.N. Maity, B. de Crombrugghe, Biochemical analysis of the B subunit of the heteromeric CCAAT-binding factor. A DNA-binding domain and a subunit interaction domain are specified by two separate segments, J. Biol. Chem. 267 (12) (1992) 8286–8292.
[34] Y. Xing, J.D. Fikes, L. Guarente, Mutations in yeast HAP2/HAP3 define a hybrid CCAAT box binding domain, EMBO J. 12 (1993) 4647–4655.
[35] Y. Xing, S. Zhang, J.T. Olesen, A. Rich, L. Guarente, Subunit interaction in the CCAAT-binding heteromeric complex is mediated by a very short alpha-helix in HAP2, Proc. Natl. Acad. Sci. U. S. A. 91 (1994) 3009–3013.
[36] A. Farina, I. Manni, G. Fontemaggi, M. Tiainen, C. Cenciarelli, M. Bellorini, R. Mantovani, A. Sacchi, G. Piaggio, Down-regulation of cyclin B1 gene transcription in terminally differentiated skeletal muscle cells is associated with loss of functional CCAAT-binding NF-Y complex, Oncogene 18 (18) (1999) 2818–2827.
[37] A. Gurtner, I. Manni, P. Fuschi, R. Mantovani, F. Guadagni, A. Sacchi, G. Piaggio, Requirement for down-regulation of the CCAAT-binding activity of the NF-Y transcription factor during skeletal muscle differentiation, Mol. Biol. Cell 14 (7) (2003) 2706–2715.
[38] V. Basile, F. Baruffaldi, D. Dolfini, S. Belluti, P. Benatti, L. Ricci, V. Artusi, E. Tagliafico, R. Mantovani, S. Molinari, C. Imbriano, NF-YA splice variants have different roles on muscle differentiation, Biochim. Biophys. Acta 1859 (4) (2016) 627–638.
[39] J. Kahle, E. Piaia, S. Neimanis, M. Meisterernst, D. Doenecke, Regulation of nuclear import and export of negative cofactor 2, J. Biol. Chem. 284 (14) (2009) 9382–9393.
[40] P. Walker, D. Doenecke, J. Kahle, Importin 13 mediates nuclear import of histone fold-containing chromatin accessibility complex heterodimers, J. Biol. Chem. 284 (17) (2009) 11652–11662.
[41] M. Kimura, Y. Morinaka, K. Imai, S. Kose, P. Horton, N. Imamoto, Extensive cargo identification reveals distinct biological roles of the 12 importin pathways, Elife 6 (2017) (pii: e21184).
[42] C. El Chartouni, C. Benner, M. Eigner, M. Lichtinger, M. Rehli, Transcriptional effects of colony-stimulating factor-1 in mouse macrophages, Immunobiology 215 (2010) 466–474.
[43] J.H. Koo, H. Yoon, W.J. Kim, S. Lim, H.J. Park, J.M. Choi, Cell membrane penetrating function of the nuclear localization sequence in human cytokine IL-1α, Mol. Biol. Rep. 41 (12) (2014) 8117–8126.
[44] A.D. Ragin, R.A. Morgan, J. Chmielewski, Cellular import mediated by nuclear localization signal peptide sequences, Chem. Biol. 9 (8) (2002) 943–948.
[45] E. Harreither, H.A. Rydberg, H.L. Amand, V. Jadhav, L. Fliedl, C. Benda, M.A. Esteban, D. Pei, N. Borth, R. Grillari-Voglauer, O. Hommerding, F. Edenhofer, B. Nordén, J. Grillari, Characterization of a novel cell penetrating peptide derived from human Oct4, Cell Regen. 3 (1) (2014) 2.
[46] N. Sachini, J. Papamatheakis, NF-Y and the immune response: dissecting the complex regulation of MHC genes, Biochim. Biophys. Acta 1860 (5) (2017) 537–542.
[47] G. Marziali, E. Perrotti, R. Ilari, U. Testa, E.M. Coccia, A. Battistini, Transcriptional regulation of the ferritin heavy-chain gene: the activity of the CCAAT binding factor NF-Y is modulated in heme-treated friend leukemia cells and during monocyte-to-macrophage differentiation, Mol. Cell. Biol. 17 (1997) 1387–1395.
[48] G. Marziali, E. Perrotti, R. Ilari, E.M. Coccia, R. Mantovani, U. Testa, A. Battistini, The activity of the CCAAT-box binding factor NF-Y is modulated through the regulated expression of its a subunit during monocyte to macrophage differentiation: regulation of tissue-specific genes through a ubiquitous transcription factor, Blood 93 (1999) 519–526.
[49] A. Gurtner, I. Manni, G. Piaggio, NF-Y in cancer: impact on cell transformation of a gene essential for proliferation, Biochim. Biophys. Acta 1860 (5) (2017) 604–616 (pii: S1874-9399(16)30283-8).

“The Switch from NF-YAl to NF-YAs Isoform Impairs Myotubes Formation”

Contribution to the published article

The article “The NF-YA long isoform governs early phases of myotubes differentiation” submitted to Cells Journal and attached above is related to the work performed during my PhD thesis. My contribution to this study started from strategy design to specifically ablate exon 3 in mouse C2C12 cells through the CRISPR system using the Cas9 nickase in order to minimize off-target effects. I designed two couple of gRNAs checking for absence of common genomic targets. I screened 335 clones and I found two positive NF-YAl-KO clones (#83 and #117). Once confirmed morphological similarity, NF-YAl-KO clones were analyzed for their capability to differentiate: 72 hours after differentiation induction, only wt cells showed the classical myotubes formation while NF-YAl-KO clones remained single cells with no expression of the muscle differentiation marker MyHCs. I performed all gene expression analyses checking for up or downregulated genes in NF-YAl-KO clones compared to the control in GM and 24h DM. Interestingly, mRMA levels of two genes necessary for muscle cells fusion (Mymk and Mymx) result downregulated both in GM and 24h DM, as Myogenin. Analysis on Myogenin and MyoD protein levels revealed a downregulation of these two muscle-specific TFs, but Chip experiments using NF-YB antibody revealed that the downregulation is due to an indirect regulation.

The data produced during my PhD thesis and showed in this manuscript highlighted that NF-YA long isoform governs early phases of myotubes differentiation trough an indirect regulation.

cells

MDPI

Article

The Switch from NF-YAl to NF-YAs Isoform Impairs Myotubes Formation

Debora Libetti, Andrea Bernardini, Sarah Sertic, Graziella Messina, Diletta Dolfini and Roberto Mantovani *

Dipartimento di Bioscienze, Università degli Studi di Milano, Via Celoria 26, 20133 Milano, Italy; debora.libetti@unimi.it (D.L.); andrea.bernardini@unimi.it (A.B.); sarah.sertic@unimi.it (S.S.); graziella.messina@unimi.it (G.M.); diletta.dolfini@unimi.it (D.D.)

* Correspondence: mantor@unimi.it

Received: 14 February 2020; Accepted: 21 March 2020; Published: 24 March 2020

Abstract: NF-YA, the regulatory subunit of the trimeric transcription factor (TF) NF-Y, is regulated by alternative splicing (AS) generating two major isoforms, "long" (NF-YAl) and "short" (NF-YAs). Muscle cells express NF-YAl. We ablated exon 3 in mouse C2C12 cells by a four-guide CRISPR/Cas9n strategy, obtaining clones expressing exclusively NF-YAs (C2-YAl-KO). C2-YAl-KO cells grow normally, but are unable to differentiate. Myogenin and—to a lesser extent, MyoD— levels are substantially lower in C2-YAl-KO, before and after differentiation. Expression of the fusogenic Myomaker and Myomixer genes, crucial for the early phases of the process, is not induced. Myomaker and Myomixer promoters are bound by MyoD and Myogenin, and Myogenin overexpression induces their expression in C2-YAl-KO. NF-Y inactivation reduces MyoD and Myogenin, but not directly: the Myogenin promoter is CCAAT-less, and the canonical CCAAT of the MyoD promoter is not bound by NF-Y in vivo. We propose that NF-YAl, but not NF-YAs, maintains muscle commitment by indirectly regulating Myogenin and MyoD expression in C2C12 cells. These experiments are the first genetic evidence that the two NF-YA isoforms have functionally distinct roles.

Keywords: splicing isoforms; CRISPR-Cas9; exon deletion; NF-Y; muscle differentiation; C2C12 cells

1. Introduction

Cell specification and differentiation during development of multicellular organisms is a complex set of events resulting in the formation of organs, whose physiology is maintained by a balance of cell proliferation and differentiation. A paradigmatic example of these phenomena is formation of skeletal muscle. In the case of mammals—mouse in particular—the process begins at early developmental stages, proceeding through embryonic, fetal and adult stages [1,2]. Sequence-specific transcription factors—TFs—play a central role in specifying the identities of myoblasts, their migration to different body locations, organization and the capacity to self-renew and differentiate into myotubes. These properties are key to guarantee maintenance and functionality of the different muscles throughout the lifespan of the organism, including repair after injury in adult life. A set of four key TFs —MyoD, Myf5, Myogenin, MRF4, termed myogenic regulatory factors (MRFs)—have been identified and thoroughly studied by genetic and biochemical means for their capacity to specify myoblasts identity [3,4]. During development, PAX3/7 are located upstream of MRFs [5]; downstream are many TFs, such as the MADS box MEF2A/C/D [6,7], the bHLH ID1/3 [8–10] and SNAI1 [11], the HOX SIX1/4/5 [12–15], STAT3 [16], NFIX [17,18] and the ZNF KLF2/4/5 [19,20]. Unlike MRFs, most of these TFs are not expressed predominantly in muscle cells and are equally important for development and differentiation of other tissues and organs [21–25].

NF-Y is an evolutionarily conserved heterotrimer formed by the sequence-specific NF-YA and the Histone Fold Domain—HFD—NF-YB/NF-YC [26]. The sequence recognized by NF-Y is the CCAAT box, which plays an important role in the activation of 25%–30% of mammalian genes. NF-Y has been classified as "pioneer" TF, in mammals and plants [27–31]. NF-YA is the regulatory subunit; it is alternatively spliced, generating two major isoforms "short" (NF-YAs) and "long" (NF-YAl), differing in 28 amino acids coded by exon 3 [32]. This stretch is located at the N-terminus of the protein, in the Gln-rich transactivation domain (TAD). NF-YAs and NF-YAl have identical subunits-interactions and DNA-binding properties *in vitro*; ChIP-seq from cells harboring predominantly either one of the two isoforms showed recovery of peaks enriched in CCAAT. The isoforms are expressed at various levels in different tissues and cell lines [32,33]. Importantly, no cell line has been so far described lacking NF-YA—nor the HFDs—and NF-YA inactivation was reported to be fatal to cells [28,34]. NF-YAl is the predominant isoform in muscle C2C12 cells: it is abundant in proliferating cells, but it drops to low levels following terminal differentiation to myotubes, unlike the HFD partners [35–37]. Highly reduced NF-YA protein was found in myotubes of adult mice [38]. This suggested that genes up-regulated in the terminal phases of muscle differentiation are either CCAAT-less or not NF-Y-dependent, whereas the trimer activates cell-cycle and growth-promoting genes required during the proliferative state. However, overexpression of NF-YAl led to improved differentiation of C2C12 [39], suggesting that NF-YAl does take part in the differentiation process.

For decades, C2C12 myoblast cells have represented an informative tool to identify genes involved in muscle differentiation [40]. Ablation of the whole NF-YA gene is early embryonic lethal [41], and KO in stable cell lines could not be generated so far. We investigated the role of NF-YAl by genetically ablating exon 3, leading to the production of an intact NF-YAs. We successfully generated homozygous C2C12 lines expressing only NF-YAs and went on to study differentiation properties.

2. Materials and Methods

2.1. Cell Culture and Proliferation Assay

Mouse myoblast cells (C2C12, ATCC) were cultured in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% Fetal Bovine Serum (FBS, Gibco-Thermo Fisher Scientific), 4 mM L-Glutamine, 100 units/mL penicillin and 100 µg/mL streptomycin (GM, growth medium), in a humidified 5% CO_2 atmosphere at 37 °C. C2C12 cells differentiation was induced plating cells in DMEM with 2% horse serum (Gibco-Thermo Fisher Scientific), 4 mM L-Glutamine, 100 units/mL penicillin and 100 µg/mL streptomycin (DM, differentiation medium). Proliferation assay was performed by plating 1.5×10^5 cells into a 12-well plate and counting every 24 h for 3 days, using the Trypan Blue dye exclusion test. All data were gathered from at least three independent biological replicates. Multiple comparisons were performed using the One-way ANOVA test.

2.2. Derivation of C2-YAl-KO Clones

To delete the exon 3 of NF-YA gene in C2C12 cells, four guide RNAs (gRNAs) were designed to simultaneously target the two flanking introns by using the online tool https://zlab.bio/guide-design-resources. Potential off-target sites were monitored by the online tool https://crispr.cos.uni-heidelberg.de: Table S1 shows the results of such analysis for the four guides. The selected gRNAs had no common off-target sites and were cloned in the two plasmids pX330A_D10A-1x2_ac and pX330A_ D10A-1x2_bd, following the Multiplex CRISPR/Cas9n Assembly System Kit protocol [42]. 1×10^6 C2C12 cells were transfected with 3 µg of the two gRNAs/CRISPR/Cas9n plasmids by electroporation and plated at low density. 72 h after transfection, single clones were picked, expanded and screened.

For DNA extraction, cells from the individual clones were washed with PBS, collected by scraping, lysed in 100 µL ice-cold lysis buffer (40 mM Tris-HCl, 2 mM EDTA, 0.08% SDS, 80 mM NaCl, 0.5 µg/µL Proteinase K) and incubated overnight at 37 °C in agitation. To precipitate DNA, 100 µL of ice-cold 2-propanol and 0.3 M NaAc were added, samples were mixed and centrifuged at 13,000 rpm for 30 min

at 4 °C. The pellet was washed with 150 µL of 70% ethanol, centrifuged at 13,000 rpm for 30 min at 4 °C. Supernatant was discarded, the pellet was dried and resuspended in 30 µL H_2O. The resulting DNAs were then screened for positive exon 3 deletion by PCR.

We screened a total of 335 individual clones and obtained 2 independent homozygously edited clones.

2.3. Protein Extraction and Western Blot Analysis

For Whole Cell Extracts preparation, cells were pelleted by centrifugation, resuspended in ice-cold RIPA buffer (10 mM TrisHCl pH 8.0, 1 mM EDTA, 0.5 mM EGTA, 0.1% SDS, 0.1% sodium deoxycholate, 140 mM NaCl, 1% Triton X-100, 1 mM PMSF, Protease inhibitor cocktail) and incubated for 30 min on ice, with occasional shaking. Samples were centrifuged at 13,000 rpm for 10 min at 4°C and the supernatant recovered and quantified using the Bradford protein assay.

20 µg of extracts were loaded on a 4–10% SDS-polyacrylamide gel and analyzed by Western blot using primary antibodies and a peroxidase-conjugate secondary antibody (Sigma-Aldrich). Primary antibodies: anti-NF-YA (G-2, Santa Cruz), anti-NF-YB (GeneSpin), anti-NF-YC (home-made) anti-Vinculin (H-10, Santa Cruz), anti-MyHCs (MF20, DHSB), anti-Myogenin (IF5D, DHSB), anti-MyoD (C-20, Santa Cruz), anti-Myf5 (C-20, Santa Cruz), anti-Pax3 (DHSB), anti-Snai1 (C15D3, Cell Signaling). Western blot experiments were performed on three independent biological replicates.

2.4. Reverse Transcriptase PCR and Real-Time PCR

RNA was isolated by the Tri Reagent (Sigma-Aldrich) protocol according to the manufacturer's instruction. The cDNA was produced starting from 1 µg of total RNA using the MMLV Reverse Transcription Mix (GeneSpin) and used for real-time PCR (SYBR® Green Master Mix, Bio-rad Laboratories) analysis. Real-time PCRs were performed with oligonucleotides designed to amplify 100–200 bp fragments (Table S2). The housekeeping gene Rsp15a was used to normalize expression data. The relative sample enrichment was calculated with the formula $2^{-(\Delta\Delta Ct)}$, where $\Delta\Delta Ct = [(Ct\ sample - Ct\ Rps15a)_x - (Ct\ sample - Ct\ Rps15a)_y]$, x = sample and y = sample control. RT-qPCR analyses were performed on three independent biological replicates. For ChIP experiments, we figured out the percentage of input immunoprecipitated by NF-YB and nc (negative control) antibodies. Results of three independent experiments were represented as Fold change (Fc) between NF-YB sample and nc sample as: %Input NF-YB/%Input nc.

2.5. Flow Cytometry Analyses

Cells were harvested by trypsinization and washed in PBS, fixed in ice-cold 70% ethanol and stored at 4 °C at least 24 hours. Cells were then washed with 1% BSA in PBS and resuspended in 500 µL of PI-staining solution (20 µg/mL Propidium Iodide, 10 µg/mL RNaseA, 1X PBS) at room temperature for 30 minutes, light protected. FACS analyses were performed using the BD FACSCantoII, analyzed with FACSDiva software and quantified with FlowJo. A total of 10^4 events were acquired for each sample. Three independent FACS experiments were performed.

2.6. Immunofluorescence

For immunofluorescence analyses, cells were washed three times with PBS and fixed 10 min with ice-cold acetone-methanol (1:1) at room temperature. After three washes, cells were permeabilized with 0.25% Triton X-100 in PBS for 5 min and incubated 1 h with the primary antibody anti-sarcomeric MyHCs (MF20, DHSB) at room temperature. Cells were washed three times, permeabilized 5 min with 0.25% Triton X-100 in PBS and incubated with secondary FITC anti-mouse antibody (1:500, Sigma-Aldrich) plus DAPI (2 µg/mL) for 40 min at room temperature, light protected. The acquisition was performed by using the inverted microscope Leica DMI6000 B. Three independent immunofluorescence experiments were performed.

2.7. Overexpression and RNA Interference Experiments

Myogenin overexpression was performed by electroporating 1×10^6 C2C12 cells with 3 µg of plasmid (pEMSV-Empty/pEMSV-Myog) and plating them in DM for 96 h. Cells were then collected and analyzed. Three independent biological replicates were performed.

For small interfering RNA (siRNA)-mediated knockdown of NF-YB [29], 2×10^6 C2C12 cells were transfected by electroporation with 100 nM of NF-YB [29] or scrambled control siRNA (Qiagen, SI01327193) and plated into a 10 cm plate in GM condition. 72 h after transfection, cells were collected by scraping for total protein preparation and RNA extraction. Gene expression was analyzed performing real-time PCR. Two independent biological replicates of siRNA interference were performed.

2.8. Chromatin Immunoprecipitation Assay (ChIP)

ChIPs were performed as described previously [43] with the following modifications. Briefly, 2×10^7 cells were crosslinked using 1% formaldehyde for 7 min, the reaction was quenched with 125 mM glycine and cells were collected by scraping. After lysis, nuclei were resuspended in Sonication buffer (50 mM Tris-HCl pH 8, 10 mM EDTA, 0.1% SDS, 0.5% sodium deoxycholate, protease Inhibitor cocktail) and sonicated (Bioruptor, Diagenode) to obtain fragments of approximately 150–300 bp, verified on agarose gel electrophoresis. Samples were centrifuged at 13,000 rpm for 10 min at 4 °C and supernatants recovered and quantified by Bradford assay. One hundred micrograms of chromatin were immunoprecipitated with 5 µg of anti-NF-YB (GeneSpin) and anti-FLAG (Sigma-Aldrich) antibodies. Protein-G beads (KPL) were used for recovery of antibody-bound proteins. Crosslinking was reversed by incubation at 65 °C overnight. Reactions were digested with RNase A and Proteinase K and DNA purified using the DNA purification kit (PCR clean Up, GeneSpin). The DNA was eluted in TE (10 mM Tris-HCl pH 8, 1 mM EDTA) and used in real-time PCR. Three independent biological replicates of ChIP experiments were performed.

3. Results

3.1. Ablation of NF-YA Exon 3 in C2C12 Cells by a Four Guides CRISPR/Cas9n Strategy

Mouse C2C12 cells mostly express NF-YAl [35–38]. To study the role of this isoform in maintenance and differentiation of C2C12, we set out a strategy to selectively eliminate exon 3, coding for the 28 extra amino acids present in NF-YAl and absent in NF-YAs. We figured that the use of four guides flanking precisely the exon 3 regions and of the single strand-cutting Cas9-nickase (Cas9n) would minimize off-target effects, which potentially affect the outcome of this technology [44]. Figure 1A shows the design of the four guide oligonucleotides, two couples targeting the 5′ and 3′ intronic DNA flaking exon 3, respectively. The two couples of oligos were first checked for absence of common genomic targets (Table S1) and cloned unpaired in the final pX330A_D10A-1x2_ad and pX330A_D10A-1x2_cb (Figure S1A), also expressing the Cas9n gene. The two plasmids were transfected in growing C2C12 cells by electroporation. Individual clones were isolated, expanded and analyzed by PCR, employing the amplicons shown in Figure 1B. As expected, the strategy was less efficient if compared to the standard use of two guides plus wt Cas9: 335 clones were individually screened and two were positive for correct ablation in homozygosity, as shown in Figure 1C. The results of PCRs of the two positive clones, #83 and #117, show the expected bands for the A, B and C amplicons, absent in the DNA of the parental C2C12 cells. The regions surrounding exon 3 in both clones were then amplified and sequenced: Figure S1B confirms the deletion of coding sequences of exon 3, with somewhat different ends in the two clones. In summary, we successfully ablated NF-YA exon 3, deriving two clones termed C2-YAl-KO. To the best of our knowledge, this is the second system of genome editing describing a clean deletion of an individual exon [45] and the first one employing the Cas9 nickase system coupled with four gRNAs.

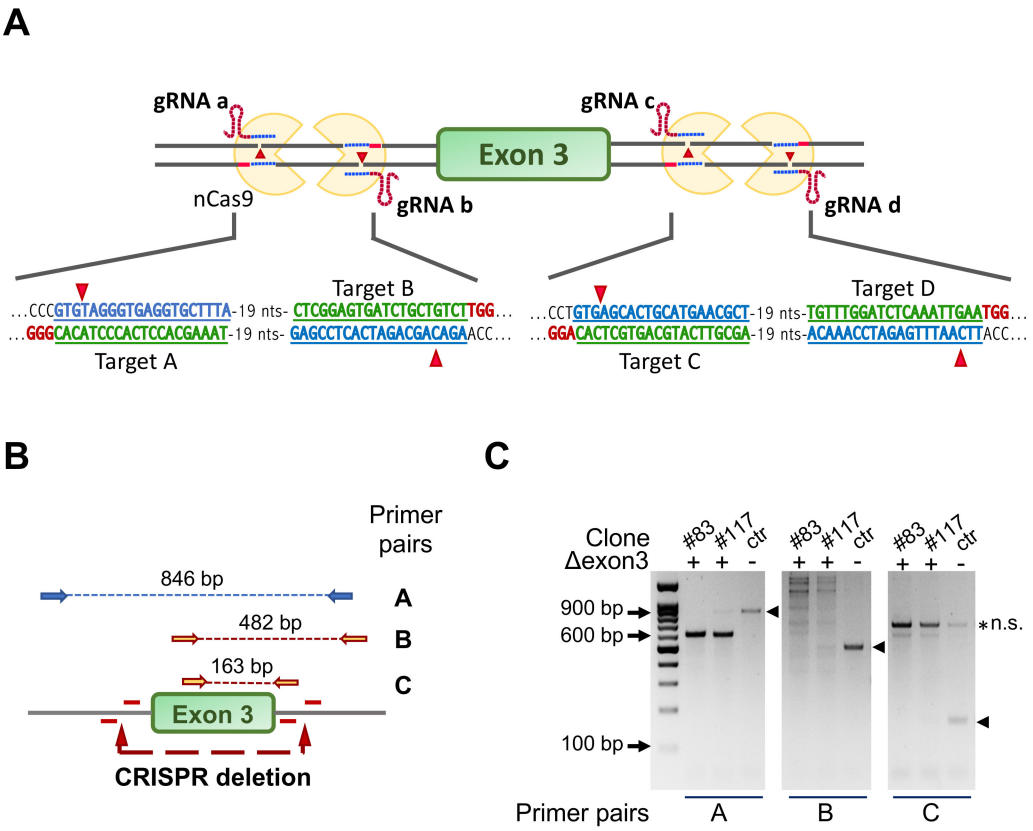


Figure 1. Strategy for ablation of NF-YA exon 3 in C2C12 cells using CRISPR/Cas9n and four gRNAs. (**A**) Gene editing strategy for NF-YA exon 3 deletion using the Cas9-nickase (Cas9n) and four guide RNAs. The targeted sequence by each guide RNA and the deletion sites are shown. Note that Cas9n cuts only the DNA strand that is complementary to and recognized by the gRNA, making necessary the simultaneous presence of two gRNAs/Cas9n complexes to induce a double-strand break (DSB). (**B**) The three primer pairs used to check for positive C2-YAl-KO clones are shown with the specific amplification products highlighted by the dashed lines. (**C**) Example of PCR products run into a 1.2% agarose gel. The expected bands in control cells (ctr) are marked with arrowheads; clones #83 and #117 represent positive C2-YAl-KO clones.

3.2. *Characterization of C2-YAl-KO Cells*

The two C2-YAl-KO clones were characterized first for expression of NF-YA. We performed qRT-PCR analysis with oligos specific for the individual isoforms [46]; Figure 2A shows that the NF-YAl mRNA is absent in the C2-YAl-KO clones. Extracts were prepared and Western blots performed: as expected, the parental C2C12 cells show expression of NF-YAl (Figure 2B). Instead, the clones express uniquely the NF-YAs isoform. We exposed the blots for long times to verify that no NF-YAl is visible in the two KO clones. Note that the levels of the two isoforms in parental cells—NF-YAl—and edited clones—NF-YAs—are essentially identical, as are the levels of NF-YB and NF-YC: since there is an important level of autoregulation among NF-Y subunits [47], this result indicates that HFD subunits are available for trimer formation and DNA-binding in C2C12 and C2-YAl-KO cells. In summary, genetic ablation of exon 3 in C2C12 was effective, leading to generation of clones that express uniquely the short isoform of NF-YA at physiological levels.

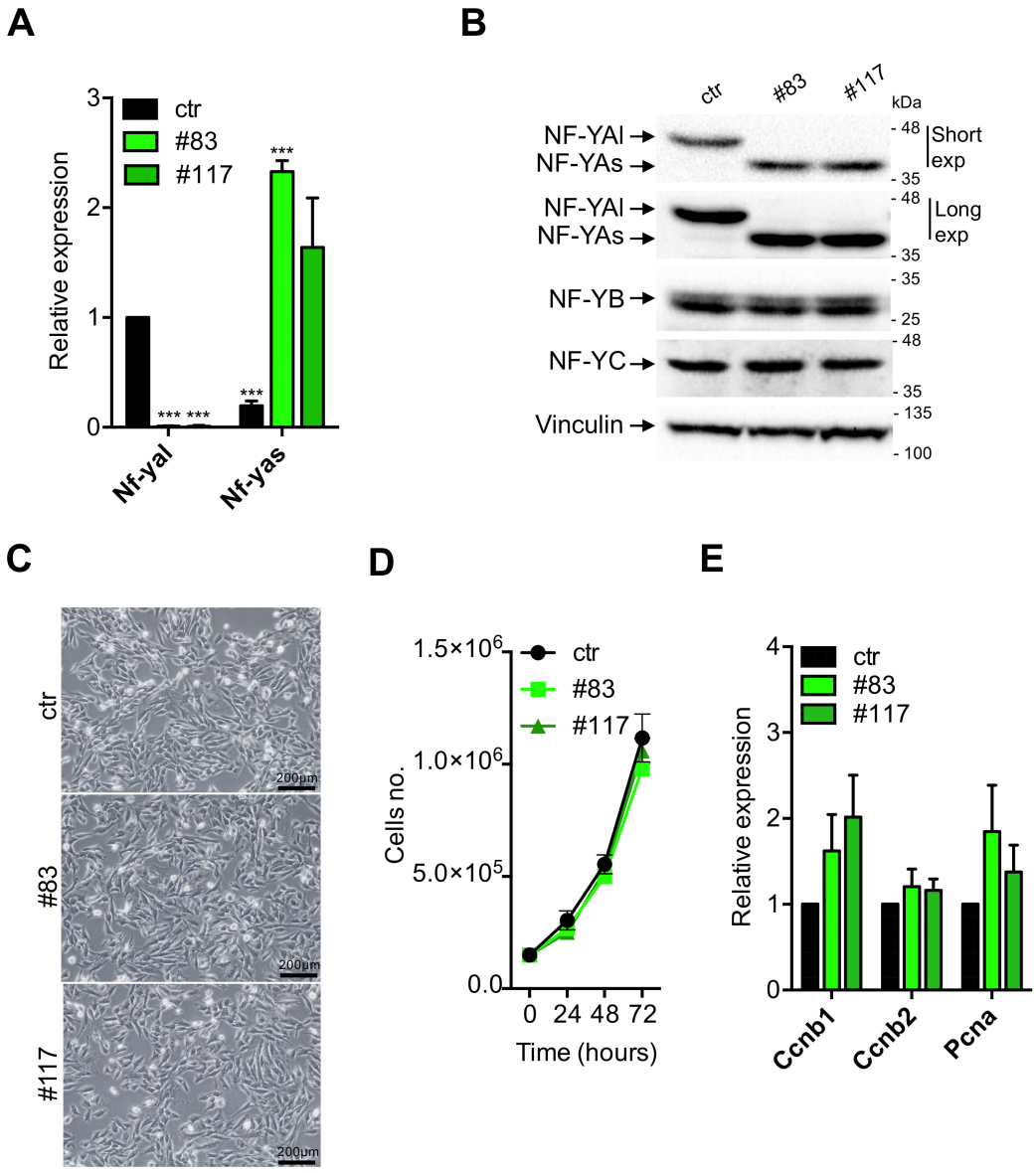


Figure 2. C2-YAl-KO clone characterization. (**A**) Gene expression analysis of NF-YA short and long levels in ctr and C2-YAl-KO clones (#83 and #117) in growth medium (GM) condition. Error bars represent the SD of three independent experiments. P-values were calculated using the one-sample t-test. (**B**) Western blot analysis of NF-Y protein subunits (NF-YA, NF-YB, NF-YC) in ctr cells and C2-YAl-KO clones (#83 and #117) in GM condition. For NF-YA isoforms analysis, short and long exposures are shown. Vinculin was used as loading control. (**C**) Phase contrast analysis of myoblast cells (ctr and C2-YAl-KO clones) morphology in GM condition. Scale bar 200 µm. (**D**) Proliferation assay performed in GM condition counting every 24 h for 3 days using the Trypan Blue dye exclusion test. Error bars represent the SD of three independent experiments. P-values were calculated using the one-way ANOVA test. (**E**) Gene expression analysis of key cell-cycle regulators in ctr and C2-YAl-KO clones (#83 and #117) in GM condition. Error bars represent the SD of three independent experiments. P-values were calculated using the one-sample t-test.

Next, we started to analyze the phenotype of the KO clones: they are stable upon repeated cycles of freezing and thawing and their morphology looks apparently similar to the parental C2C12 cells (Figure 2C). In mouse embryonic stem cells, expression of NF-YAs is associated with growth, and NF-YAl to differentiation [43]: in theory, NF-YAs-expressing C2C12 clones could be enhanced in proliferation. Cells were compared for growth under standard conditions: Figure 2D shows that

growth curves are similar, with the two edited clones being marginally slower. In FACS analysis, we did notice some differences: a higher number of S-phase and G2/M cells in the two clones (Figure S2, 21% and 28%, with respect to 18% in C2C12). We checked the mRNA levels of PCNA, Cyclin B1/B2: a slight increase of Cyclin B1 and PCNA in the KO clones is observed (Figure 2E); although not statistically significant, this is consistent with the FACS data. The most noticeable difference, however, was the lower number of sub-G1 cells: 6%–7% in the two clones compared to 12% in the parental C2C12 cells (Figure S2): such non cycling cells are possibly undergoing cell death, suggesting that the switch to NF-YAs is not provoking negative effects on cellular vitality, and, if anything, the opposite. In summary, C2-YAl-KO clones expressing NF-YAs have an apparently normal morphology, grow well, but not faster, with the expected partitioning in cell cycle phases, bar slightly elevated G2/M and decreased sub-G1 populations.

3.3. C2-YAl-KO Cells Fail to Differentiate and Fuse into Myotubes

The levels of NF-YAl drop following terminal differentiation of C2C12 cells and myotubes of mouse muscles show low-to-nil levels of NF-YAl [35–39]. To ascertain whether NF-YAs-expressing cells could form myotubes, we switched the parental C2C12 and the two C2-YAl-KO clones at 70%–80% confluence to a differentiation medium. Before and after 72 h, we monitored cell morphology, performed Immunofluorescence experiments and derived whole cell extracts. Figure 3A shows that parental C2C12 form well organized, multinucleated myotubes, as expected (Upper Panels). The average number of nuclei per fiber is 15, in keeping with an efficient process (Figure 3B). On the other hand, the two edited clones showed a dramatic lack of myotubes formation: cells did not fuse; they were disorganized (Figure 3A, lower panels). We reasoned that the process could be simply slower in these cells and prolonged differentiation up to 5 days: this did not lead to formation of myotubes, nor cell fusion in the C2-YAl-KO clones (not shown). Immunofluorescence and Western blot data are consistent: the MyHCs marker is clearly visible in IFs (Figure 3A, right panels) and WB (Figure 3C) in C2C12 cells after differentiation, but not in the two edited clones. Interestingly, the levels of Myogenin and MyoD were substantially lower both in growing cells and at these late stages of differentiation in C2-YAl-KO clones. As previously reported, NF-YAl, in C2C12, and NF-YAs, in the edited clones, are down-regulated after 72 h of differentiation; NF-YB remained unchanged (Figure 3C). In summary, we conclude that terminal differentiation is completely blocked in C2C12 cells expressing NF-YAs instead of NF-YAl.

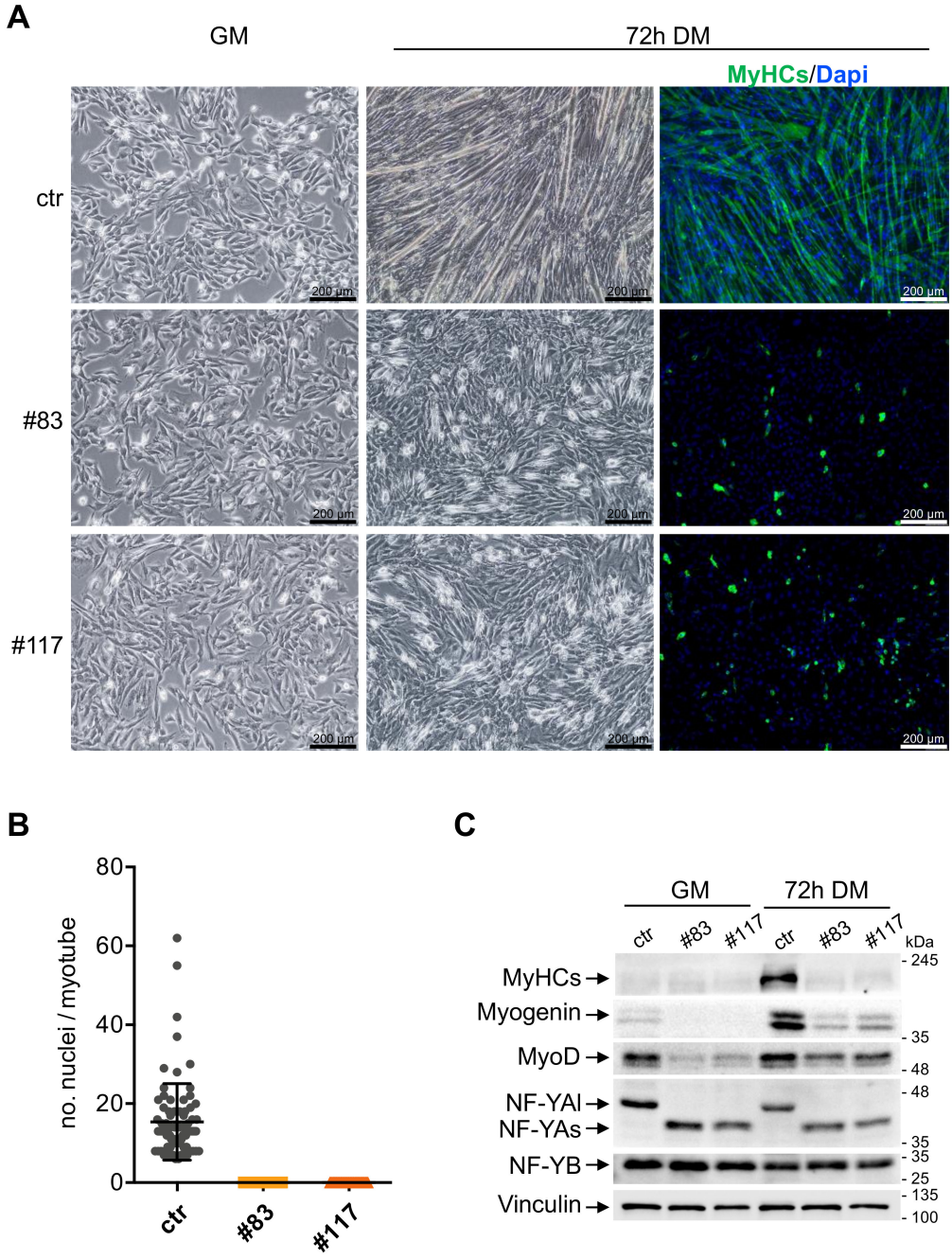


Figure 3. C2-YAl-KO clones fail to differentiate into myotubes. (**A**) Phase contrast analysis of myoblast cells (ctr and C2-YAl-KO clones) before and after 72 h of differentiation (differentiation medium (DM) condition) and immunofluorescence analyses after 72 h of differentiation. Antibody against all sarcomeric MyHCs and DAPI were used. (**B**) Fusion index was calculated as the number of nuclei in each myotube (with three or more nuclei). (**C**) Western blot analysis of key muscle differentiation regulators (MyHCs, MyoD), NF-YA isoforms (NF-YAl, NF-YAs) and NF-YB proteins, before (GM) and after 72 h of differentiation (72 h DM). Vinculin was used as loading control. The experiment was performed three times.

3.4. Expression of TFs in C2-YAl-KO

We analyzed expression of MRFs and TFs with a proven role in differentiation, in the parental and in the C2-YAl-KO cells under growing conditions and 24 h after differentiation. Profiling experiments established this as an early time point to detect significant changes in gene expression [48]. Note that

most of the TFs analyzed have CCAAT in promoters and some formally shown to be under NF-Y control. First, we verified expression levels of MRFs in parental C2C12 (Figure S3): Myogenin is robustly induced; MyoD is modestly increased; Myf5 is modestly decreased after differentiation; Mef2C, but not Mef2D, is robustly increased. These changes are in agreement with expectations [49]. At the same time, we analyzed other TFs shown to be important for muscle differentiation: Six1/4/5, Snai1, Stat3 and Klf5 are all increased upon C2C12 differentiation, Id1/3 are modestly decreased, Pax3 is unchanged (Figure S3). These results are also in agreement with published data. Having established that our differentiation program runs normally in C2C12 cells, we monitored expression of these genes in the C2-YAl-KO clones. The results are shown in Figure 4A for growing conditions and Figure 4B for differentiation. MRFs show the most conspicuous differences: Myogenin is almost undetectable in growing C2-YAl-KO clones and marginally increased upon differentiation. MyoD basal levels are normal, but induction is reduced upon differentiation, compared to parental C2C12. Myf5 expression is basally similar in the edited clones and higher after differentiation (Figure 4A,B). Mef2C levels are similar in growing conditions, but lower after differentiation: note that the levels are very low basally and differences with parental C2C12 cells are not statistically significant. Mef2D expression is identical in C2C12 and edited clones. As for the other TFs, Six1/4/5, Klf5 and Pax3 show similar expression patterns (Figure 4A,B). Minor changes are observed in growing conditions for Snai1, Stat3 and Id1 (one clone only) and for Id1 (same clone) after differentiation. Finally, Id3 shows somewhat higher levels before and after differentiation, but again, these changes are variable in the three experiments and thus not statistically significant.

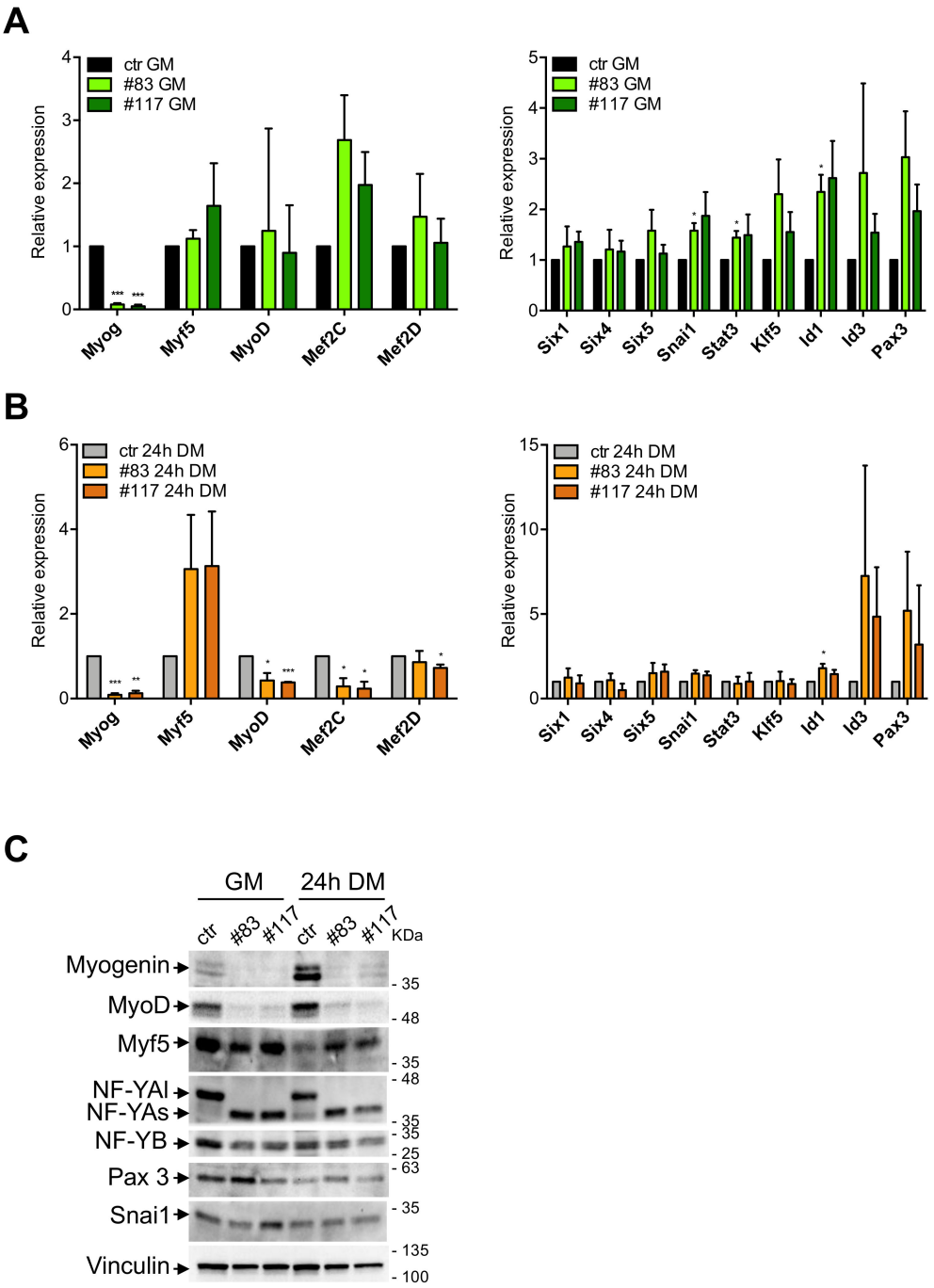


Figure 4. MRFs are downregulated in C2-YAl-KO clones. (**A**) Gene expression analysis of key muscle differentiation regulators (left panel) and other TFs shown to be important for muscle differentiation (right panel) in GM condition. Error bars represent the SD of three independent experiments. P-values were calculated using the one-sample t-test. (**B**) Gene expression analysis of key muscle differentiation regulators (left panel) and other TFs shown to be important for muscle differentiation (right panel) 24 h after differentiation (24 h DM). Error bars represent the SD of three independent experiments. P-values were calculated using the one-sample t-test. (**C**) Western blot analysis of key muscle differentiation regulators (Myogenin, MyoD, Myf5), NF-YA isoforms (NF-YAl, NF-YAs) and NF-YB proteins and other TFs shown to be important for muscle differentiation (Pax3, Snai1), in GM and 24 h DM. Vinculin was used as loading control.

To substantiate these results, protein expression of selected TFs was monitored by Western Blot analysis. Figure 4C shows that Myogenin levels are consistent with the mRNA data, being much lower in C2-YAl-KO clones than in parental cells, both in growing cells and after 24 h of differentiation. MyoD is substantially reduced in growing and differentiating clones, compared to parental C2C12. Note that protein levels were far lower than expected based on the mRNA levels, especially under growing conditions: this calls for post-transcriptional control in edited clones. Myf5 protein is downregulated in C2C12 after differentiation, as expected; in edited clones, it shows lower levels in growing cells, but higher after induction. NF-YA and NF-YB show the expected patterns; Pax3 is very modestly increased in C2-YAl-KO clones and Snai1 is unchanged. In summary, C2-YAl-KO cells have substantial differences in MRFs levels with respect to C2C12 cells, both before and after differentiation, whereas the other TFs showed rather minor changes.

3.5. Expression of Myomaker and Myomixer Is Activated by Myogenin and It Is Impaired in C2-YAl-KO

We were intrigued by the lack of cell fusion of the C2-YAl-KO clones after induction of differentiation. Myomaker—Mymk—and Myomixer—Mymx—are genes induced transcriptionally during muscle terminal differentiation, including in the C2C12 system [50,51]. Specifically, their expression is essential for the process of myocytes fusion [52]. We checked expression by qRT-PCR in parental C2C12 and in the two edited clones 24 h after differentiation. Figure 5A shows a strong induction—20-fold—of both Myomaker and Myomixer in C2C12 cells. C2-YAl-KO have much lower levels in growing cells (Figure 5B) and even more after differentiation (Figure 5C).

The obvious hypothesis was that these genes are under direct NF-Y control. We surveyed their promoter sequences and verified that no bona fide CCAAT box is present, notably within the evolutionary conserved areas: given the specificity of NF-Y CCAAT recognition, we considered unlikely that it acts directly on their expression. Genetic experiments in zebrafish have recently shown that Myomaker and Myomixer are directly activated by Myogenin [53]. We analyzed ENCODE datasets of C2C12 cells and found that Myogenin and MyoD target both promoters. Myomixer has apparently one promoter, Myomaker has two promoters, some 4 kb distant from each other: Figure 5D shows the overlapping peaks of Myogenin and MyoD. Myogenin binds exclusively after 24 h of differentiation, in accordance with its induced expression. One MyoD peak is visible already under growing conditions on Myomaker, and two additional peaks are found at 24 h. Importantly, the regions bound by MyoD and Myogenin in these two promoters are conserved across vertebrates, as shown by PhastCons data in Figure S4A: this corroborates the functional relevance proven in zebrafish [53]. To verify whether Myogenin activates Myomaker and Myomixer, we overexpressed it in parental C2C12 and in one of the C2-YAl-KO clones (#83) and induced to differentiate: Western blot of Figure 5E shows the increased levels of Myogenin compared to cells transfected with an Empty vector control; q-RT-PCR of Figure 5F shows that Myogenin overexpression has negligible effects on expression of the endogenous Myomaker and Myomixer in parental C2C12, but it increases expression of both genes in the C2-YAl-KO cells. Finally, morphological observation of the edited cells shows —incomplete—improvement in differentiation (Figure S4B).

In essence, we find that the marginal levels of Myogenin in C2-YAl-KO cells could result in lack of induction of the Myomaker and Myomixer targeted genes, entailing lack of cell fusion in NF-YAs-expressing clones.

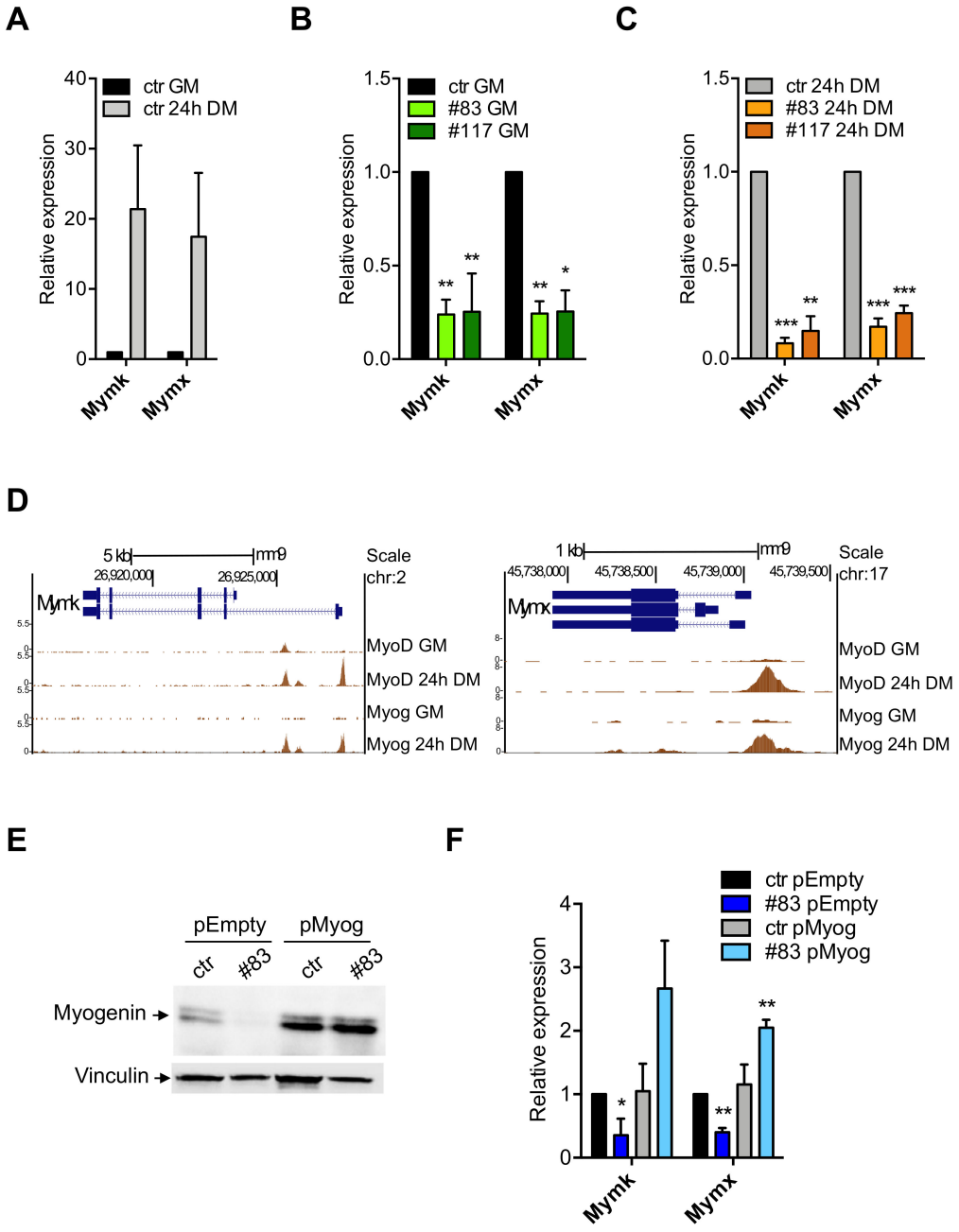


Figure 5. Myogenin directly regulates Myomaker and Myomixer expression. (**A**) Relative expression levels of Myomaker (Mymk) and Myomixer (Mymx) in C2C12 cells before and after 24 h of differentiation (24 h DM). Error bars represent the SD of three independent experiments. P-values were calculated using the one-sample t-test. (**B**,**C**) Relative expression levels of Myomaker (Mymk) and Myomixer (Mymx) in C2C12 cells before (**B**) and after 24 h of differentiation (**C**) in ctr and the two C2-YAl-KO clones. Error bars represent the SD of three independent experiments. P-values were calculated using the one-sample t-test. (**D**) ChIP-seq peaks of MyoD and Myogenin on Mymk and Mymx promoters in GM and after 24 h of differentiation (24 h DM) (UCSC-genome browser available tracks). Vertical viewing range Mymk: min 0, max 5.5. Vertical viewing range Mymx: min 0, max 8. (**E**) Western blot analysis of Myogenin protein levels in C2C12 cells transfected with a control plasmid (pEmpty) and the Myogenin-overexpressing plasmid (pMyog) 96 h after differentiation induction. Vinculin was used as loading control. (**F**) Relative expression levels of Mymk and Mymx in C2C12 Myog-overexpressing cells after 96 h of differentiation. Error bars represent the SD of three independent experiments. *p*-values were calculated using the one-sample t-test.

3.6. Myogenin and MyoD Are—Indirectly—Regulated by NF-Y

The results shown above beg the question as to whether NF-Y directly regulates MRFs. Myogenin and Myf5 promoters do not contain CCAAT boxes, MyoD does [54]. To verify the NF-Y dependence of these genes, we transitorily inactivated NF-Y activity. In our hands, NF-YA inactivations by shRNA or siRNA were rather inefficient in C2C12 cells (not shown). We thus turned to NF-YB by treating C2C12 cells with an siRNA previously shown to be active and very specific, including in profiling experiments [29]. NF-YB is a necessary component of the DNA-binding trimer: this allows us to inhibit CCAAT-binding activity, upon siRNA treatment. Most importantly, unlike NF-YA, NF-YB inactivation does not trigger apoptosis [29,34], making this a suitable choice for long differentiation processes. Figure 6 shows the results of experiment 1, Figure S5 those of experiment 2: in both, RT-qPCR (Figure 6A and Figure S5A) and Western blots (Figure 6B and Figure S5B) show far lower expression of NF-YB in C2C12 cells treated with NF-YB siRNA, with respect to the control siRNA. In mRNA analysis, Myogenin, MyoD and Mef2C, but not Myf5 nor Mef2D, are substantially downregulated upon NF-Y inactivation; Myomaker and Myomixer are also reduced. Six1/4/5 are reduced: for Six4, this in keeping with an NF-Y dependence predicted from previous data on NF-Y binding to a canonical promoter CCAAT [39]. As for Id1 and Id3, they are somewhat reduced, but the results are borderline significant: Id1 in experiment 2 and Id3 in experiment 1. We conclude that NF-Y removal entails a reduction of MRFs, which, in turn, could explain the observed drop of Myomaker and Myomixer. We also show that members of the Six family are NF-Y targets. Analysis of proteins levels in extracts of siRNA-inactivated cells by Western blots confirmed these results: the levels of NF-YB were lower (although not to the extent of the mRNA) and paralleled by somewhat lower levels of NF-YA. Myogenin is substantially decreased and MyoD is also affected, to a lesser extent (Figure 6B and Figure S5B). We conclude that NF-Y regulates the expression of MyoD and Myogenin in C2C12 cells.

The Myogenin promoter is CCAAT-less and was not bound by NF-Y in C2C12 cells [39] and, despite the presence of a canonical CCAAT, the MyoD promoter was also not bound [39]. To understand whether the positive effect of NF-Y on MyoD is direct, we checked the parental C2C12 cells for the presence of NF-Y in ChIP experiments. Three independent experiments are shown in Figure 6C and Figure S5C. The absence of enrichment of NF-Y on MyoD is indeed confirmed, whereas the Stard4 positive control promoter is clearly bound. Equally positive was the promoter of Id1, but not that of Id3. Note that there is variability in the fold-enrichments in the three experiments: as this is high (from 60 to 800-folds), we consider quantitative changes difficult to interpret, especially when compared to completely negative promoters such as MyoD and Id3. Therefore, we conclude that NF-Y does not regulate MyoD directly—and despite promoter binding—NF-Y has modest effects on Id1 transcription in C2C12 cells.

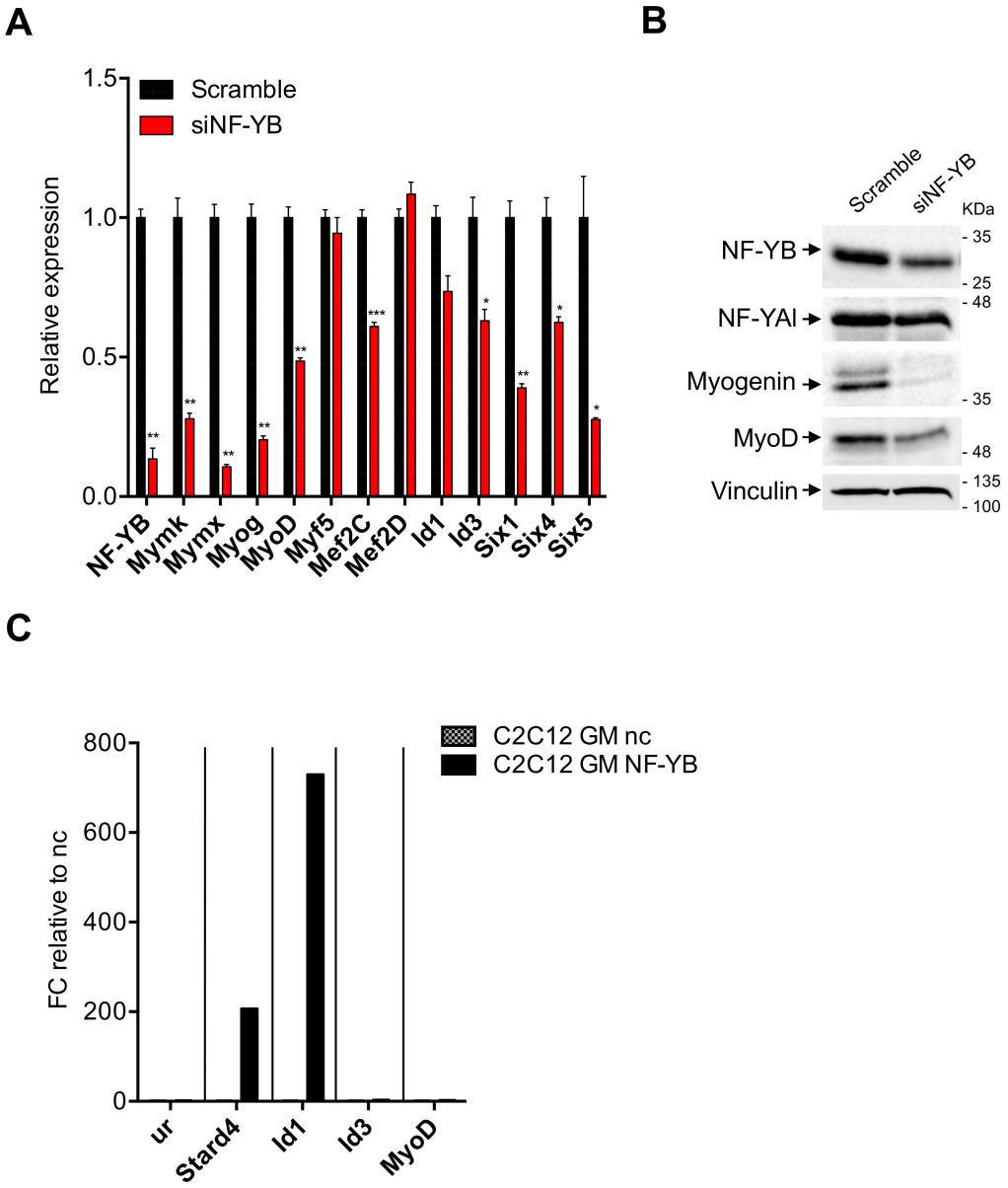


Figure 6. Analysis of NF-Y involvement in muscle specific genes expression. (**A**) Gene expression analysis of NF-YB and key muscle differentiation regulators in C2C12 cells 72 h after NF-YB silencing (siNF-YB) and scrambled siRNA control. Error bars represent the SD of two different RT-qPCR replicates. P-values were calculated using the one-sample t-test. (**B**) Western blot analysis of NF-YB, NF-YA and key muscle differentiation regulators (Myogenin, MyoD) protein levels 72 h after NF-YB silencing (siNF-YB) and the scrambled siRNA control. Vinculin was used as loading control. (**C**) ChIP experiment performed on C2C12 ctr cells in GM condition using NF-YB and negative control (nc) antibodies. The unrelated region (ur) and Stard4 were used as negative and positive control, respectively. Results are represented as the input percentage of each sample normalized to the input percentage of the nc antibody.

4. Discussion

By genome editing, we derived clones of C2C12 cells that express NF-YAs instead of NF-YAl. We verified that NF-YAs—and companion HFDs—are expressed at comparable levels and that it decreases after differentiation. The edited C2C12 clones are stable, grow normally, yet they are completely deficient in differentiation. We report defects of basal and induced expression of Myomaker and Myomixer, early response-genes likely responsible for lack of cell fusion. Their promoters are

targeted by MyoD and Myogenin. In turn, we find low—basal and induced—levels of MyoD and Myogenin in the NF-YAs-expressing clones. Finally, expression of both MRFs are indirectly controlled by NF-Y.

4.1. Role of NF-YA Alternative Splicing in Muscle Cells

Specific isoforms of TFs have long been known to impact heavily on transcriptional regulation. Paradigmatic examples are the members of the p53/p63/p73 families, whose isoforms, produced by multiple promoters and alternative splicing, have different targets and often opposing transcriptional effects [55]. The muscle system is no exception [56,57]. Mef2C and Mef2D undergo alternative splicing during muscle differentiation [57,58]: a muscle-specific isoform of Mef2D contains exon α2 rather than α1, both expressed in muscle cells. Growing and early differentiating cells harbors MEF2Dα1; the switch to MEF2Dα2 occurs in terminal stages of C2C12 differentiation, leading to activation of late genes. MEF2Dα1 is phosphorylated at two serines by PKA [59], which mediate association with HDACs, resulting in repression. MEF2Dα2 lacks these residues, functioning as a transcriptional activator. Parallel molecular mechanisms appear to be operating for the related MEF2Cα1/α2 alternative splicing isoforms [58]. The key issue in Mef2 splicing regulation is involvement in late stages of differentiation. Alternative splicing was reported for the master TFs of muscle commitment PAX3 and PAX7, but the functional roles of the single isoforms are less well characterized [60–64].

We show here that a switch from NF-YAl to NF-YAs causes a major difference in the differentiation properties of C2C12 cells. The major NF-YA isoforms, originally reported decades ago [32], are only recently attracting the attention they deserve. In part, this was due to the elusive logic of their expression patterns: in some systems, cells have NF-YAs before—and NF-YAl after—differentiation; in others, such as in muscle cells, NF-YAl is mostly found. In part, it was because of the rather unimpressive nature of the exon 3 amino acids incorporated into NF-YAl: a short stretch rich in glutamines and hydrophobic residues amid the larger transactivation domain. Overexpression experiments suggested differences in gene activation [39,65], but these experiments are to be taken with a grain of salt, because of the large amount of proteins produced, targeting the large number of potential NF-Y sites in the genome. NF-YA AS is likely more complex than what is shown here. First, NF-YAx is another alternatively spliced isoform, recently reported in glioblastomas, devoid of exons 3 and 5: this greatly reduces the activation domain, with important functional consequences [66]. Expression of NF-YAx will have to be monitored in normal cells, to verify whether it is specific for glioblastomas. Second, there are micro differences—6 amino acids—produced in many cell types within the acceptor site of exon 5. Third, some cells show the inclusion of an additional Gln residue at the acceptor splicing site of exon 3, producing a 29 amino acids insertion [32]. Note that a similar situation was reported for PAX3, in which an extra Gln causes differences in DNA-binding affinity [59]. Precise editing techniques, as we have started to use here in C2C12 cells, could sort out the functionality of the various isoforms.

4.2. NF-Y Does Not Target Directly Genes Involved in C2C12 Differentiation

Sequence-specific TFs target specific genomic sites, driven by the discriminatory power of their DNA-binding Domains. However, they are also known to be binding indirectly, being tethered by other TFs or complexes: analysis of genomic locations by ENCODE has shown that this latter mechanism is far from marginal [67]. In addition to ENCODE, several independent ChIP-seq of TFs—and cofactors—identified binding to CCAAT locations [68]. One such example regards the orphan receptor Rev-Erb, important for muscle regeneration, targeting NF-Y sites in C2C12 cells [69]. The reverse, namely NF-Y being tethered to CCAAT-less locations by other TFs, has yet to be described. The issue could theoretically be relevant, since the genes down-regulated after NF-Y removal, or by switching from NF-YAl to NF-YAs, have generally no CCCAT in promoters. The effects appear to be largely indirect, but we do not favor the promoter tethering hypothesis. Rather, we report binding of Myogenin and MyoD to the promoters of Myomaker and Myomixer and show that Myogenin overexpression leads to recovery of their expression in C2-YAl-KO cells. This extends to mouse cells

genetic experiments made in zebrafish [53]. It also indicates that NF-Y does not regulate other TFs essential for expression of these two genes. In summary, NF-Y/CCAAT interactions in promoters, which are structurally identical for NF-YAl and NF-YAs, are likely not crucial for genes induced during myotubes formation: rather, the focus is shifted to the control of MRFs, or other TFs.

We have analyzed expression of TFs involved in myoblast/C2C12 differentiation. The majority are not dramatically altered in edited clones. Mef2C induction is impaired, but previous studies indicated that NF-Y is bound to the Mef2D, not to Mef2C promoter [39]. We find that Mef2C, not Mef2D, is regulated by NF-YB RNAi interference. Note that these TFs are also targeted by MyoD and Myogenin, as they play a role in the final stages of differentiation [7,59]. This suggests indirect regulation by NF-Y via MRFs. Id1/Id3 do have bona fide functional CCAAT in promoters [70], bound in cancer cells as per ENCODE data (M. Ronzio, A.B., D.D., R.M., in preparation) and in NTera2 cells [71]: Id1, but not Id3, is bound in vivo by NF-Y in C2C12, parental cells and edited clones. The levels are decreased in C2-YAl-KO upon differentiation, but NF-Y-inactivation brings very marginal decrease in Id1 expression. PAX3, which acts upstream of MyoD, shows variable, somewhat increased mRNA levels in the edited clones, but this is not supported by analysis of protein levels. In summary, there is no clear CCAAT-driven TF that could explain the phenotype: instead, we propose that the decrease of Myogenin and MyoD expression entails a cascade of transcriptional events leading to failure of differentiation (Figure 7).

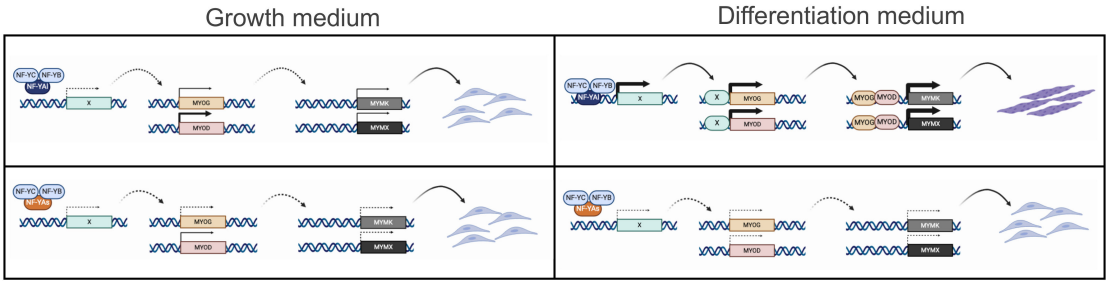


Figure 7. NF-YA isoforms involvement in regulation of expression of muscle genes. Model for NF-YA isoforms mediated regulation of expression of muscle genes in growth condition (left panel) and differentiation condition (right panel).

4.3. NF-Y Regulates MRFs Expression Indirectly

Switching from NF-YAl to NF-YAs—and NF-YB inactivation—negatively affects MRFs expression. Myf5 is moderately down in growing cells, remaining somewhat higher after differentiation. NF-Y-inactivation leads to a severe drop in Myogenin expression and a decrease of MyoD, which indicates an impact of NF-Y on their expression. The regulation appears to be transcriptional for Myogenin, not for MyoD, whose mRNA levels are variable, but overall similar. The Myogenin promoter is CCAAT-less and an indirect effect of NF-Y must be invoked. As for MyoD, the promoter harbors a high affinity NF-Y site, extremely conserved in evolution [54] and at the expected position (at -70 from TSS). Yet, NF-Y is not bound in vivo (Figure 6C). This is the only such example in nearly 200 promoters for which genetic analysis was reported [72]. The combination of an evolutionarily conserved, canonical CCAAT in a standard promoter position might function through NF-Y somewhen during the physiological activation of MyoD in development, while it has become expendable in the C2C12 system. Thus, down-regulation of MyoD in NF-YAs-expressing cells is also an indirect effect. It was proposed that MyoD serves as "pioneer" TF predisposing chromatin configurations for Myogenin to act as powerful activator of terminal differentiation genes and repressor of cell-cycle genes [73]. The latter function might be robustly counteracted by NF-YAs, but we have no evidence of that (Figure 2). It is now clear that the focus is set on transcriptional regulation of the MyoD and Myogenin units and on which activator TF(s)—or cofactor(s)—is under NF-YAl—but not NF-YAs—direct control.

For the time being, the "candidate" TFs approach used here failed to offer a plausible explanation on how NF-YAl regulates MRFs expression, thereby muscle differentiation. We must resolve to more systematic analysis, such as RNA-seq, to identify potential NF-Y-mediated regulators in C2C12. In light of the low intrinsic levels of muscle-commitment by MRFs in C2-YAl-KO clones, such analysis could also shed light on the actual identity of these cells.

Supplementary Materials: The following are available online at http://www.mdpi.com/2073-4409/9/3/789/s1, Figure S1: CRISPR/Cas9n system and ablation of NF-YA exon 3 in C2C12 cells. (A) Schematic representation of plasmids construction, following the Multiplex CRISPR/Cas9n Assembly System Kit protocol (Yamamoto lab) [40]. (B) Sequencing of the two C2-YAl-KO clones (#83, #117) compared to the control (ctr). Deleted sequence, targeted sequence and exon 3 sequence are highlighted. Figure S2: Cell-cycle analysis of C2-YAl-KO clones. Flow cytometry analysis of ctr C2C12 cells and the two C2-YAl-KO clones in GM condition. The analysis of three independent experiments and the average of percentage of cells in each cell-cycle phase are shown. Figure S3: Gene expression analysis of TFs in growing and differentiated C2C12 cells. Gene expression analysis by RT-qPCR of key muscle differentiation regulators (left panel) and other TFs shown to be important for muscle differentiation (right panel) in GM condition and 24 h after differentiation in C2C12 ctr cells. Error bars represent the SD of three different experiments. P-values were calculated using the one-sample t-test. Figure S4: Myomaker and Myomixer expression are regulated by MyoD and Myogenin. (A) UCSC view of Mymk and Mymk *loci* showing alignment of ChIP-Seq data and DNA regulatory motifs conserved across Vertebrates by PhastCons. (B) Phase-contrast analysis of C2C12 cells (ctr and #83) morphology transfected with pEmpty or pMyog, 96 h after differentiation. Figure S5: Analysis of NF-Y involvement in muscle specific genes expression. (A) Gene expression analysis by RT-qPCR of NF-YB and key muscle differentiation regulators in C2C12 cells 72 h after NF-YB silencing (siNF-YB) and the scrambled siRNA control (II° experiment). Error bars represent the SD of two different q-PCR replicates. *p*-values were calculated using the one-sample t-test. (B) Western blot analysis of NF-YB, NF-YA and key muscle differentiation regulators (Myogenin, MyoD) protein levels 72 h after NF-YB silencing (siNF-YB) and the scrambled control. Vinculin was used as loading control. (C) Analysis of II° and III° ChIP experiments performed on C2C12 ctr cells in GM condition using NF-YB antibody and the negative control (nc). The unrelated region (ur) and Stard4 were used as negative and positive control, respectively. Results are represented as the input percentage of sample normalized to the nc. Table S1: Off-targets analysis. Analysis of possible off-target sites of each gRNA using the online tool https://crispr.cos.uni-heidelberg.de. For each gRNA the off-target gene name, gene id, position (intronic, intergenic, exonic), mismatches (MM) and the PAM sequence are reported. Table S2: Primers used. The specific sequence of each primer (forward and reverse) used for RT-qPCR and ChIP analysis are reported.

Author Contributions: D.L.: investigation, formal analysis and visualization; A.B.: data curation and methodology; S.S.: investigation; G.M.: resources, writing (review and editing); D.D.: methodology, data curation, writing (review and editing); R.M.: supervision, funding acquisition, writing (Original draft and editing). All authors have read and agreed to the published version of the manuscript.

Funding: The work was supported by AFM Téléthon Trampoline Grant (n.16408), AIRC Grant (n.19050) and PRIN 2015 to R.M. and AFM Téléthon Research Grant (n. 18364) to G.M.

Acknowledgments: The authors would like to thank prof. Carol Imbriano (Università degli Studi di Modena e Reggio Emilia) for providing antibodies against Myod and Myf5.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Buckingham, M. Skeletal muscle formation in vertebrates. *Curr. Opin. Genet. Dev.* **2001**, *11*, 440–448. [CrossRef]
2. Buckingham, M.; Rigby, P.W. Gene Regulatory Networks and Transcriptional Mechanisms that Control Myogenesis. *Dev. Cell* **2014**, *28*, 225–238. [CrossRef]
3. Hernández-Hernández, J.M.; García-González, E.G.; Brun, C.E.; Rudnicki, M.A. The myogenic regulatory factors, determinants of muscle development, cell identity and regeneration. *Semin. Cell Dev. Boil.* **2017**, *72*, 10–18. [CrossRef] [PubMed]
4. Zammit, P.S. Function of the myogenic regulatory factors Myf5, MyoD, Myogenin and MRF4 in skeletal muscle, satellite cells and regenerative myogenesis. *Semin. Cell Dev. Boil.* **2017**, *72*, 19–32. [CrossRef] [PubMed]
5. Buckingham, M.; Relaix, F. PAX3 and PAX7 as upstream regulators of myogenesis. *Semin. Cell Dev. Boil.* **2015**, *44*, 115–125. [CrossRef] [PubMed]
6. Black, B.L.; Olson, E.N. Transcriptional Control Of Muscle Development by Myocyte Enhancer Factor-2 (MEF2) Proteins. *Annu. Rev. Cell Dev. Boil.* **1998**, *14*, 167–196. [CrossRef]

7. Taylor, M.V.; Hughes, S.M. Mef2 and the skeletal muscle differentiation program. *Semin. Cell Dev. Boil.* **2017**, *72*, 33–44. [CrossRef]
8. Kumar, D.; Shadrach, J.L.; Wagers, A.J.; Lassar, A.B. Id3 Is a Direct Transcriptional Target of Pax7 in Quiescent Satellite Cells. *Mol. Boil. Cell* **2009**, *20*, 3170–3177. [CrossRef]
9. Wu, J.; Lim, R.W. Regulation of inhibitor of differentiation gene 3 (Id3) expression by Sp2-motif binding factor in myogenic C2C12 cells: Downregulation of DNA binding activity following skeletal muscle differentiation. *Biochim. et Biophys. Acta (BBA) Gene Struct. Expr.* **2005**, *1731*, 13–22. [CrossRef]
10. Atherton, G.T.; Travers, H.; Deed, R.; Norton, J.D. Regulation of cell differentiation in C2C12 myoblasts by the Id3 helix-loop-helix protein. *Cell Growth Differ. Mol. Boil. J. Am. Assoc. Cancer Res.* **1996**, *7*, 1059–1066.
11. Soleimani, V.D.; Yin, H.; Jahani-Asl, A.; Ming, H.; Kockx, C.; Van Ijcken, W.F.J.; Grosveld, F.; Rudnicki, M.A. Snail regulates MyoD binding-site occupancy to direct enhancer switching and differentiation-specific transcription in myogenesis. *Mol. Cell* **2012**, *47*, 457–468. [CrossRef] [PubMed]
12. Grifone, R.; Demignon, J.; Houbron, C.; Souil, E.; Niro, C.; Seller, M.J.; Hamard, G.; Maire, P. Six1 and Six4 homeoproteins are required for Pax3 and Mrf expression during myogenesis in the mouse embryo. *Development* **2005**, *132*, 2235–2249. [CrossRef] [PubMed]
13. Yajima, H.; Motohashi, N.; Ono, Y.; Sato, S.; Ikeda, K.; Masuda, S.; Yada, E.; Kanesaki, H.; Miyagoe-Suzuki, Y.; Takeda, S.; et al. Six family genes control the proliferation and differentiation of muscle satellite cells. *Exp. Cell Res.* **2010**, *316*, 2932–2944. [CrossRef]
14. Santolini, M.; Sakakibara, I.; Gauthier, M.; Aulinas, F.R.; Takahashi, H.; Sawasaki, T.; Mouly, V.; Concordet, J.-P.; Defossez, P.-A.; Hakim, V.; et al. MyoD reprogramming requires Six1 and Six4 homeoproteins: genome-wide cis-regulatory module analysis. *Nucleic Acids Res.* **2016**, *44*, 8621–8640. [CrossRef] [PubMed]
15. Yajima, H.; Kawakami, K. Low Six4 and Six5 gene dosage improves dystrophic phenotype and prolongs life span of mdx mice. *Dev. Growth Differ.* **2016**, *58*, 546–561. [CrossRef]
16. Guadagnin, E.; Mázala, D.; Chen, Y.-W. STAT3 in Skeletal Muscle Function and Disorders. *Int. J. Mol. Sci.* **2018**, *19*, 2265. [CrossRef] [PubMed]
17. Messina, G.; Biressi, S.A.M.; Monteverde, S.; Magli, A.; Cassano, M.; Perani, L.; Roncaglia, E.; Tagliafico, E.; Starnes, L.; Campbell, C.E.; et al. Nfix Regulates Fetal-Specific Transcription in Developing Skeletal Muscle. *Cell* **2010**, *140*, 554–566. [CrossRef]
18. Rossi, G.; Antonini, S.; Bonfanti, C.; Monteverde, S.; Vezzali, C.; Tajbakhsh, S.; Cossu, G.; Messina, G. Nfix Regulates Temporal Progression of Muscle Regeneration through Modulation of Myostatin Expression. *Cell Rep.* **2016**, *14*, 2238–2249. [CrossRef]
19. Hayashi, S.; Manabe, I.; Suzuki, Y.; Relaix, F.; Oishi, Y. Klf5 regulates muscle differentiation by directly targeting muscle-specific genes in cooperation with MyoD in mice. *eLife* **2016**, *5*, 37798. [CrossRef]
20. Sunadome, K.; Yamamoto, T.; Ebisuya, M.; Kondoh, K.; Sehara-Fujisawa, A.; Nishida, E. ERK5 Regulates Muscle Cell Fusion through Klf Transcription Factors. *Dev. Cell* **2011**, *20*, 192–205. [CrossRef]
21. Potthoff, M.J.; Olson, E.N. MEF2: a central regulator of diverse developmental programs. *Dev.* **2007**, *134*, 4131–4140. [CrossRef] [PubMed]
22. Ling, F.; Kang, B.; Sun, X.H. Id proteins: small molecules, mighty regulators. *Curr. Top Dev. Biol.* **2014**, *110*, 189–216. [PubMed]
23. Christensen, K.L.; Patrick, A.N.; McCoy, E.L.; Ford, H.L. Chapter 5 The Six Family of Homeobox Genes in Development and Cancer. *Advances in Cancer Research* **2008**, *101*, 93–126. [PubMed]
24. Bialkowska, A.; Yang, V.W.; Mallipattu, S.K. Krüppel-like factors in mammalian stem cells and development. *Development* **2017**, *144*, 737–754. [CrossRef]
25. Piper, M.; Gronostajski, R.; Messina, G. Nuclear Factor One X in Development and Disease. *Trends Cell Boil.* **2019**, *29*, 20–30. [CrossRef]
26. Dolfini, D.; Gatta, R.; Mantovani, R. NF-Y and the transcriptional activation of CCAAT promoters. *Crit. Rev. Biochem. Mol. Boil.* **2011**, *47*, 29–49. [CrossRef]
27. Fleming, J.D.; Pavesi, G.; Benatti, P.; Imbriano, C.; Mantovani, R.; Struhl, K. NF-Y coassociates with FOS at promoters, enhancers, repetitive elements, and inactive chromatin regions, and is stereo-positioned with growth-controlling transcription factors. *Genome Res.* **2013**, *23*, 1195–1209. [CrossRef]
28. Sherwood, R.I.; Hashimoto, T.; O'Donnell, C.P.; Lewis, S.; A Barkal, A.; Van Hoff, J.P.; Karun, V.; Jaakkola, T.; Gifford, D.K. Discovery of directional and nondirectional pioneer transcription factors by modeling DNase profile magnitude and shape. *Nat. Biotechnol.* **2014**, *32*, 171–178. [CrossRef]

29. Oldfield, A.; Yang, P.; Conway, A.E.; Cinghu, S.; Freudenberg, J.; Yellaboina, S.; Jothi, R. Histone-fold domain protein NF-Y promotes chromatin accessibility for cell type-specific master transcription factors. *Mol. Cell* **2014**, *55*, 708–722. [CrossRef]
30. Oldfield, A.; Henriques, T.; Kumar, D.; Burkholder, A.B.; Cinghu, S.; Paulet, D.; Bennett, B.D.; Yang, P.; Scruggs, B.S.; Lavender, C.A.; et al. NF-Y controls fidelity of transcription initiation at gene promoters through maintenance of the nucleosome-depleted region. *Nat. Commun.* **2019**, *10*, 3072. [CrossRef]
31. Lu, F.; Liu, Y.; Inoue, A.; Suzuki, T.; Zhao, K.; Zhang, Y. Establishing Chromatin Regulatory Landscape during Mouse Preimplantation Development. *Cell* **2016**, *165*, 1375–1388. [CrossRef] [PubMed]
32. Li, X.Y.; Van Huijsduijnen, R.H.; Mantovani, R.; Benoist, C.; Mathis, D. Intron-exon organization of the NF-Y genes. Tissue-specific splicing modifies an activation domain. *J. Boil. Chem.* **1992**, *267*, 8984–8990.
33. Ceribelli, M.; Benatti, P.; Imbriano, C.; Mantovani, R. NF-YC Complexity Is Generated by Dual Promoters and Alternative Splicing. *J. Biol. Chem.* **2009**, *284*, 34189–34200. [CrossRef] [PubMed]
34. Benatti, P.; Dolfini, D.; Vigano, M.A.; Ravo, M.; Weisz, A.; Imbriano, C. Specific inhibition of NF-Y subunits triggers different cell proliferation defects. *Nucleic Acids Res.* **2011**, *39*, 5356–5368. [CrossRef] [PubMed]
35. Farina, A.; Manni, I.; Fontemaggi, G.; Tiainen, M.; Cenciarelli, C.; Bellorini, M.; Mantovani, R.; Sacchi, A.; Piaggio, G. Down-regulation of cyclin B1 gene transcription in terminally differentiated skeletal muscle cells is associated with loss of functional CCAAT-binding NF-Y complex. *Oncogene* **1999**, *18*, 2818–2827. [CrossRef] [PubMed]
36. Gurtner, A.; Manni, I.; Fuschi, P.; Mantovani, R.; Guadagni, F.; Sacchi, A.; Piaggio, G. Requirement for Down-Regulation of the CCAAT-binding Activity of the NF-Y Transcription Factor during Skeletal Muscle Differentiation. *Mol. Boil. Cell* **2003**, *14*, 2706–2715. [CrossRef] [PubMed]
37. Gurtner, A.; Fuschi, P.; Magi, F.; Colussi, C.; Gaetano, C.; Dobbelstein, M.; Sacchi, A.; Piaggio, G. NF-Y Dependent Epigenetic Modifications Discriminate between Proliferating and Postmitotic Tissue. *PLOS ONE* **2008**, *3*, e2047. [CrossRef]
38. Goeman, F.; Manni, I.; Artuso, S.; Ramachandran, B.; Toietta, G.; Bossi, G.; Rando, G.; Cencioni, C.; Germoni, S.; Straino, S.; et al. Molecular imaging of nuclear factor-Y transcriptional activity maps proliferation sites in live animals. *Mol. Boil. Cell* **2012**, *23*, 1467–1474. [CrossRef]
39. Basile, V.; Baruffaldi, F.; Dolfini, D.; Belluti, S.; Benatti, P.; Ricci, L.; Artusi, V.; Tagliafico, E.; Mantovani, R.; Molinari, S.; et al. NF-YA splice variants have different roles on muscle differentiation. *Biochim. et Biophys. Acta (BBA) - Gene Regul. Mech.* **2016**, *1859*, 627–638. [CrossRef]
40. Mauro, A. Satellite cell of skeletal muscle fibers. *J. Cell Boil.* **1961**, *9*, 493–495. [CrossRef]
41. Maity, S.N. NF-Y (CBF) regulation in specific cell types and mouse models. *Biochim. et Biophys. Acta (BBA) Bioenerg.* **2016**, *1860*, 598–603. [CrossRef] [PubMed]
42. Sakuma, T.; Nishikawa, A.; Kume, S.; Chayama, K.; Yamamoto, T. Multiplex genome engineering in human cells using all-in-one CRISPR/Cas9 vector system. *Sci. Rep.* **2014**, *4*, 5400. [CrossRef] [PubMed]
43. Dolfini, D.; Minuzzo, M.; Pavesi, G.; Mantovani, R. The Short Isoform of NF-YA Belongs to the Embryonic Stem Cell Transcription Factor Circuitry. *STEM CELLS* **2012**, *30*, 2450–2459. [CrossRef] [PubMed]
44. Cullot, G.; Boutin, J.; Toutain, J.; Prat, F.; Pennamen, P.; Rooryck, C.; Teichmann, M.; Rousseau, E.; Lamrissi-Garcia, I.; Guyonnet-Duperat, V.; et al. CRISPR-Cas9 genome editing induces megabase-scale chromosomal truncations. *Nat. Commun.* **2019**, *10*, 1136. [CrossRef] [PubMed]
45. Min, Y.-L.; Bassel-Duby, R.; Olson, E.N. CRISPR Correction of Duchenne Muscular Dystrophy. *Annu. Rev. Med.* **2018**, *70*, 239–255. [CrossRef] [PubMed]
46. Bungartz, G.; Land, H.; Scadden, D.T.; Emerson, S.G. NF-Y is necessary for hematopoietic stem cell proliferation and survival. *Blood* **2012**, *119*, 1380–1389. [CrossRef] [PubMed]
47. Belluti, S.; Semeghini, V.; Basile, V.; Rigillo, G.; Salsi, V.; Genovese, F.; Dolfini, D.; Imbriano, C. An autoregulatory loop controls the expression of the transcription factor NF-Y. *Biochim. et Biophys. Acta (BBA) Bioenerg.* **2018**, *1861*, 509–518. [CrossRef]
48. Moran, J.; Li, Y.; Hill, A.A.; Mounts, W.M.; Miller, C.P. Gene expression changes during mouse skeletal myoblast differentiation revealed by transcriptional profiling. *Physiol. Genom.* **2002**, *10*, 103–111. [CrossRef]
49. Clever, J.L.; Sakai, Y.; Wang, R.A.; Schneider, D.B. Inefficient skeletal muscle repair in inhibitor of differentiation knockout mice suggests a crucial role for BMP signaling during adult muscle regeneration. *Am. J. Physiol. Physiol.* **2010**, *298*, C1087–C1099. [CrossRef]

50. Salizzato, V.; Zanin, S.; Borgo, C.; Lidron, E.; Salvi, M.; Rizzuto, R.; Pallafacchina, G.; Donella-Deana, A. Protein kinase CK2 subunits exert specific and coordinated functions in skeletal muscle differentiation and fusogenic activity. *FASEB J.* **2019**, *33*, 10648–10667. [CrossRef]
51. Millay, D.P.; Gamage, D.G.; Quinn, M.E.; Min, Y.-L.; Mitani, Y.; Bassel-Duby, R.; Olson, E.N. Structure–function analysis of myomaker domains required for myoblast fusion. *Proc. Natl. Acad. Sci.* **2016**, *113*, 2116–2121. [CrossRef] [PubMed]
52. Petrany, M.J.; Millay, D.P. Cell Fusion: Merging Membranes and Making Muscle. *Trends Cell Boil.* **2019**, *29*, 964–973. [CrossRef] [PubMed]
53. Ganassi, M.; Badodi, S.; Quiroga, H.P.O.; Zammit, P.S.; Hinits, Y.; Hughes, S.M. Myogenin promotes myocyte fusion to balance fibre number and size. *Nat. Commun.* **2018**, *9*, 4232. [CrossRef] [PubMed]
54. Pedraza-Alva, G.; Zingg, J.M.; Jost, J.P. AP-1 binds to a putative cAMP response element of the MyoD1 promoter and negatively modulates MyoD1 expression in dividing myoblasts. *J. Boil. Chem.* **1994**, *269*, 6978–6985.
55. Murray-Zmijewski, F.; Lane, D.P.; Bourdon, J.C. p53/p63/p73 isoforms: an orchestra of isoforms to harmonise cell differentiation and response to stress. *Cell Death Differ.* **2006**, *13*, 962–972. [CrossRef]
56. Imbriano, C.; Molinari, S. Alternative Splicing of Transcription Factors Genes in Muscle Physiology and Pathology. *Genes* **2018**, *9*, 107. [CrossRef]
57. Nakka, K.; Ghigna, C.; Gabellini, D.; Dilworth, F.J. Diversification of the muscle proteome through alternative splicing. *Skelet. Muscle* **2018**, *8*, 8. [CrossRef]
58. Zhang, M.; Zhu, B.; Davie, J. Alternative Splicing of MEF2C pre-mRNA Controls Its Activity in Normal Myogenesis and Promotes Tumorigenicity in Rhabdomyosarcoma Cells*. *J. Boil. Chem.* **2014**, *290*, 310–324. [CrossRef]
59. Sebastian, S.; Faralli, H.; Yao, Z.; Rakopoulos, P.; Palii, C.; Cao, Y.; Singh, K.; Liu, Q.-C.; Chu, A.; Aziz, A.; et al. Tissue-specific splicing of a ubiquitously expressed transcription factor is essential for muscle differentiation. *Genome Res.* **2013**, *27*, 1247–1259. [CrossRef]
60. Vogan, K.; Underhill, D.A.; Gros, P. An alternative splicing event in the Pax-3 paired domain identifies the linker region as a key determinant of paired domain DNA-binding activity. *Mol. Cell. Boil.* **1996**, *16*, 6677–6686. [CrossRef]
61. Barber, T.D.; Barber, M.C.; Cloutier, T.E.; Friedman, T.B. PAX3 gene structure, alternative splicing and evolution. *Gene* **1999**, *237*, 311–319. [CrossRef]
62. Pritchard, C.; Grosveld, G.; Hollenbach, A.D. Alternative splicing of Pax3 produces a transcriptionally inactive protein. *Gene* **2003**, *305*, 61–69. [CrossRef]
63. Charytonowicz, E.; Matushansky, I.; Castillo-Martin, M.; Hricik, T.; Cordon-Cardo, C.; Ziman, M. Alternate PAX3 and PAX7 C-terminal isoforms in myogenic differentiation and sarcomagenesis. *Clin. Transl. Oncol.* **2011**, *13*, 194–203. [CrossRef] [PubMed]
64. Vorobyov, E.; Horst, J. Expression of two protein isoforms of PAX7 is controlled by competing cleavage-polyadenylation and splicing. *Gene* **2004**, *342*, 107–112. [CrossRef] [PubMed]
65. LiBetti, D.; Bernardini, A.; Chiaramonte, M.L.; Minuzzo, M.; Gnesutta, N.; Messina, G.; Dolfini, D.; Mantovani, R. NF-YA enters cells through cell penetrating peptides. *Biochim. et Biophys. Acta (BBA) Bioenerg.* **2019**, *1866*, 430–440. [CrossRef] [PubMed]
66. Cappabianca, L.; Farina, A.R.; Di Marcotullio, L.; Infante, P.; De Simone, D.; Sebastiano, M.; Mackay, A. Discovery, characterization and potential roles of a novel NF-YAx splice variant in human neuroblastoma. *J. Exp. Clin. Cancer Res.* **2019**, *38*, 1–25. [CrossRef]
67. Wang, J.; Zhuang, J.; Iyer, S.; Lin, X.; Whitfield, T.W.; Greven, M.C.; Pierce, B.G.; Dong, X.; Kundaje, A.; Cheng, Y.; et al. Sequence features and chromatin structure around the genomic regions bound by 119 human transcription factors. *Genome Res.* **2012**, *22*, 1798–1812. [CrossRef]
68. Zambelli, F.; Pavesi, G. Genome wide features, distribution and correlations of NF-Y binding sites. *Biochim. et Biophys. Acta (BBA) Bioenerg.* **2017**, *1860*, 581–589. [CrossRef]
69. Welch, R.D.; Guo, C.; Sengupta, M.; Carpenter, K.J.; Stephens, N.A.; Arnett, S.A.; Meyers, M.J.; Sparks, L.M.; Smith, S.R.; Zhang, J.; et al. Rev-Erb co-regulates muscle regeneration via tethered interaction with the NF-Y cistrome. *Mol. Metab.* **2017**, *6*, 703–714. [CrossRef]
70. Van Wageningen, S.; Ridder, M.C.B.-D.; Nigten, J.; Nikoloski, G.; Erpelinck-Verschueren, C.A.J.; Löwenberg, B.; De Witte, T.; Tenen, D.G.; Van Der Reijden, B.A.; Jansen, J.H. Gene transactivation without direct DNA binding defines a novel gain-of-function for PML-RARα. *Blood* **2008**, *111*, 1634–1643. [CrossRef]

71. Moeinvaziri, F.; Shahhosseini, M. Epigenetic role of CCAAT box-binding transcription factor NF-Y onIDgene family in human embryonic carcinoma cells. *IUBMB Life* **2015**, *67*, 880–887. [CrossRef] [PubMed]
72. Dolfini, D.; Mantovani, R.; Zambelli, F.; Pavesi, G. A perspective of promoter architecture from the CCAAT box. *Cell Cycle* **2009**, *8*, 4127–4137. [CrossRef] [PubMed]
73. Singh, K.; Dilworth, F.J. Differential modulation of cell cycle progression distinguishes members of the myogenic regulatory factor family of transcription factors. *FEBS J.* **2013**, *280*, 3991–4003. [CrossRef]