UNIVERSITÀ DEGLI STUDI DI MILANO

Scuola di Dottorato in Scienze Biologiche e Molecolari

XXVIII Ciclo

Biochemical and Functional Analysis of NF-YA Cell Penetrating Peptide Properties in Different Cell Systems

Maria Luisa Chiaramonte

PhD Thesis

Scientific tutor: Roberto Mantovani

Academic year: 2014-2015

SSD: BIO/18

Thesis performed at the Department of Biosciences, University of Milan.

Index

PART I ... 2
1. Abstract ... 3
2. State of art ... 5
2.1. Overview on regulation of gene expression ... 5
2.2 Eukaryotic transcriptional machinery and the transcription regulatory elements ... 8
2.3 Gene expression programs control ... 10
2.4 The CCAAT-box ... 12
2.5 Nuclear Factor-Y (NF-Y) ... 14
2.5.1 NF-YB and NF-YC: the histone like dimer ... 15
2.5.2 The different isoforms of NF-Y subunits: NF-YB and NF-YC subunits ... 16
2.5.3 NF-YA isoforms ... 17
2.5.4 NF-YA subunit regulation ... 18
2.5.5 Effects of NF-YA inactivation ... 18
2.5.6 NF-YA: cell cycle regulation and apoptosis ... 20
2.5.7 NF-YA in embryonic stem cells ... 21
2.5.8 NF-Y in human diseases and metabolism ... 23
2.6 Cell penetrating peptides ... 26
2.6.1 Discovery and properties of cell penetrating peptides ... 26
2.6.2 Classification of cell penetrating peptides. ... 28
2.6.3 Cell penetrating peptides applications ... 30
2.6.3.1 Peptides, proteins and enzymes ... 30
2.6.3.2 Nucleic acids and liposomes ... 30
2.6.4 Mechanism of internalization ... 31
3. Aim of the projects and main results ... 34
3.1 Characterization of NF-YA as Cell Penetrating Peptide ... 34
3.2 NF-Y modulates stemness circuitry both in mouse and in human stem cells ... 36
3.3 NF-Y regulates different metabolic pathways ... 37
4. Conclusions and future perspectives ... 38
PART II ... 50
Accepted paper ... 50
PART III ... 51
Manuscript in preparation ... 52
Report ... 72

PART I

1. Abstract

The transcription factor NF-Y is a trimeric complex composed by three subunits NF-YA, NF-YB and NF-YC, which specifically recognizes and binds the CCAAT box sequence, present in almost 30% of eukaryotic promoters. NF-YB and NF-YC contain a histone fold domain, structurally related to H2A and H2B and they dimerize head to tail; NF-YA exists in two different isoforms and it is thought to be the regulatory subunit of the trimer.

NF-Y is a key regulator of cell cycle progression and it regulates different metabolic pathways in proliferating cells. In the last years, it also emerged as important component of stemeness circuitry.

The projects performed during my PhD were focused on the analysis and characterization of different biological and functional properties of NF-Y.

My main project has been inspired by the inspection of NF-YA 3D structure that suggested us the presence, in the evolutionary conserved domain of the protein, of Cell Penetrating Peptide (CPP) features: the presence of two alpha helices and the abundance of arginines residues. Previous works used a functional recombinant GST-TAT-NFYA fusion protein (TAT is a well-known CPP), to transfer NF-YA in hematopoietic stem cells (HSCs), demonstrating its ability to repopulate the bone marrow of immunocompromized animals and, in mESCs, demonstrating its cruciality to maintain cell pluripotency.

In this study, we demonstrated NF-YA capability to enter cells and translocate into the *nuclei* in a TAT-independent manner. Mutational analyses, in the evolutionary conserved domain of the protein, led to the identification of basic aminoacids required for NF-YA cellular uptake. Thanks to the stability of recombinant NF-YA, after transduction, in NIH3T3 and mESCs, we provided evidences about the functionality of the protein *in vivo.* In NIH3T3, NF-YA affects the expression of CCAAT cell cycle progression genes causing cell growth delay and induces the activation of the apoptotic pathway mediated by E2F1; mESCs, instead, after recombinant NF-YA transduction, maintain their pluripotency properties when grown without leukemia inhibitory factor (LIF).

In the two last decades, the use of CPPs has become an important tool for therapeutic applications. In this context, the discovered NF-YA properties, to overpass cell membranes, could be exploited to make recombinant NF-YA a new reagent to study its biological effects in different cell contexts, avoiding limitations of canonical methods of gene delivery, and laying the groundwork for its potential application in clinical trials.

The cruciality of NF-Y in mESCs pluripotency maintenance, and the possibility to use recombinant NF-YA in stem cell systems, pushed us to investigate NF-Y also in human stem cells. We reprogrammed adult human fibroblasts to induced pluripotent stem cells (iPSCs). Preliminary data

suggest that, similarly to mESCs, NF-Y is involved in stemness circuitry regulation of human stem cells. Indeed, we observed a down modulation of master stemness genes after NF-Y subunits silencing in human iPSCs.

The second project was focused on a detailed analysis on the regulation of NF-Y metabolic target genes, altered in cancer cells.
By integrating different studies on NF-Y genomic binding with data of gene expression profiling, coming from different cell types, after NF-Y subunits inactivation, we demonstrated that NF-Y plays an important role in the regulation of genes involved in lipid pathways interacting with master lipogenic TFs, SREBPs. Moreover, we found that NF-Y directly and positively regulates genes involved in glycolysis, and in the biosynthesis of amino acids (serine, glycine and glutamine), purines and polyamines. Differently from its activity in yeast metabolic pathways, in mammalian cells NF-Y enhances energy production through anaerobic pathways; the analysis performed in this study, indeed, highlighted that NF-Y negatively regulates genes involved in TCA cycle.
Most of metabolic genes regulated by NF-Y, identified in this study, are altered in different tumors. This work contributes to unravel the regulatory circuitry of cancer cells and to better understand how the alteration of metabolic pathways affects the progression and growth of cancer cells.

2. State of art

2.1. Overview on regulation of gene expression

The regulation of gene expression is a fine mechanism that orchestrates a plethora of biological processes involving development, cell proliferation, differentiation, aging and apoptosis.

The expression of coding-protein genes is controlled by different steps including chromatin organization, transcription, mRNA processing, transport, translation and protein stability.

In eukaryotic cells, genetic information is encoded in DNA that is packaged in chromatin and confined into the nucleus. The chromatin is a complex of macromolecules consisting of 146 base-pairs of duplex DNA wrapped around a histone octamer consisting of two of each histone proteins, H2A, H2B, H3, H4, forming the basic unit of the chromatin: the nucleosome.

Nucleosomes represent the first level of eukaryotic DNA organization and they are connected by linker DNA of variable length. The fifth histone protein, H1, with other non-histone proteins, plays an important role to wrap nucleosomes into a 30 nm fiber while others fine and not well understood mechanisms, contribute to higher order of compaction.

Although DNA is highly packaged in the nucleus, its accessibility is fundamental for different biological processes: transcription, replication, recombination and DNA repair. DNA accessibility is favored by two classes of enzymes: nucleosome remodelers and histone modifying enzymes.

Chromatin remodeling is promoted by enzymes that, in ATP-dependent process, change the state of the chromatin disrupting nucleosome structure and making free the access to transcription factors or other protein complexes involved in DNA metabolism. (ATP-dependent) The remodeling complexes include two main classes of enzymes: one contains the SWI2/SNF2 family proteins (SWI/SNF and RSC) and the other is represented by ISWI-containing complex (NURF, ACF and CHRAC) (Tsukiyama *et al*., 1995), (Dirscherl and Krebs, 2004).

It has been reported that mammalian SWI/SNF complex plays an important role as tumor suppressor in many human malignancy (Versteege *et al*., 1998); the association with coactivators as well with corepressors suggests that remodeling complex are involved not only in transcriptional activation, but also in transcriptional repression (Eberharter and Becker, 2004).

On the other hand, we have to consider that the amino acids of the histones undergo a remarkable variety of covalent modifications associated to transcriptional activation or repression contributing to establish a condensed or decondensed chromatin state that determines gene expression. Among the different histone modifications, acetylation, phosphorylation, methylation, ubiquitylation, sumoylation, ADP-ribosylation, the first three are the best characterized (Fig. 1).

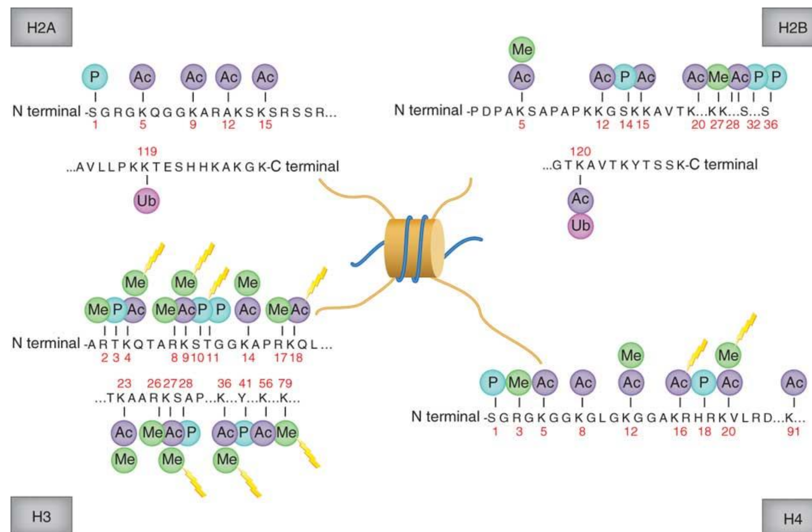


Figure 1. Representation of the main modifications of histone tails. Ac, acetylation; Me, methylation; P, phosphorylation; Ub, ubiquitination. (Manuel Rodríguez-Paredes and Manel Esteller, 2011).

The acetylation of lysine is a dynamic process regulated by the action of two different protein families: histone acetyl-transferases (HATs) and histone deacetylases (HDACs) (Xhemalce B, et *al.,* 2011). The HATs use the cofactor Acetyl CoA to transfer the acetyl group to the ε-amino group of lysine chain side. In this way the positive charge of lysine is neutralized creating a weaker interaction between histone and DNA. Two main classes of HATs have been characterized: type-A and type-B. The type-B HATs are mainly cytoplasmic and they acetylate free histones, before they are included in nucleosome. In particular, type-B HATs acetylate H4 at K5 and K12 and also some residues on H3, and these modifications are important for the deposition of histones, to be removed after histones localization (Parthun MR, 2007).

Type-A HATs are composed by a class of enzymes that can be classified in three different groups:

- GNATs (GCN5- related N-acetyltransferases)
- MYST (including MOZ, Ybf2/Sas3,Sas2 and Tip60)
- P300/CBP (CREB binding protein) (Hodawadekar SC *et al.*, 2007).

These enzymes own the ability to modify sites different from N-terminal histone tails. Indeed, a part from histone tails, other additional sites are present in the globular histone core, whose modification contribute to disrupting the electrostatic interaction that stabilize the nucleosome structure, making chromatin much more accessible to transcriptional activators. HGCN5, for example, acetylates H3K56 (Tjeertes JV et *al*., 2009) affecting histone/DNA interaction, and similarly the knockdown of p300 HAT is associated with the loss of HK56ac, suggesting that also p300 may target this site (Das C. *et al*., 2009).

HDACs play the opposite role to HATs: they remove acetyl group from lysine restoring the positive charge and stabilizing the local chromatin architecture. In agreement with their activity, HDACs, contrary to the HATs, are predominantly classified as transcriptional repressors.

HDACs are divided in four classes, Class I, Class II, Class III and Class IV, and in general, they have a low substrate specificity, so that single enzyme can deacetylate different sites within the histone. Moreover, HDACs were found in multiple complex with other enzymes and other HDAC family members (i.e. HDAC1 was found in a complex with HADC2 in the NuRD, Sin3a and Co-REST complexes) (Yang *et al.,* 2008) making much more difficult to understand which HADC is responsible for a specific effect.

Not only histone acetylation, but also histone phosphorylation reduces the positive charge disrupting electrostatic interaction histone/DNA. Phosphorylation occurs, in a highly dynamic way, on serine, threonine and tyrosine residues of N-terminal histone tails (Xhemalce B *et al.,* 2011), thanks to the action of kinases that transfer a phosphate group from ATP to the hydroxyl group of targeted residues. On the other hand, the action of phosphatases contributes to restore the "*close*" chromatin environment.

Histone methyltransferases are the class of enzymes that transfer a methyl group from S-adenosylmethionine (SAM) to lysine and arginine of histone proteins. While lysine residues can be mono- (me1), di- (me2) or three-(me3) methylated on their amine group, arginine residues are only mono- or di- methylated on their guanidinyl group in a symmetric (me2s) or asymmetric (me2a) configuration (Ng SS *et al.,* 2009).

Specific lysine and arginine residues are methylated on histone H3 and H4, in particular: K4, K9, K27, K36, K79, R2, R8 and R17 are methylated on H3 while H4 methylation markers occur on K20 and R3.

Starting from the first lysine methyltransferase (KMT) discovered, SUV39H1, many others have been identified. Most of them contain a SET domain carrying the catalytic activity and they act on relatively specific target residues. Set1 is the only specific histone methyltransferase for H3K4 in yeast and it works in a complex (COMPASS) containing Bre2, Spp1, Swd1, Swd2, Swd3 and Sdc1 proteins. In mammals six Set1 family members have been described: KMT2F, KMT2G, KMT2A, KMT2B, KMT2C and KMT2D. All of them carry a specific histone methyltransferase activity on H3K4 playing an important role in transcriptional activation (Dou *et al.,* 2005; Steward *et al*., 2006). However, Dot1 is an exception: this enzyme specifically methylates H3K79 in the histone core and it doesn't carry the SET domain. Histone arginine methylation occurs thanks to the protein arginine N-methyltransferase (PRMT) containing two classes of enzimes: type-I enzyme and type-II enzyme.

While type-I generate the Rme1 and the asymmetric dimethylation of arginine (Rme2as), type-II generate Rme1 and Rme2s.

Modifications provide interaction surfaces for protein complexes that control transcription.
When enzymes remove modifications that negatively regulate transcription, it is induced a chromatin remodeling that allow RNA pol II to be recruited and transcription starts.

2.2 Eukaryotic transcriptional machinery and the transcription regulatory elements

Eukaryotic transcription involves a machinery of factors that include: general transcription factors (GTFs), activators and coactivators.
Apart from RNA polymerase II, a series of auxiliary factors including TFIIA, TFIIB, TFIID, TFIIE TFIIF and TFIIH play an important role for an accurate and correct transcriptional process (Figure 2).
The assembly of GTFs on the core promoter to for a preinitiation complex (PIC), directs RNA polymerase II to bind the transcription start site (TSS).

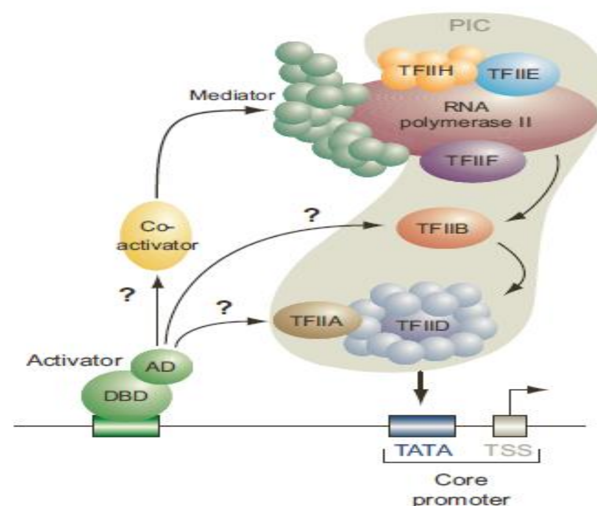


Figure 2. The eukaryotic transcriptional machinery involves: General Transcription Factos (GTFs), activators and coactivators (Maston *et al.,* 2013).

Transcription activity is influenced and stimulated by activators typically characterized by a DNA binding domain, by which they recognize and bind a sequence-specific upstream of the core promoter, and an activation domain required to stimulate transcription (Ptashne M, Gann A., 1997). The activators work promoting the formation of PIC formation and the process subsequent such as

transcription initiation, elongation and reinitiation and their activity can be modulated by other factors important for eukaryotic transcription, the coactivators.

Unlike the activators, coactivators do not have a DNA binding region but they are recruited by protein-protein interactions with one or more activators.

A peculiar feature of activators regards their ability to stimulate transcription synergistically. This is obtained from the cooperation of multiple copies of the same factor or from the cooperation between different factors (Carey *et al*., 1990) (Lin *et al*., 1990). The mechanism underlying transcription synergy is still unclear; it may occur through the simultaneous interaction of activators with recruited GTFs or by the different actions of activators involved to modify chromatin structure and in the regulation of transcription steps, initiation and elongation.

Among transcriptional regulatory elements, the core promoter is the region located at the start of a gene that provides a docking site for the assembly of PIC, define the position of TSS and the direction of transcription. The first described element of core promoter was the TATA box. In addition to TATA box, many other elements were identified in metazoan core promoters such as Initiator element (Inr), Downstream Promoter Element (DPE), Downstream Core Element (DCE), TFIIB-Recognition Element, Motif Ten Element (MTE) and CCAAT box (Lim *et al*., 2004)

Statistical analysis on 10000 human promoters highlighted that the above mentioned regulatory sequences are not universal (e.g. TATA box in present in only 24% of human promoters) (Yang et *al*., 2007) suggesting that other core promoter elements have to be discovered and described.

The region upstream from core promoter, defined proximal prompter element, hold multiple binding site for activator elements.

Roughly 60% of human promoters are near to CpG islands, short DNA sequences rich in C-G di-nucleotides; because of this, the presence of CpG island is used to predict the presence of a gene. CpG islands typically occur near to the TSS of housekeeping genes as well as regulated genes (Bird, 1987) (Gardiner, Frommer, 1987). The CpG islands scattered along the genome are usually methylated, while those near to the promoter are normally unmethylated. Indeed, DNA methylation, and in particular the methylation of CpG islands in gene promoters, leads to transcription silencing by blocking the ability of transcription factors to bind their specific sequences. CpG island methylation, therefore, represents an important mechanism of transcriptional regulation (Jones and Laird, 1999).

Transcription factors regulate gene expression in cooperation with other factors not only by binding core promoter regions where transcription starts but also by binding enhancers.

Enhancers are long distance transcriptional control elements that can be located several hundred kilobase upstream or downstream of a promoter that positively increase gene transcription (Lettice *et*

al., 2003). The control of distant genes is allowed by physical contact that involve DNA bending between enhancers and core promoters (Krivega and Dean, 2012). Enhancers emerged as very important regions for transcription regulation and mutations in their sequences change significantly the binding of transcription factors and therefore the regulation of gene expression of target genes; genetic variations in these regulatory regions are often associated with diseases (Maurano *et al.,* 2012).
Unlike enhancers, the silencers are negative transcription regulatory elements sharing binding sites for negative transcription factors. Transcriptional repression occurs thank to the action of repressors that or prevent the access of GTFs to the promoter, establishing a repressive chromatin structure (Srinivasan and Atchison, 2004) or by blocking the assembly of PIC (Chen and Widom, 2005).
In many cases, gene transcriptional state is repressed and becomes active after specific stimuli.
It has emerged that a class of silencing elements, the Polycomb group Response Elements (PREs), are involved in the process that regulate the promoter switching from repressive to active state. PREs work as silencer or activator elements depending on the type of proteins that they bind (Schmitt S *et al*., 2005).
Finally, among the transcriptional regulatory elements, we find insulators and locus control regions.
While insulators main work is to limit the action of enhances towards a specific set of genes, "protecting" other genes from the transcriptional activity of neighboring genes, locus control regions are a class of regulatory elements involving enhancers, silencers and insulators that by binding TFs, activators, repressors and chromatin modification factors, regulate the transcription of an entire locus or gene cluster (Li *et al.*, 2002).
Therefore, it is the action of different regulatory elements to determine gene expression pattern in a finely tuned dynamic process.

2.3 Gene expression programs control

It is clear that the identity of a cell is defined by genes that are transcribed and they involve housekeeping genes (active in most cells) and cell-specific genes (genes expressed in one cell type).
The gene expression program of a particular cell type is the result of the activity of specific transcription factors that in turn regulate transcription of a selective subset of genes. What emerged from studies on TFs involved in establishing and maintaining cell identity, is that only few key TFs are necessary to determine cell type specific gene expression programs.
In the figure 3 are reported some examples.

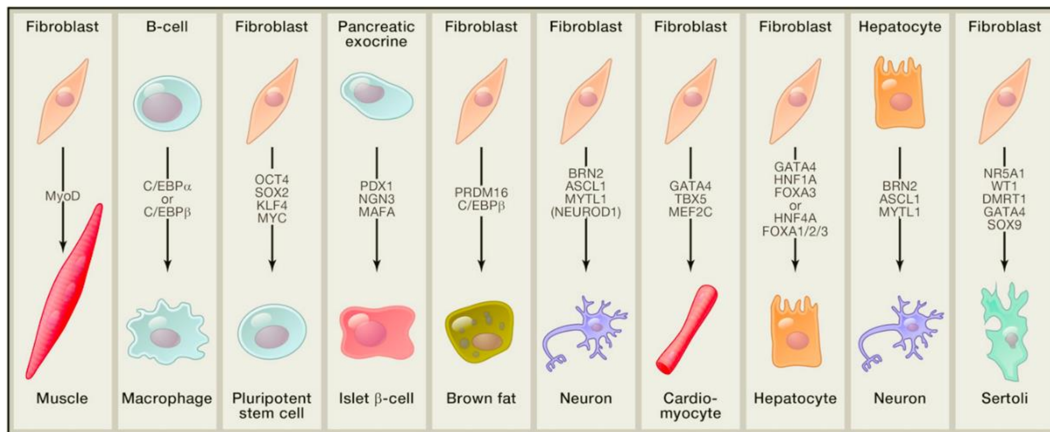


Figure 3. Transcription factors that control specific cell states that are able to reprogramming cell states when ectopically expressed in different cell types (Lee and Young, 2013).

Among the different transcription factors involved in embryonic stem cells (ESCs) circuitry, only few of them, SOX2, OCT4, KLF4 and MYC are involved and are necessary in reprogramming cells to induced pluripotent stem cells (iPSCs), with the some properties of ESCs (Yamanaka, 2012).
The ESCs are characterized by the ability to self-renewal and maintain pluripotency ensuring in this way enough cells during organogenesis, the homeostasis control of tissue, regeneration and repair.
Both human and mouse ESCs are able to maintain their pluripotent state thank to the regulatory feedback circuitry formed by SOX2, OCT4 and NANOG, activating target genes that encode pluripotency and self-renewal mechanisms. Simultaneously, in this regulatory circuitry the three TFs regulate themselves and repress signaling pathways that induce differentiation interacting with Polycomb-mediated repression complex (Young, 2011).
From different studies, what emerged is that TFs that control cell state or that are involved in reprogramming have a broad effect on gene expression program by regulating the initiation or the elongation transcription processes of regulated genes.
The c-Myc transcription factor can stimulate increased elongation from the entire active gene expression program in different cell lines (Lin *et al.*, 2012; Nie *et al.*, 2012).
OCT4 and NANOG, e.g., interact with p300 and Mediator coactivators to promote the formation of open chromatin and recruit transcriptional apparatus; similarly, it happens for other TFs involved in cell trans-differentiation such as MYOD, C/EBPβ, HNF1α, HNF4α, BRN2 and GATA4 (Borgrefe and Yue, 2011).
Among TFs involved in elongation transcription control we can mention hematopoietic TIF1γ transcription factor that controls erythroid cell fate interacting with P-TEFb and regulating transcription elongation at a specific set of target genes (Bai *et al.*, 2010).

Genetic variations and mutations in the transcription factors contribute to the onset of diseases due to the deregulation of gene expression. Some transcription factors control RNA polymerase II pause, release and elongation, therefore, an altered expression or altered function may produce aggressive tumor cells or autoimmune diseases (Lin *et al.,* 2012; Malecki, 2005). Similarly, mutations in coactivators complexes that integrate information from many transcription factors can cause developmental diseases (Ding *et al.,* 2008; Kaufmann *et al*., 2010).
Alteration in specific chromatin regulators contribute to development of cancer and many other diseases (Tsurusaki *et al*., 2012).
New studies have highlighted that mutations in ncRNA species can play important role in cardiovascular diseases (Han *et al.,* 2011; Papageorgiou *et al.,* 2012).
Addition insights into the role of transcription misregulation in human disease, will require improved genome annotation, knowledge of DNA sequences whose alteration contribute to disease and further understanding of the role of cofactors, chromatin regulators and ncRNAs.

2.4 The CCAAT-box

Transcription regulation involves a series of events defined by DNA sequences positioned both in proximity of the genes, promoter, and at distance sites, enhancers, that together with chromatin conformation and modifications define the territories of activity of transcription factors, expressed in a particular cell type.
Promoters and enhancers that activate the transcription of genes by RNA polymerase II are characterized by sequences that are specifically recognized and bound by sequence-specific transcription factors. The CCAAT box is one of the most widespread *cis*-acting elements, identified in human, yeast and plant promoters, playing a very important role in transcriptional regulation.

Figure 4. Sequence Logo of NF-Y consensus optimized with ChIP on ChIP data (Dolfini *et al.,* 2009).

It is present in almost 30% of eukaryotic promoters and it is found both in forward and reverse orientation (Dolfini *et al.*, 2009) between -60 and -100 bps from transcription start sites (TSS).

Although the CCAAT box was long considered only a promoter element important for the recruitment of Pol II, genomic data revealed that CCAAT box is contained in not only the “classic” core promoters but also it correlates with CpG island promoters, intergenic regions and introns (Dolfini *et al.,* 2009). Apart from CCAAT box others well defined *cis*-acting elements common to eukaryotic promoters include TATA box and the GC box. A negative correlation between CCAAT box and TATA box was recently found by bioinformatics analysis: the numbers of promoters containing both CCAAT box and TATA box is significantly lower compared to the promoters containing only one of them.
In higher eukaryotes, the CCAAT box has been found in promoters of many classes of genes, developmentally controlled and tissue specific genes, housekeeping and inducible genes. A class of genes particularly enriched in CCAAT box are genes involved in cell cycle progression and regulation: CCNA, CCNB1, CCB2, CDC2, CDC25A, PLK1 (Mantovani, 1998) (Lin *et al.,* 2014). Figure 5 shows the list of some cell cycle genes regulated by NF-Y

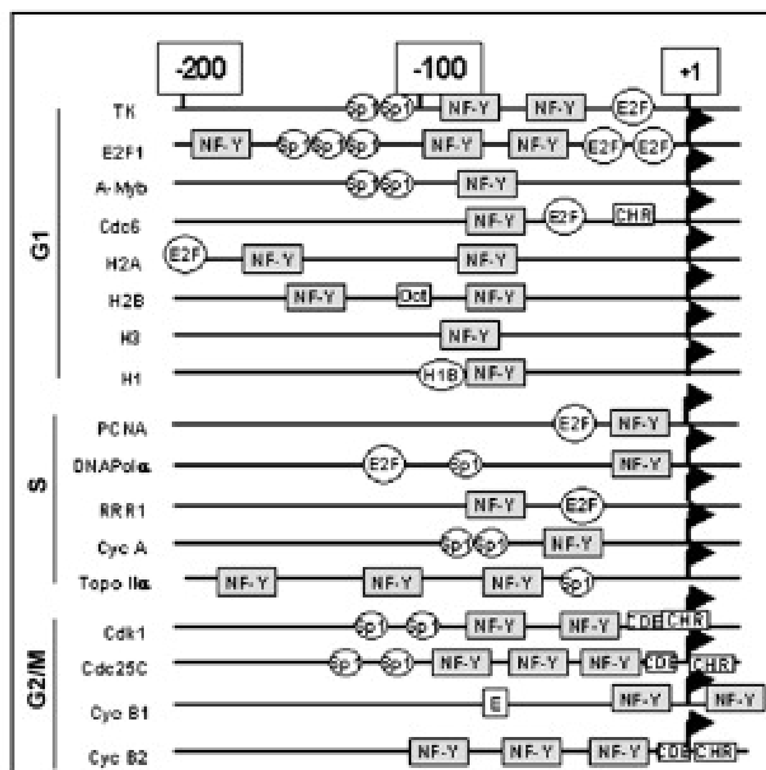


Figure 5. List of some cell-cycle genes regulated by NF-Y. Representation of their promoter regions and transcription factors binding sites are illustrated (Di Agostino, *et al.,* 2006).

These genes lack the TATA box but present multiple CCAAT boxes in their promoters: CCNB2 for instance, has three CCAAT boxes in its proximal promoter region. The major transcription factor that

specifically recognizes and binds the pentanucleotide is NF-Y, which absolutely require all five nucleotides for efficient DNA binding.

2.5 Nuclear Factor-Y (NF-Y)

NF-Y (Nuclear Factor Y), also known as CBF (CCAAT Binding Factor), CP1 or HAP2/3/5 in yeast, is a trimeric transcription factor composed by NF-YA, NF-YB and NF-YC, all indispensable for DNA binding. Each NF-Y subunit contains an evolutionary conserved domain, sufficient for CCAAT binding and subunit interaction, indicated as HAP for the homology with the yeast proteins. 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) (Fig.6).

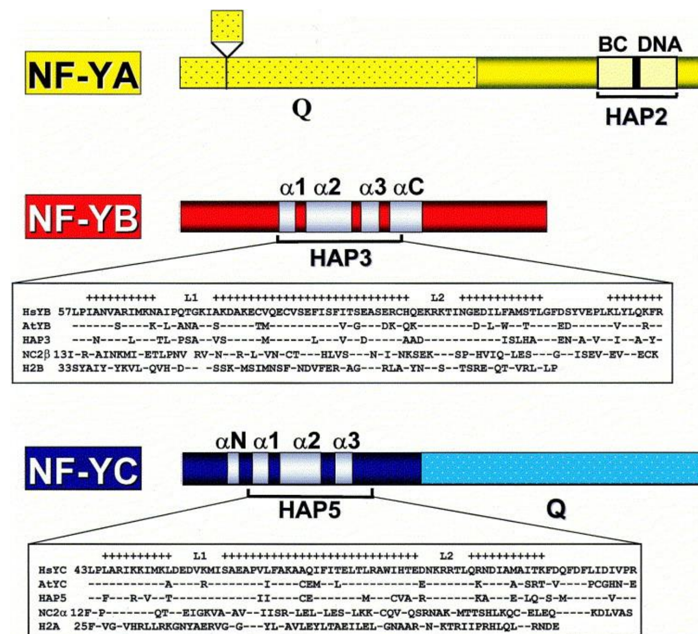


Figure 6. Schematic representation of the NF-Y genes. The yeast homology domains are indicated by brackets. White boxes in NF-YB and NF-YC indicate the position of the four α-helices of the histone fold domains. Below the NF-YB and NF-YB schemes are shown the sequences of the conserved domains of the human NF-YC, *A. thaliana* NF-YCb, *S. cerevisiae* HAP5, NC2a and *Xenopus* H2A (lower panel) and of the corresponding sequences of NF-YB, *A. thaliana* NF-YBb, HAP3, human NC2b and *Xenopus* H2B (upper panel). The + symbol indicates the position of the α-helices in the histone folds; the L1 and L2 are the loop regions. (Mantovani, 1999).

The N-terminal domain of NF-YA and the C-terminal domain of NF-YC are characterized by a large glutamine rich portion (Q-rich) important for transactivation function (Mantovani, 1999).

NF-YB and NF-YC subunits have a histone-like domain characterized by three alpha helices separated by short strand regions that associate forming a dimer important for the interaction with NF-YA that dictates the association to DNA. The resulting functional trimer binds to the CCAAT containing double stranded DNA with high affinity and specificity, inducing a global DNA bending of 80°. This is a peculiarity of NF-Y that can reflect functional consequences for the spatial

organization of its target promoters. Indeed, the transcriptional regulation of NF-Y target genes is not only due to NF-Y CCAAT binding. Different studies highlighted functional and physical interaction of NF-Y with other transcription factors, both *in vitro* and *in vivo* (Zwicker *et al.,* 1995).

NF-Y was identified as a factor associated to the conserved promoter element Y-box of the major histocompatibility complex class II genes (Dorn *et al.*, 1987). Since, many other eukaryotic promoters (type I collagen, albumin, globin, thombospondin 1 and human thymidine kinase gene pomoters) were discovered to be activated by NF-Y (Chodosh *et al.,* 1988, Delvoye *et al*., 1993, Framson and Bornstein, 1993; Mantovani *et al.,* 1992; Raymondjean *et al.,* 1988).

Now it is known that the human genome harbors sites for NF-Y, scattered along the whole genome, with most of them residing in annotated gene bodies and intergenic region (45%); localization within promoters can be attested at nearly 10% of NF-Y sites. Within promoters, the majority of NF-Y sites falls nearby the TSS. For the remnant part of the genome, one third of NF-Y sites is associated to mRNA expression area, regions of Pol II activity, enhancers and transcriptional units whose RNAs have not been annotated yet (Ceribelli *et al.,* 2008).

2.5.1 NF-YB and NF-YC: the histone like dimer

NF-YB and NF-YC core region sequences present a high degree of homology with the H2A and H2B histones, therefore, they belong to the Histone Fold Domain protein family.

Mutational analysis in the histone-fold domain of NF-YB and NF-YC highlighted that the integrity of these motives is fundamental for subunits dimerization and DNA binding (Maity and Crombrugghe, 1998).

As demonstrated by human NF-YB/NF-YC structure, the two proteins interact with each other head to tail, forming a dimer that is structurally similar to H2A/H2B (Romier *et al.,* 2003) (Fig.7).

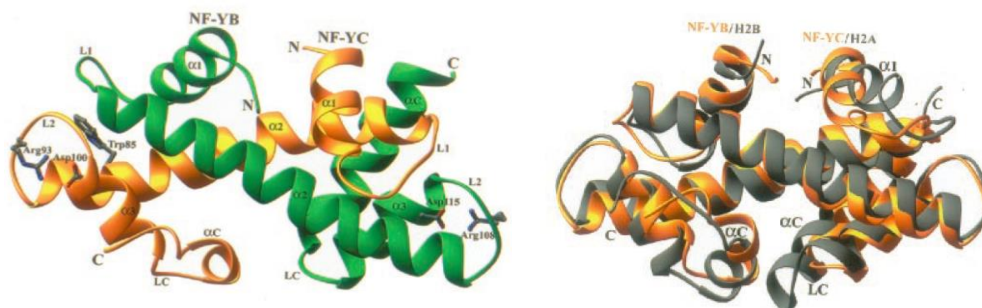


Figure 7. Comparison of NF-YC/NF-YB and H2A/H2B dimers. The left panel show the ribbon representation of NF-YC/NF-YB dimer; the right panel represent the superimposition of NF-YC/NF-YB (orange) and H2A/H2B (gray) histone pairs (Romier *et al.,* 2003).

The conserved histone fold motif of NF-YB and NF-YC is important for the interaction of NF-Y with other histone like proteins such as different TAF(II)s subunits of TFIID. In particular, while TAF11, TAF12 and TAF13-TAF11 complex interact with the histone fold domain of NF-Y dimer, TAF6 and TAF9-TAF6 complex interact only with NF-Y trimer (Frontini *et al.,* 2002).
Compared to the general histone fold domain (HFD) structure, characterized by three alpha helices linked by two loops, NF-YB and NF-YC core regions carry an additional alpha helix (α-C) at the C-terminal region of their HFDs that is important for the interaction with other transcription factors.
NF-YC α-C, indeed, is crucial to establish crosstalk and interactions between NF-Y and other transcription factors, for instance c-MYC or p53 (Izumi *et al.,* 2001).
The formation of NF-YB/NF-YC dimer, offer the correct docking site for the association of NF-YA subunit. In this way, in addition to HFD, NF-YA DNA contact further stabilizes DNA binding of the composed trimer and adds a layer of high specificity to the trimer.

2.5.2 The different isoforms of NF-Y subunits: NF-YB and NF-YC subunits

As it is well known, most of the mammalian genes undergo alternative splicing so that the same gene generates different protein isoforms; NF-Y subunits undergo the same process.
While NF-YB gene, in mammals, generates a single protein product of 32 KDa (Fig.8), the transcriptional regulation of the NF-YC locus is much more complex. NF-YC protein isoforms, indeed, originate not only from events of alternative splicing but also by the activity of alternative promoters, P1 and P2. The P1 promoter has substantial housekeep activity as demonstrated by transient transfection, ChIP experiments and RT-PCR analysis; the promoter P2 is activated following DNA damage in a p53 dependent pathway.

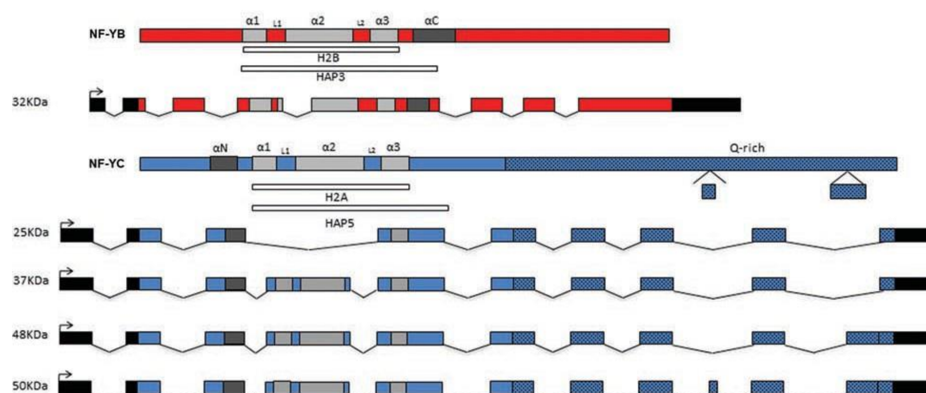


Figure 8. The structure of NF-YB and NF-YC genes. Functional domains and differential splicing are reported (Dolfini *et al.*, 2012).

The Q-rich transactivation domain of NF-YC is the region affected by alternative splicing. The splicing involves exon 8 and exon 9, generating three different protein products of 37, 48 and 50 KDa (Fig.8).

RT-PCR analysis performed in 13 different human cell lines highlighted that the three different splicing isoforms are transcribed in the cells but the protein expression level is different: the relative abundance of each protein isoform is cell dependent and, in particular, the 37 and 50 KDa isoforms are mutually exclusive. These data indicate that strong post-transcriptional mechanism affecting mRNA stability and translation (Ceribelli *et al.,* 2009).

Moreovere, these findings strongly correlate NF-YC to NF-YA subunit, described in the following section.

2.5.3 NF-YA isoforms

NF-YA subunit, differently from NF-YB and NF-YC, does not belong to the histone fold domain family.

NF-YA is composed of a Q-rich trans-activation domain in its N-terminal domain of 180 amino acids rich in hydrophobic residues and an evolutionary conserved domain in the C-terminal domain. This core region is composed of two alpha helices each specifying a function: A1 mediates the interaction with NF-YB and NF-YC and A2 is involved in DNA binding and CCAAT recognition. The two alpha helices are connected by a linker and they are rich in basic amino acids (lysines and arginines) with the C-terminal domain of A2 helix holding two serines –S292 and S298-, target of CDK2 in a cell cycle dependent manner (Yun *et al.*, 2003; Chae *et al.*, 2004; Chan *et al.,* 2010).

NF-YA gene undergoes alternative splicing involving exon 3 in the Q-rich domain. The exon skipping event generates two different protein isoforms: the long (NF-YAl) and the short (NF-YAs) lacking the 28 amino acids codified by third exon (Fig. 9).

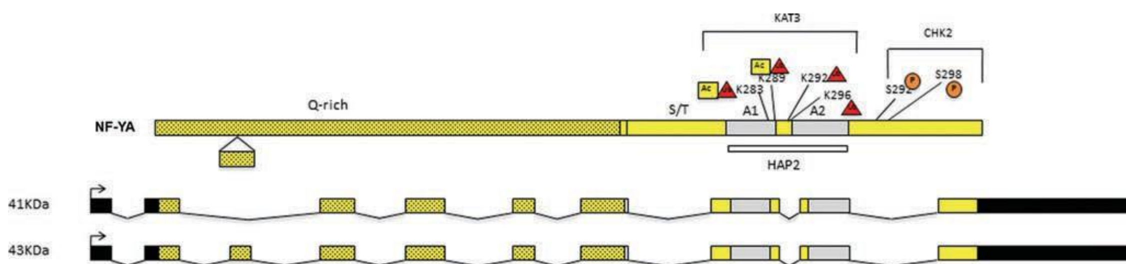


Figure 9. Structure of NF-YA gene with its functional domains, alternative splicing and post-translational modifications. (Dolfini *et al*., 2012)

Different cell lines preferentially express the long or the short isoform of NF-YA, but never the expression of one isoform completely abolish the expression of the other. It was found, as above

mentioned, a tight correlation between NF-YA and NF-YC subunits expression: 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 (Ceribelli *et al.*, 2009). These data suggest that different cell contests harbor different NF-Y complexes, but it is not still clear how and whether they activate different set of genes.

2.5.4 NF-YA subunit regulation

NF-YA is the limiting subunit of the trimer. While the amount of NF-YB and NF-YC subunits are relatively constant, it was observed that NF-YA expression level fluctuates during cell cycle highlighting its role of regulatory subunit of the trimer. NF-YA accumulation control occurs at post transcriptional level. Differently from NF-YB and NF-YC subunits, NF-YA protein is regulated by the ubiquitin/proteasome pathway. The short half-life of the protein, estimated about two hours in proliferating cells, increases after proteasome inhibition and mutational analysis identified four lysines (K283, K289, K292 and K296) in the C-terminal domain of the protein as important targets of ubiquitylation. Moreover, it was observed that two of the lysines ubiquitylated are also acetylated by the p300 histone acetyl-transferase (Manni *et al.,* 2008). Therefore, the competition between acetylation and ubiquitylation on the same lysines residues emerges as mechanism to regulate NF-YA stability.

The differential expression of NF-YA, that determines the amount of active trimer available for CCAAT-binding activity, was observed in different tissues and cell lines. The expression level of NF-YA fluctuates during cell cycle and it is high in G1 and S phase while, it decreases in G2/M (Bolognese *et al.,* 1999). The amount of NF-YA protein can be modulated by specific *stimuli* and cell treatments: while deprivation of serum induces a reduction of NF-YA subunit expression level in human IMR-90 fibroblasts (Chang *et al.,* 1994), MMS treatment increases it (Lee *et al.,* 2004). 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 (Farina *et al*., 1999; Gurtner *et al.,* 2003). All these data show that the level of NF-YA governs the function of NF-Y trimer.

2.5.5 Effects of NF-YA inactivation

The NF-Y complex is a fundamental player in the regulation of cell proliferation, supporting the basal transcription of cell cycle genes.

The function of NF-Y in gene transcription and cell cycle regulation took advantage by loss of function approaches, such as the expression of dominant negative NF-YA mutants or by conditional inactivation of NF-YA.

The stable expression of dominant negative NF-YA in mouse fibroblast cells induces a delay in S phase progression (Hu and Maity, 2000); a truncated NF-YA mutant, lacking the C-terminal transcription activation domain, leads cells to cycle block in G2/M phase without apoptosis (Hu et *al*., 2006) while conditional inactivation of NF-YA alleles in mouse embryonic fibroblasts (MEFs) completely blocks cells in S phase and then induces apoptosis (Bhattacharya *et al*., 2003).

The importance of NF-Y is not just limited to the regulation of genes involved in cell cycle progression, but it extends to the control of genes involved in different metabolic pathways. It was demonstrated that the conditional inactivation of NF-YA in liver of newborn mice caused a severe liver injury with progressive degeneration of hepatocytes and induction of an aberrant endoplasmic reticulum stress pathway, demonstrating that NF-YA, and therefore NF-Y, is a key transcription factor controlling ER function and metabolic processes in mature hepatocytes (Luo *et al*., 2011).

NF-Y was also identified as important factor for preadipocytes and as an adipogenic factor that regulates leptin gene expression. *In vitro,* NF-Y is expressed in adipocytes and its levels decrease during adipogenesis. NF-Y knockdown reduces adipocyte differentiation suggesting that it plays an important role prior to differentiation or that have functions early in the process before the point of adipocytes development when it is no longer expressed.

RNAi mediated knockout of NF-Y in preadipocytes and adipocyte-specific knockout of NF-YA results in a decreased expression of leptin gene and of other fat specific genes; it was observed that NF-Y knockout mice develop a moderately severe lipodystrophy, remediable with leptin therapy (Yi-Hsueh Lu *et al.,* 2015). Further studies are in any case required to elucidate the mechanisms by which NF-Y controls leptin gene expression, to identify mechanisms that modulate adipocyte function and leptin production in response to changes in adipocyte lipid context.

A recent study demonstrated that, although NF-Y often loses its activity in differentiated nonproliferative cells, it is active in adult neurons. The conditional deletion of NF-YA subunit in post mitotic neurons induce progressive neurodegeneration accompanied by a failure of incorporation of Ubiquitin and p62 in filamentous inclusions that co-accumulated with insoluble membrane around endoplasmic reticulum (ER). Neurodegeneration is characterized by an aberrant increase of smooth ER in the perinuclear region without the activation of ER stress response. These data, therefore, place NF-Y as an important regulator of ER organization in adult neurons and its ablation causes neuropathology with abnormal accumulation of ubiquitin/p62 (Yamanaka *et al.,* 2014).

2.5.6 NF-YA: cell cycle regulation and apoptosis

Cell proliferation and apoptosis are two processes finely related, as two sides of the same coin. Most of important effectors of apoptosis are expressed in cells and under appropriate stimuli can alter the transcriptome and induce the expression of proapoptotic proteins.

Different genes, such as *c-myc, E2F1* and cyclins, were described as genes involved in both cell proliferation, cell-cycle arrest and cell death. The choice and the activation of different pathways depend on cell type, cellular environment and genetic background (Vermeulen *et al.,* 2003). Although NF-Y is known to play a central role in cell proliferation by controlling the expression of important genes involved in cell cycle progression such as *cyclin A, cyclin B1, cyclin B2, cdc25A, cdc25C and cdk1,* different reports described NF-Y as transcription factor having role also in apoptosis. Gurtner et *al*. demonstrated that NF-Y overexpression, in particular NF-YA overexpression in mouse embryonic fibroblasts (MEFs) and human cells inhibits cell proliferation and induces apoptosis by activating E2F1-dependent p53 apoptotic pathway. Apoptosis is not induced in E2F1$^{-/-}$ or p53 ablated cells and an increasing in cell proliferation is observed after NF-Y overexpression in mutant p53 cells (Gurtner *et al.,* 2010).

It is reasonable to assume that it is the amount of NF-YA, that is the limiting subunit of the trimer, to be decisive for cell fate, because NF-YA depletion also leads cells to apoptotic response (Benatti *et al*., 2011).

Gene expression profilings, ChIP-on-ChIP and bioinformatics analysis identified pro-survival, pro-growth and anti-apoptotic genes (BCl-xl and BCl2) to be under the direct control of NF-Y (Benatti *et al.,* 2006) but also apoptotic genes are under NF-Y transcription control: Bim, a strong pro-apoptotic genes, for example, is transcriptionally activated by NF-Y in sympathetic neurons (Hughes et *al.,* 2011). Bim is a critical mediator of apoptosis in different cell types and in sympathetic neurons is required for NGF withdrawal- induced death.

The inverted CCAAT box present in *bim* promoter is bound by NF-Y and its activity is essential to activate the reporter gene *bim*-Luciferase in absence of NGF. Moreover, it has been shown that NF-Y and FOXO3a cooperate and interact with coactivators CBP and p300 for bim promoter activation after withdrawal of NGF (Hughes *et al.,* 2011).

On the other hand, it was demonstrated that NF-YA regulates the transcription of Polo Kinase1 (PLK1), an important kinase of mitotic phase (Hamanaka *et al.,* 1995). In this paper, the authors propose a link between NF-Y, CDK2 a p21. In unstressed cells, NF-YA is associated with CDK2 to the CCAAT box of PLK1 promoter to regulate cell cycle progression and cell division. When cells are treated with DNA damage agents, p53 is activated, the expression levels of p21 increased and p21

associates with NF-YA replacing CDK2 on PLK1 promoter, that is repressed, preventing mitotic catastrophe and inducing mitotic death (Lin *et al.,* 2014). The molecule model is shown in figure 10.

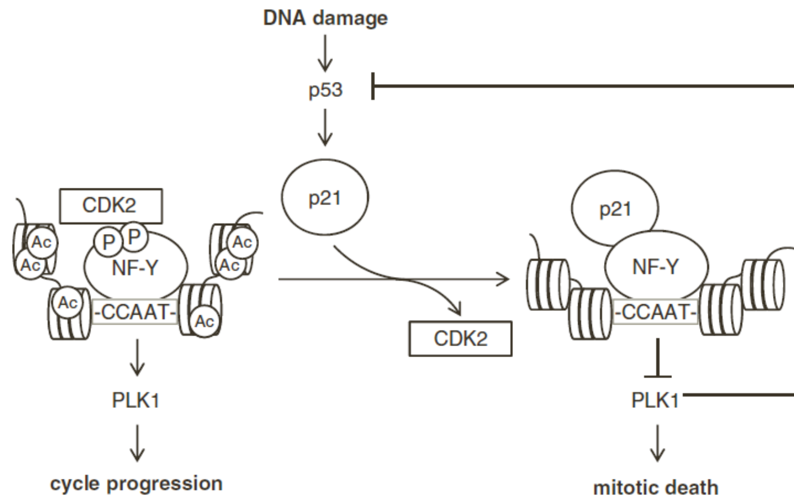


Figure 10. Model in which the dynamic interplay among CDK2, p21 and NF-YA in the promoter fine tunes *PLK1* expression. In unstressed cells, CDK2 associates with NF-YA bound on the CCAAT box in the PLK1 promoter to allow cell cycle progression. DNA damage activates p53 and induces p21 to replace CDK2 in binding to NF-YA, which causes PLK1 repression and prevents mitotic cell death. (Lin *et al.,* 2014).

In agreement with the role of NF-Y to drive cellular growth, NF-YC subunit was recently classified as driver oncogene (Tong *et al.,* 2015). However, the role of NF-YA in apoptosis remain ambiguous because both overexpression and depletion of NF-YA subunit induce cells to apoptosis and both pro-apoptotic and anti-apoptotic genes are modulated by NF-Y. A better understanding about the implication of NF-Y in apoptotic processes will come from genomic analysis of NF-Y sites, positive and negative, and the matching with the regulome of other TFs involved in apoptosis induction.

2.5.7 NF-YA in embryonic stem cells

Consistent with its role in cell cycle regulation, NF-Y is required for embryonic stem cells (ESCs) and hematopoietic stem cells (HSCs) proliferation.

It was found that while NF-YA heterozygous mice are normal and fertile, NF-YA null mice die prior 8.5 dpc, suggesting that NF-Y plays a crucial role in early mouse embryonic development (Bhattacharya *et al.,* 2003).

The involvement of NF-Y in early tissue development was initially suggested by Stathopoulos t *al.* in *Drosophila,* where he demonstrated that the homologous if NF-Y regulates dorsal-ventral

patterning (Stathopoulos *et al.*, 2002). Emerson team, later, found that NF-Y acts as master regulator of HSCs self-renewal and differentiation (Zhu *et al.*, 2005).

Of the two NF-YA isoforms, the short one is considered the stemness isoform as demonstrated by different experiments concerning both adult and embryonic stem cells (ESCs).

The overexpression through retroviral infection of NF-YAs in HSCs induces the expression of HOXB4 and HOX4 paralogs (HOXC4, HOXD4 and HOXA4) as well as Hes-1, LEF-1, Notch1, p27 and telomerase and it address HSCs toward self-renewal rather than differentiation. This aspect is highlighted by the prominent ability of HSCs, overexpressing NF-YAs, to repopulate the bone marrow of immunocompromized animals (Zhu *et al.,* 2005).

Similar results were confirmed by Domashenko, using TAT-NF-YAs protein as alternative to using retroviral expression of NF-YA in HSCs. TAT-mediated transduction of NF-YAs peptide induces the *ex vivo* proliferation and increased engraftment of human hematopoietic progenitor cells (Domashenko *et al.,* 2010).

Human and mouse ESCs transcriptomes analysis identified the CCAAT box enriched in important regulatory regions of embryonic stem specific genes (Grskovic *et al.,* 2007).

Between the two isoforms, NF-YAs predominates in mouse ESCs and plays a global role in maintenance of stemness. A switch between the short and the long isoform was observed upon differentiation of mESCs to embryo bodies (EBs): NF-YAl isoform is present at very low levels in mESCs, while NF-YAs is the major isoform expressed before and decrease to undetectable levels after differentiation (Dolfini *et al.,* 2012). Further experiments demonstrate that the use of dominant negative of NF-YAs negatively affects the expression of important stemness genes in mESCs: *Klf4, Klf5, Arid1a, Fgf4, Sall4* and *Jarid2.*

Moreover, the transduction of TAT-fusion NF-YAs activates stemness genes containing CCAAT box accelerating growth of mESCs when they are cultured in normal conditions and maintaining cells in a pluripotent state when they are growing in absence of LIF (Dolfini *et al.*, 2012).

A strong correlation between NF-Y and NANOG emerged from bioinformatics analysis highlighting 30%-50% of NANOG peaks containing NF-Y sites. In mESCs, the *Nanog* promoter is not a direct target of NF-Y, but it is controlled by SOX2 and OCT4, which are instead directly regulated by NF-Y; in this way, indirectly, NF-Y regulates mRNA and protein expression levels of NANOG (Dolfini *et al.,* 2012).

After NF-Y depletion, the ESCs gene expression profile is similar to that obtained after inactivation of pluripotent master genes *Oct4, Sox2, Nanog.* The silencing of one of NF-Y subunits leads to self-renewal defects and start of differentiation toward the trophoblast lineage.

Even if NF-Y function has been widely studied in relation to its property to bind CCAAT box in proximity to transcription starting sites (TSS), genome wide analysis revealed that NF-Y sites are particularly enriched not only in promoters but also in active enhancer regions of ES specific genes. It has emerged that at least half of NF-Y binding sites are located at 500 bps from TSS in mESCs where NF-Y colocalizes with *OCT4, Nanog* and *Prdm14.* As pioneer TF, NF-Y promotes, in these sites, chromatin accessibility for other master TFs *(*Oldfield *et al.,* 2014).
It is very important to underline that genes regulated by NF-Y from distal binding site are enriched in pathways involved in embryonic development. In agreement with the role of NF-Y in ESCs maintenance, NF-Y binding to distal sites is lost when cells leave pluripotent state and are committed to differentiation (Oldfield *et al.,* 2014).
All these findings define NF-Y as an essential factor for ESCs pluripotent state regulation.

2.5.8 NF-Y in human diseases and metabolism

From different studies, NF-Y emerged as a transcription factor involved in the regulation of many genes involved in different types of cancers (Li *et al.,* 1992) and its activity plays an important role both in the regulation of genes involved in human diseases and in the regulation of cell metabolism.
NF-Y was not found to be muted or altered in cancers but, different expression profiling experiments show that the activation of some CCAAT genes is important to change the normal transcriptome profile of the cells, inducing them to transformation.
Data on gene expression profiles coming from thyroid (Salvatore *et al*., 2007), breast (Scafoglio *et al*., 2006), colon (Jürchott *et al*., 2010), prostate carcinoma (Blum *et al.,* 2009; Calvo *et al.,* 2010) and leukemias (Forsberg et al., 2010), all indicate that promoters of cancer genes, particularly those of most aggressive tumors, are enriched in NF-Y sites (Thomassen *et al*., 2008). The treatment of cells with antitumor drugs, led to repression of NF-Y driver cancer genes (Blum *et al.,* 2007; Yamanaka *et al.,* 2009). Analyses extended to more than 6000 expression profiles of tumors found NF-Y and E2F sites as the only enriched motifs in genes overexpressed in human cancers (Rhodes *et al*., 2005; Sinha *et al*., 2008; Goodarzi *et al.,* 2009).
Therefore, a link between NF-Y and E2F emerged as an important point in the regulation of cancer genes and, a detailed analysis, has to be undertaken in order to better understand mechanisms governing these processes.
The correlation between NF-Y and human disorders emerged from different studies. Several reports demonstrated the involvement of NF-Y in polyglutamine diseases, among these: Huntington disease-HD- (Yamanaka *et al*., 2008) and Spincerebellar Ataxia Type 17 -SCA17- (Landsberger and Wolffe, 1995).

HD is a neurodegenerative disorder due to the accumulation of a mutated Huntingtin protein (HTT), with an extended polyglutamine tract, with other TFs containing Gln-rich domain, in post mitotic neurons of cortex and striatum. The accumulation of HTT with NF-YA and NF-YC subunits subtract NF-Y from its activity of transcription factor. Heat shock genes –HSP40 and HSP70- coding for important chaperonins that provide the correct folding of other proteins preventing aggregation, and are controlled by NF-Y. Thus, a negative loop is trigged, enhancing cell toxicity and neurological disorders (Yamanka *et al*., 2008).

SCA17, instead, is due to the expansion of Gln-rich domain of TBP gene a TF binding the TATA box. Although TATA boxes are mainly present in CCAAT less promoters, NF-Y and TBP interact for the transcription of chaperonins (Bellorini *et al.,* 1997), therefore, similar to HD, poly-Gln TBP causes NF-YA aggregation and thus neuron degeneration (Huang *et al.,* 2011).

Lagor *et al.* study, links NF-Y to diabetes disease. Insulin induces the production of hepatic 3-hydroxy-3-methylglutaryl-CoA reductase – HMGR – CCAAT gene that is normally bound by NF-Y but not in diabetic animals (Lagor *et al.,* 2005). However, it is not still clear whether the loss of binding of NF-Y to the CCAAT box is the cause of the illness or whether it is a secondary effect.

Different studies report that SNPs in the NF-Y target sequence can affect the expression of genes involved in different pathways and related to human disorders. Among genes owning SNPs in CCAAT box, were identified genes coding for enzymes involved in the control of telomeres length, detoxification (von Richter *et al*., 2004) and in genes related to Hereditary persistence

of fetal hemoglobin – HPFH – and β-thalassemia. In HPFH, CCAAT genes coding for fetal globin continue to be expressed after the birth, in adult life. The proposed model about this regulation is that mutations around CCAAT box prevent the binding of a repressor TF for fetal globin genes, maintaining gene expression (Chassanidis *et al.,* 2009; Liberati *et al*., 2001).

Finally, a mutation in the CCAAT box of β-globin gene promoter was found to be associated to β-thalassemia intermedia, characterized by a reduced transcription of this mRNA (Chen *et al*., 2007).

NF-Y was also identified as an important regulator of genes involved in different metabolic pathways. Two important cholesterogenic genes are regulated by NF-Y: farnesyl diphosphate (FPP) synthase and 3-hydroxy-3-methylglutaryl-coenzyme A (HMG-CoA). An inverted CCAAT box was identified in the promoter of FPP synthase, and its mutation leads to a complete transcriptional inhibition while the transfection of the NF-YA dominant negative, affects both FPP and HMG-CoA synthase transcription (Jackson *et al.,* 1995). In addition, the 7-dehydrocholesterol reductase (Dhcr7), the terminal enzyme in the pathway of cholesterol biosynthesis, contains, on its promoter, binding sites for NF-Y and Sp1 and mutational analysis highlighted that they play a cooperative role in the transcription activation sterol-mediated (Jai-Hyun Kim *et al*., 2001). Moreover, NF-Y interacts with

a potent lipogenic transcription factor, SREBP-1, in particular with the hepatic SREBP-1c, positively enhancing the transcription of fatty acid synthase (FAS) (Teran-Garcia *et al.,* 2002).

In agreement with its ability to work as transcription activator and repressor, NF-Y negatively regulates the transcription of pyruvate carboxylase (PC), an enzyme involved in different metabolic pathways such as that involved in the secretion of insulin by pancreatic β-cells under glucose stimulation. Indeed, it was demonstrated that, the mutation of CCAAT box located in the promoter of PC, increased the transcription of a reporter gene transfected in insulin secreting cells (Sunyakumthorn *et al*., 2005).

In the last years, a tight correlation emerged between cancer and metabolism: different metabolic pathways are usually altered tumors.

A detailed analysis about the link between NF-Y, metabolism and cancer, is reported in the paper attached in the second part of this thesis.

2.6 Cell Penetrating Peptides

The integrity of cell membrane is fundamental to protect cells, to favorite their communication and to maintain tissue homeostasis. Cell membrane organization, indeed, prevents the exit and the uptake of small molecules, proteins and genetic material in unspecific way. However, the presence of cells and tissues barriers, such as the blood-brain barrier (in the case of neurodegenerative diseases), represent an important obstacle for drug delivery to pelt their targets.

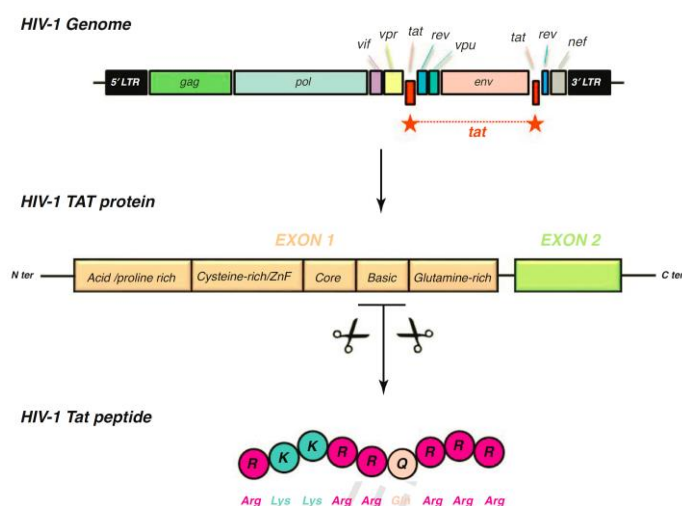
An efficient and relatively new strategy to deliver pharmaceutical molecules into the cells and to increase their tissue distribution emerged in the last years and it is based on the use of cell penetrating peptides (CPPs).

CPPs are relatively short peptide (less of 30 amino acids) able to ubiquitously over cross the plasma membrane in a receptor free manner. In this way, therapeutic molecules, antibodies, peptides, nanoparticles, nucleic acids, conjugated to a cell penetrating peptide, can be efficiently distributed, increasing their concentration, in difficult to access target tissues. Therefore, CPPs represent a revolutionary tool, to treat different human diseases, in a relatively low toxic and non-invasive way.

2.6.1 Discovery and properties of cell penetrating peptides

The first cell-penetrating peptides discovered, Tat-peptide and Penetratin, are still today among the most studied (Frankel and Pabo, 1988).

After the characterization and the description of HIV to infect cells, TAT protein was identified as protein able to over cross cell membranes. TAT is an RNA-binding protein of 14 KDa, that recognizes the TAR sequence (transactivator response element) of HIV genome (Feng *et al.,* 1988)



,

Figure 10. Tat peptide derived from TAT protein codified by HIV-1 genome (Rizzuti *et al.,* 2014).

it stimulates HIV-1 gene expression enhancing the processivity of RNA pol II (Ott *et al.,* 2004). TAT protein is composed of a transcriptional activation domain containing a cysteine rich region and of a hydrophobic motif rich in arginines that specifies the binding of the protein to RNA (Fig. 10).
Generating deleted mutants in the N-terminal and C-terminal domain of TAT protein, Park et *al.* arrived to define the minimum Tat peptide sequence (TAT 49-57) needed to translocate cells without losing efficiency (Park *et al.,* 2002). This region, required for translocation, is rich in basic amino acids, lysines and arginines, and it is important for DNA binding.
Contemporary to the discovery of Tat peptide, *Drosophila* Antennapedia homeobox protein (pAntp) was identified as CPP able to penetrate differentiated neurons and to accumulate into the *nuclei* after cell translocation (Joliot *et al.,* 1991). Generally speaking, homeoproteins are transcription factors that bind specific DNA sequences through a 60 amino acid sequence composed by three α-helices and a β-turn between helix 2 and 3 (Qian *et al.,* 1989). The analyses aimed to the identification of the region important for cellular uptake, reveled a sequence of 16 amino acids, named Penetratin, in the third α-helix of the protein Antennapedia, sufficient for the uptake of the whole protein (Fig. 11).

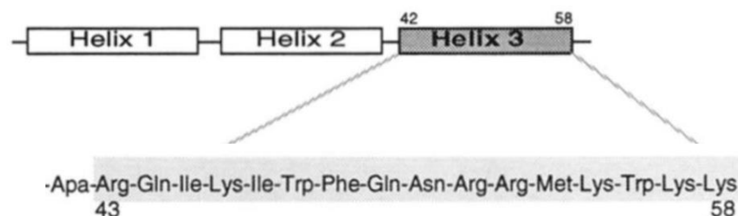


Figure 11. Sequence corresponding to the Antennapedia homeodomain third helix (Derossi *et al*., 1994).

The field of CPPs evolved rapidly and new peptides able to cross the cell membranes are emerging daily marking hard to have a definition embedding all the cell-penetrating peptides discovered. In general, we can define a CPP as a short peptide rich in basic amino acids, with the ability to over cross plasma membranes without the necessity to recognize a receptor. The uptake occurs in an ATP dependent or independent manner, with very low toxic effects for the cells. A secondary α-helix structure is important for protein cellular uptake but not required.
Many other homeo-box domains were characterized for their ability to internalize the whole protein such as Hoxa5, Hoxc8 and PDX-1 (Kong *et al.,* 2008) (Chatelin *et al.,* 1996) (Han *et al.,* 2000).
Pax4, instead, even if has a homeodomain region, holds the property of CPP in its paired domain (Lu *et al.,* 2007).

The transcription factor Oct4, diffusely described for its cruciality to maintain the pluripotent state of embryonic stem cells (Schöler *et al.,* 1990) and for its involvement in somatic cells reprogramming (Nichols *et al*., 1998), owns a protein transduction domain (PTD) in the third helix of its homeodomain. It is represented by a sequence of 16 amino acids sharing a high level of homology with Penetratin sequence of Antennapedia (Fig. 12).

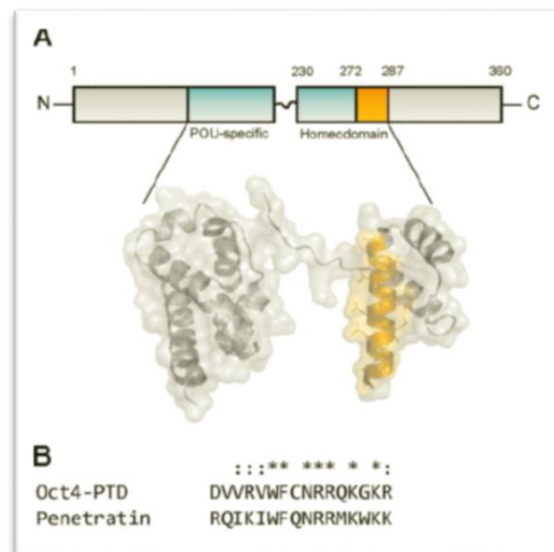


Figure 12. (A): The protein transduction domain (PTD) of Oct4 within its homeodomain is highlighted in orange. (B): Sequence alignment between Oct4-PTD and Penetratin; identic amino acid (*); same amino acid charge (* or :) (Harreither *et al*., 2014).

Differently from Penetratin that accumulates mainly into the cytoplasm, Oct4-PTD diffuses from cytoplasm to the *nucleus.* Moreover, the peptide maintains the ability to translocate the full-length Oct4 protein both in human and mouse cell lines (Harreither *et al,* 2014). This data lays the basis for the use of recombinant Oct4 protein in reprogramming process in order to produce iPSCs in a gene integration free-manner.

2.6.2 Classification of cell penetrating peptides.

CPPs can be classified, according to their origin, in three different classes: peptide derived from natural proteins, including viral proteins, DNA/RNA binding protein, antimicrobial peptides and homeoproteins; chimeric peptides derived from the fusion of natural proteins or sequences and synthetic peptides that are artificially designed peptides based on the sequence of natural proteins.

Peptide	Origin	Sequence
Protein-derived		
Penetratin	Antennapedia (43–58)	RQIKIWFQNRRMKWKK
Tat peptide	Tat(48–60)	GRKKRRQRRRPPQ
pVEC	Cadherin(615–632)	LLIILRRRIRKQAHAHSK
Chimeric		
Transportan	Galanine/Mastoparan	GWTLNSAGYLLGKINLKALAALAKKIL
MPG	HIV-gp41/SV40 T-antigen	GALFLGFLGAAGSTMGAWSQPKKKRKV
Pep-1	HIV-reverse transcriptase/SV40 T-antigen	KETWWETWWTEWSQPKKKRKV
Synthetic		
Polyarginines	Based on Tat peptide	$(R)_{n;}$ $6 < n < 12$
MAP	de novo	KLALKLALKALKAALKLA
R_6W_3	Based on penetratin	RRWWRRWRR

Table 1. Origin and sequences of same CPPs (Bechara and Sagan, 2013).

Another classification method is based on the physic-chemical properties of the sequences. Sequences are classified in cationic, amphipathic and hydrophobic (Milletti, 2012).

The class of cationic peptides are particularly rich in arginine residues and they can easily penetrate cell membranes at physiological pH thanks to the arginine guanidine group. The lysines, instead, because lack the guanidine group, show a reduced capability to penetrate cells, even if, similarly to arginines, they present a net positive charge.

Belong to the class of amphipathic CPPs different peptides such as the murine sequence of vascular endothelium cadherin, pVEC, peptides characterized by a secondary α-helix structure owing both hydrophobic and hydrophilic amino acids distributed on different faces of the helix, like the chimeric MAP and peptides rich in prolines such as Bac7 and SAP (Milletti 2012). Hydrophobic CPPs include both natural and chemical designed peptides containing prenyl groups that facilitate the attachment to cell membranes.

The relevance of cationic peptides able to over cross the cell membranes push to the search of other classes of peptides owing this property. Transcription factors and nuclear proteins are characterized by sequences, typically of 12 residues, rich in basic amino acids important for the transport of the proteins from cytoplasm into the *nucleus*. These sequences, known as nuclear localization signals (NLSs), coming from transcription factors NF-kB, OCT-6, TFIIE-β, TCF1-α, SV40, HATF3 and C. *elegans* SDC3, were tested for the ability of cellular uptake and subcellular localization. What emerged was that all the NLS peptides analyzed, showed a rapid accumulation inside to the cells when added to culture media expanding the list of cationic peptides available for delivery of therapeutic agents in a noninvasive way (Ragin *et al.*, 202).

2.6.3 Cell penetrating peptides applications

Thanks to the capability to over cross cell membranes, CPPs represent a very important tool to transport inside to the cells different covalently or non-covalently linked cargoes ranging from small molecules (drugs and imaging molecules) to nucleic acids, proteins and liposomes.

2.6.3.1 Peptides, proteins and enzymes

As it known, many proteins are unable to passively penetrate cell membrane so, CPPs represent an effective method to delivery inside to the cells proteins with a pharmacological/therapeutic activity.
The transport of small and large 120 KDa proteins through CPPs was achieved both in *vivo* and in *vitro*, and it was employed to treat different diseases. CPPs, for instance, were used to hit cancer cells in different studies. The herpes virus E2 protein was fused to TAT peptide to induce apoptosis in cervical cancer cells (Roeder *et al.*, 2004); the tumor suppressor p53 associated to a synthetic polyarginines peptide was used to block the growth of bladder cancer cells (Araki *et al.*, 2010). Moreover, CPPs were used to release inside to the cells endonucleases able to remove genes causing diseases (Liu *et al.,* 2014) and in the treatment of inflammatory diseases combining inhibitors of NF-kB to different cell penetrating peptides (Orange *et al.,* 2008). Enzymes able to prevent oxidative damage, such as glyoxalase and superoxide dismutase, were coupled to TAT peptide in order to prevent ischemic injury and protect cells from oxidative damages (Shin *et al.,* 2014). TAT peptide was successfully used to transport NF-YA (the regulatory subunits of the trimeric transcription factor NF-Y) inside to human hematopoietic cells inducing the activation of HOXB4 gene and the proliferation of primitive hematopoietic progenitor cells (Domashenko *et al.,* 2010) providing an alternative approach to retroviral stem cells therapy.
Applications of CPPs for protein delivery were used not only to target human cells but also bacteria and fungi. This represents a new approach to inhibit the growth of pathogenic agents bringing benefits to human well-being. Considering the growing antibiotic resistance of bacteria, the possibility to use CPPs as antimicrobial agents or as tool to transport antimicrobial molecules in non-mammalian cells is of great interest and represent a new developing research field.

2.6.3.2 Nucleic acids and liposomes

Cell penetrating peptides are widely used to transport into the cells nucleic acids. Nucleic acids, for their hydrophilic nature, are very inefficient to over-cross plasma membranes. The conjugation of CPPs to nucleic acids improves the delivery of genetic material into the cells exceeding problems related to classic methods of gene delivery: toxicity, low transfection efficiency and low safety profile of viral infection. CPPs were associated also to other molecules to further improve the delivery of

genes after cell entry. As it was recently demonstrated, the association of TAT peptide with the fusion protein dynein light chain Rp3, improved DNA delivery in the nucleus rather than seven fold compared to the transport protein alone (Favaro *et al.,* 2014).

Although cell penetrating peptides have the ability to reach cells also bypassing physical barriers, the main limitation of their application is due to their low cell specificity. For this reason, to better control and direct molecules toward specific cells, CPPs have been incorporated to targeting ligands. Fang demonstrated in his study how the association of TAT peptide with the vascular endothelial growth factor receptor 1, selectively delivered siRNAs against specific targets, in tumor cells (Fang *et al.,* 2013).

Thanks to CPPs, therapeutic gene delivery became possible and much easier to target tissues difficult to access like skin and brain.

Cell penetrating peptides modified micelles, were used to transport siRNA and camptothecin in the brain of rats with malignant glioma through intranasal cavity, prolonging their life (Kanazawa et *al*., 2014) while SPACE peptide was used to topically transport into the skin siRNAs encapsulated in liposomes (Chen *et al.*, 2014).

The combination of cell penetrating peptides with liposomes represents indeed a very promising system for gene delivery. New liposomes, with a good pharmacokinetic profile, have been designed and their association with CPPs, significantly improve their poor intracellular delivery.

2.6.4 Mechanism of internalization

The mechanism used by CPPs to enter cells is still not well understood. However, several uptake mechanisms have been identified. The origin and the properties of peptides provide information about their way to over-cross cell membranes.

Positively charged arginine-rich peptides, generally take advantages from the basic residues to penetrate cells. In fact, the electrostatic interactions with the negatively charged plasma membrane, allow to the peptide to find their way to be internalized.

The cruciality of positively charged residues for cell uptake, was demonstrated by different experiments. The truncation or the substitution of positively charged amino acids with alanine, dramatically reduced the efficiency of TAT peptide to enter cells (Wender *et al.,* 2000). A similar effect was observed for Penetratin after the mutation of basic residues to alanines.

More detailed studies demonstrate that arginine residues are much more effective to favor cell uptake rather than lysines, indeed, the replacement of lysine residues with arginines, is associated to a more efficient uptake. Therefore, the efficiency with which peptides enter cells is associated to the guanidine group of arginine side chain rather than to the only positive charge (Rothbard *et al.,* 2004).

Moreover, it was also demonstrated that the number of arginine residues is fundamental in order that the peptide maintains the best uptake level.
In addition, the secondary structure of CPPs was studied in order to explain the effects of cell membrane perturbation when the peptide contacts the cell surface. However, the relationship between secondary structure and ability to penetrate cells remain an open question. In fact, it was observed that, although a secondary α-helix structure characterizes many cell penetrating peptides, they adopt different structures following the interaction with phospholipids. This variability may be associated, to the experimental conditions such as buffer composition, pH and temperature. Penetratin, for instance, has an α-helix structure but it can also adopt β strand conformation. The orientation of secondary structure elements, have been determined using oriented circular dichroism spectroscopy. These experiments revealed that the helical structures are predominantly perpendicular to the membrane surface, while that of the β-type carbonyls is parallel to the membrane surface (Clayton *et al.,* 2006). What it seems to be relevant for internalization is the structural flexibility and CPPs malleability.
Starting from electrostatic interactions and cell membrane destabilizations, following the contact of the CPPs with the cell surface, two mechanisms have been described to explain the uptake of CPPs into the cells: the endocytosis and the direct translocation (Fig. 13)**.**

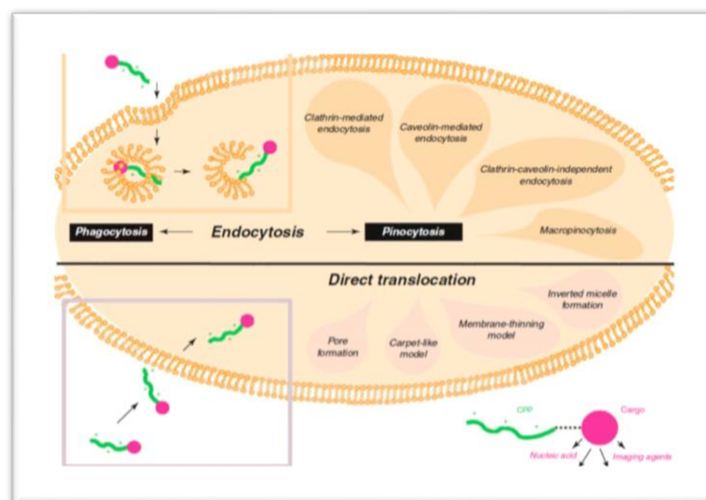


Figure 13. Two different mechanisms for CPPS of internalization: direct translocation and endocytosis.

Endocytosis represents an energy dependent mechanism of internalization that involves different pathways such as macropinocytosis, clathrin-mediated endocytosis, caveolae or lipid-raft-mediated endocytosis, and clathrin- or caveolae-independent endocytosis (Conner *et al.,* 2003).
The choice of endocytotic pathway employed for internalization depends on the nature, on the properties of the peptide and of its connected cargo as well as of the targeted cell type.

It is clear that peptide stability is fundamental to allow its conjugate to reach the intracellular target; proteolytic cleavages and degradation are avoided by the endosomal escape.
It is not so easy determine exactly the specific way that CPPs chose for cell uptake. Different studies aimed to determine the endocytotic pathways are performed using chemical inhibitors and by detecting the colocalization of the peptide with protein markers. The issue is that sometimes, the outcomes of these studies depend on the cell line used and have low specificity, so results have be considered with caution (Ivanov, 2008).
The alternative method used by CPPs to over-cross cell membrane refers to the energy and temperature independent direct translocation pathway. This internalization method is merely based on the direct interaction between peptide and cell membrane and can involve four different pathways: inverted micelle formation, pore formation model, carpet-like model and membrane thinning model. Although these processes are temperature independent, direct translocation is however decreased at 4°C because membrane fluidity is affected at low temperatures (Zaro *et al.,* 2005).
Generally speaking, direct translocation and endocytosis can occur at the same time. Different endocytotic pathways can be used by the same peptide as demonstrated for Penetratin and TAT peptides. Because the internalization of CPPs is not completely understood, many studied pointed the attention on the importance of the interaction between the peptide and the negatively charged cell membranes. Therefore, peptide internalization remains an open challenge for next studies.

3. Aim of the projects and main results

NF-Y is a pioneer transcription factor that, by binding the CCAAT sequence, governs complex genes regulatory networks such as those involved in cell cycle progression and proliferation, cell signaling, DNA repair, metabolism and stem pluripotency pathways.

The increased knowledge about gene regulation, derived from the identification of new *cis*-acting sequences in promoters and enhancers, together with the characterization of their corresponding transcription factors, have represented a valid tool to better understand the activity of NF-Y as transcription regulator and chromatin organizer. My PhD thesis contributed to illustrate new biological properties of NF-Y, through studies ranging from biochemical to functional aspects, in different cell contexts, including non-transformed, cancer and pluripotent stem cells.

3.1 Characterization of NF-YA as Cell Penetrating Peptide

In the last years, many studies have been centered on the identification of new cell penetrating peptides (CPPs) in order to solve problems of drugs delivery through cell membrane to treat and manage several diseases.

As mentioned in the previous paragraphs, CPPs are a class of short peptides that retain the ability to overpass cell plasma membranes and gain access to the cell interior. Their main properties reside in owning amino acid sequences rich in basic residues and in the organization of their secondary structure in alpha helices. CPPs are used as carrier to uptake inside to the cells, proteins, nucleic acids, drugs and molecules. A well-known and characterized CPP, HIV TAT peptide (Park et *al.,* 2002), was demonstrated to efficiently transduce into primary hematopoietic stem cells (HSCs) a functional GST-TAT- NF-YA fusion protein (Domashenko et *al*., 2010).

NF-YA is the limiting subunit of NF-Y, composed by NF-YA, NF-YB and NF-YC subunits, each of which possess an evolutionary conserved domain important for subunit interactions and DNA binding. The analysis of NF-Y/CCAAT tridimensional structure highlighted, in NF-YA subunit, the rich presence of exposed basic aminoacids in the conserved 56- aminoacids residues, organized in two alpha helices (A1 and A2) connected by a linker (Nardini et *al.,* 2013).

These observations have laid the foundations for the rationale of my work. The main project of my PhD was dedicated to i) verify the hypothesis that NF-YA could own the properties of CPP, ii) identify the aminoacidic stretches that confer CPP features, iii) analyze the functionality of the recombinant protein into the cells.

In order to verify the ability of NF-YA to penetrate cells, the two isoforms of NF-YA, NF-YAs and NF-YAl, were produced as recombinant proteins and transfected in HCT116 WT cells. Interestingly,

we noticed a rapid uptake of NF-YA into the *nuclei* at very low concentrations, suggesting that NF-YA is able to penetrate cells in a TAT independent way.
We reasoned that, the stretches of basic aminoacids, contained in the evolutionary conserved domain of NF-YA, could confer to the protein the property to overpass plasma membrane. Mutational analysis identified two aminoacidic stretches that, when mutagenized to alanines, completely abolish NF-YA ability to penetrate cells: the first stretch is in α helix 1- KRR- and the other in alpha helix 2 –RRK and RKR-.
NF-YA, as part of a transcription complex, has a typically nuclear localization. As demonstrated in previous work, basic aminoacids in A1 and A2 helices are important nuclear localization signals (Kahle et *al*., 2005) and they are the same identified in this study as cell penetrating peptides.
The aminoacidic sequences identified as CPPs, in the evolutionary conserved domain of the protein, are also involved in NF-YB/NF-YC subunit interactions and in CCAAT binding. The overlapping of sequence motifs necessary for cellular uptake with the conserved domain of the protein may suggests that the ability of NF-YA to penetrate cells is related to functional purposes. Although, at day, there are not studies about the release of NF-YA from cells, data obtained from this study leave us to assume that NF-YA might be part of a possible mechanism of protein export, alternative to paracrine signaling pathway, where released proteins are received by surrounding cells that, consequently, change their behavior. Concept that can be extended to other transcription factors described as CPPs.
Moreover, we wanted to verify that the ability of recombinant NF-YA to penetrate cells was not cell type or specie specific. For this reason, experiments of protein transduction were performed in different human and mouse cell lines. We observed that NF-YA penetrates efficiently in the *nuclei* of all tested cell types.
Whether, from one side, we noticed that the half-life of the recombinant protein was brief for many transformed cell lines such as HeLa, HCT116, HT1080 and U2OS cells, on the other, we observed that for non-transformed NIH3T3 and HaCaT cells, the trend was different, as they retain NF-YA protein in the *nuclei* for at least 24 hours from transfection. This was a very exciting data because allowed us to assess recombinant NF-YA functionality. We monitored NIH3T3 cell growth, after recombinant NF-YAs and NF-YAl transduction, for 24 and 48 hours. We observed that recombinant NF-YA protein, in NIH3T3 cells, led a cell proliferation delay and this effect was more pronounced with the long isoform.
Although the activity of NF-Y transcription factor is mainly associated to cell growth and cell proliferation, it is likewise a regulator of both anti and pro-apoptotic genes (Benatti *et al.,* 2006).
In this work, we demonstrated that recombinant NF-YA is active after cell translocation. Its overexpression induces cell growth inhibition, characterized by the down modulation of cell cycle

progression genes -*Pcna* and *Ccnb1/2*-, and the activation of apoptosis pathway as demonstrated by the increased transcription levels of *E2f1* (target of NF-Y) and its apoptotic target genes -*Apaf1* and *Bax*- such as by the reduced expression of the anti-apoptotic *Bcl2*.
We know that NF-YA is the limiting subunit of NF-Y and its expression is cell cycle regulated. Therefore, a deregulated expression of the protein, in this case the overexpression, could be perceived by the cells as a "*threat*" for their normal healthy state and the activation of apoptotic response may represent a defense response to this *stimulus*. However, these data are in line with the effects obtained by DNA-mediated NF-YA overexpression in murine fibroblasts where the activation of apoptotic genes, with a mechanism dependent on the up-regulation of E2F1, was described in Gurtner work (Gurtner *et al.,* 2010).

3.2 NF-Y modulates stemness circuitry both in mouse and in human stem cells

Results obtained in the first part of the work, and the well-known properties of NF-YA to stimulate growth and to maintain the pluripotency of stem cells (Zhu et *al.*, 2005; Dolfini et *al.*, 2012), prompted us to investigate recombinant NF-YA also in mouse embryonic stem cells (mESCs) (reported in the last part of the thesis). We observed that after transduction, recombinant NF-YA had a relatively long half-life in mESCs. This result pushed us to verify whether, similarly to GST-TAT-NF-YA, recombinant NF-YA, could counteract leukemia inhibitory factor (LIF) withdrawal. LIF is a cytokine that plays a very important role in the maintenance of mESCs pluripotency. When LIF levels drop, the cells differentiate (He *et al.,* 2006). Our preliminary data demonstrated that, mESCs, daily treated with recombinant NF-YA, and in particular with the short isoform, grow faster and maintain their pluripotency features, compared to control cells, when grown in absence of LIF. In these conditions, after NF-YA transduction, the expression of cell cycle genes and master stemness transcription factors are maintained near to the normal levels, while a drastic reduction in the expression levels of *Nanog*, *Klf4*, *Klf5* and *Ccnb2* is observed in control cells. These data are well in line with what was observed in previous studies where GST-TAT-NF-YAs was described as able to promote stem cell growth and compensate withdrawal of LIF (Dolfini et *al.*, 2012). Data obtained from these experiments, further confirm the functionality of recombinant NF-YA and suggest that a potential future application of recombinant NF-YA, in other stem cell systems, can be considered.
Considering the involvement of NF-Y in the regulation of mESCs stemness circuitry, an additional goal of my work was to investigate NF-Y in human stem cells. Starting from human dermal fibroblasts, we generated induced pluripotent stem cells using Yamanaka factors. Colonies obtained from reprogramming process were characterized for stemness properties and used as model to start this study. Preliminary data obtained from hiPSCs, from their differentiation to embryo bodies and

from NF-Y single subunits inactivation, highlighted similes between human and mouse ESCs about NF-Y activity; in particular, we observed that like mESCs, hiPSCs mainly express the short isoform of NF-YA while the relative abundance of NF-YAl isoform increases upon differentiation. Moreover, the inactivation of NF-Y subunits, and in particular NF-YA silencing, dramatically affects the expression of important stemness genes: SOX2, OCT4, KLF4, NANOG and KLF5 expression levels are indeed greatly reduced. All these evidences suggest that NF-Y influences the stemness circuitry in human stem cells as well as in mESCs.

3.3 NF-Y regulates different metabolic pathways

The manuscript reported in the second part of this thesis, (accepted paper for publication in *Oncotarget*), to which I took part, was dedicated to a study of NF-Y in the regulation of metabolic genes in cancer cells.

The rationale of this work was dictated by: i) analysis of NF-Y regulome that highlight the term "*metabolism*" as particularly enriched among the class of genes regulated by NF-Y in Gene Ontology categorization; ii) bioinformatic analysis highlighting the CCAAT box regulatory sequence enriched in promoters of genes overexpressed in cancer cells; iii) computational analysis showing different metabolic pathways altered in cancer cells.

These observations pushed us to analyze available Affymetrix gene expression profiles of HeLa S3 cells depleted for NF-YA subunit as well as those of mESCs after the inactivation of NF-Y single subunits. Moreover, we performed other experiments of gene expression profiles for HCT116 and H322 cells after NF-YA silencing and for HeLa cells after inactivation of NF-YB subunit.

Farther, we analyzed ENCODE ChIP-seq data for NF-YA and NF-YB in HeLa-S3, GM12878, K562 and mESCs in order to correlate functional data with the effective binding of NF-Y to the CCAAT box. Data collected by these analyses highlighted that NF-Y influences, positively or negatively, different metabolic pathways; while anabolic genes are down modulated after NF-Y subunits depletion, catabolic genes tend to be upregulated.

Although the expression levels of specific metabolic pathway genes are different between cancer and mESCs after NF-Y subunits inactivation, overall what emerged is that NF-Y positively controls genes of the biosynthetic pathway of lipids, cholesterol and fatty acids metabolism, most of them are overexpressed in many cancers and involved in neoplastic transformation.

NF-Y regulates also other metabolic pathways usually altered in cancer cells: it targets genes involved in amino acids metabolism, in particular, the pathways involved in *de novo* synthesis of serine and glycine (the SOCG -Serine, One Carbon, Glycine-), glutamine pathway as well as genes involved in purines and polyamines biosynthesis.

NF-Y activates the transcription of genes involved in glycolysis, but it has a negative or neutral role in regulating mitochondrial respiratory genes.

An important point, emerged from this work, is that in mammalian cells, NF-Y promotes the expression of genes involved in the production of energy through anaerobic pathways, a typical feature of cancer cells. This underlines a change of course during evolution as, in yeast *S.cerevisiae*, NF-Y promotes the expression of genes involved in mitochondrial energy production. Data emerged from this study highlighted that NF-Y does not favor anapleroris process but promotes the expression of genes involved in pathways out of Krebs cycle.

These data provide an important starting point to understand how NF-Y metabolic targets are involved in cancer growth and development. In the future, the identification of new NF-Y TFs partners, involved in metabolic genes expression, such as further information about their DNA regulatory elements, today poorly understood, will provide further news about the regulation of these processes.

4. Conclusions and future perspectives

The work performed during my PhD describes, for the first time, NF-YA as new component of the class of proteins owning the properties of cell penetrating peptide and contributes to enrich the knowledge about the properties and the features of this protein. We demonstrated that specific basic aminoacids residues, in the evolutionary conserved domain of NF-YA, are fundamental to allow to the protein cell penetration and *nucleus* translocation.

In the last two decades, the application of CPPs took hold in biotechnological applications thanks to their ability to over-cross physical barriers and transport cargos to pelt their targets.

The description of NF-YA carrying protein transduction domains, and in particular the proved functionality of the protein into the cells after translocation, represent a very important starting point for its potential application in clinical therapeutic. Data collected during my PhD suggest that, recombinant NF-YA, may represent a valid tool to undertake cell treatments and studies bypassing conventional therapy applications. The use of recombinant NF-YA represents indeed an efficient and no toxic tool to overexpress NF-YA avoiding classic methods of gene delivery.

The already tested GST-TAT-NF-YA protein, used to stimulate the growth of HSCs, in treatments of bone marrow transplantation, could be replaced by recombinant NF-YA devoid of tags or epitopes. Considering the capability of NF-YA protein to maintain the pluripotent state of stem cells and to promote their growth, further applications could relate its use to the treatment of degenerative diseases characterized by depletion of stem cells pool (e.g. muscular dystrophy).

Moreover, it will be interesting to validate whether the identified CPPs, in the evolutionary conserved domain of NF-YA, work likewise when fused to other molecules or proteins.

A question remains to be clarified: which is the biological function of NF-YA cellular uptake and, at the same time, which is the mechanism of its cell exit. A possible action of NF-YA in the biological process of monocytes maturation to macrophages, during inflammation, is an appealing hypothesis. We know that monocytes express NF-YB/YC subunits but not NF-YA. Their maturation needs the activation of genes, which are under NF-Y trimer control. Therefore, monocytes could uptake NF-YA protein released from the nearby damaged tissues during an inflammatory state, hence triggering their maturation to adult macrophages.

Finally, data obtained from hiPSCs, encourage us to further investigate the relationship between NF-Y and other human master stemness TFs, in order to better collocate the role of NF-Y in human stemness maintenance and differentiation. The employment of recombinant NF-YA may be useful to shed light on these proposes as well as ChIP experiments and cell differentiation toward a specific cell lineage.

5.Reference

Araki, D., Takayama, K., Inoue, M., Watanabe, T., Kumon, H., Futaki, S., Matsui, H., and Tomizawa, K. (2010). Cell-penetrating D-isomer peptides of p53 C-terminus: long-term inhibitory effect on the growth of bladder cancer. Urology *75,* 813-819.

Bai, X., Kim, J., Yang, Z., Jurynec, M.J., Akie, T.E., Lee, J., LeBlanc, J., Sessa, A., Jiang, H., DiBiase, A., Zhou, Y., Grunwald, D.J., Lin, S., Cantor, A.B., Orkin, S.H., and Zon, L.I. (2010). TIF1gamma controls erythroid cell fate by regulating transcription elongation. Cell. *142*, 133-143.

Bechara, C., and Sagan, S. (2013). Cell-penetrating peptides: 20 years later, where do we stand? Letters *587*, 1693–1702.

Bellorini, M., Lee, D.K., Dantonel, J.C., Zemzoumi, K., Roeder, R.G., Tora, L., and Mantovani, R. (1997). CCAAT binding NF-Y-TBP interactions: NF-YB and NF-YC require short domains adjacent to their histone fold motifs for association with TBP basic residues. Nucleic Acids Res *25*, 2174–2181.

Benatti P, Basile V, Merico D, Fantoni LI, Tagliafico E, Imbriano C. A balance between NF-Y and p53 governs the pro- and anti-apoptotic transcriptional response.

Berk, A.J. (2012). Yin and yang of mediator function revealed by human mutants. Proc Natl Acad Sci USA *109*, 19519–19520.

Bhattacharya, A., Deng, J.M., Zhang, Z., Behringer, R., de Crombrugghe, B. and Maity, S.N. (2003). The B subunit of the CCAAT box binding transcription factor complex (CBF/NF-Y) is essential for early mouse development and cell proliferation. Cancer Res *63*, 8167–8172.

Bird, A. (1987). CpG islands as gene markers in the vertebrate nucleus. Trends Genet. *3*, 342–347.

Blum, R., Elkon, R., Yaari, S., Zundelevich, A., Jacob-Hirsch, J., Rechavi, G., Shamir, R., and Kloog, Y. (2007). Gene expression signature of human cancer cell lines treated with the ras inhibitor salirasib (S-farnesylthiosalicylic acid). Cancer Res 67:3320–3328.

Blum, R., Gupta, R., Burger, P.E., Ontiveros, C.S., Salm, S.N., Xiong, X., Kamb, A., Wesche, H., Marshall, L., Cutler, G., Wang, X., Zavadil, J., Moscatelli, D., and Wilson, E.L. (2009). Molecular signatures of prostate stem cells reveal novel signaling pathways and provide insights into prostate cancer. PLoS ONE 4:e5722.

Bolognese, F., Wasner, F., Dohna, C. L., Gurtner, A., Ronchi, A., Muller, H., Manni, I., Mossner, J., Piaggio, G., Mantovani, R., and Engeland, K. (1999). The cyclin B2 promoter depends on NF-Y, a trimer whose CCAAT-binding activity is cell cycle regulated. Oncogene *18*, 1845–1853.

Borggrefe, T., and Yue, X. (2011). Interactions between subunits of the Mediator complex with gene-specific transcription factors. Semin. Cell Dev. Biol. *22*, 759–768.

Calvo, A., Perez-Stable, C., Segura, V., Catena, R., Guruceaga, E., Nguewa, P., Blanco, D., Parada, L., Reiner, T., and Green, J.E. (2010). Molecular characterization of the Ggamma-globin-Tag transgenic mouse model of hormone refractory prostate cancer: Comparison to human prostate cancer. Prostate *70*, 630–645.

Carey, M., Lin, Y.S., Green, M.R., and Ptashne, M. (1990). A mechanism for synergistic activation of a mammalian gene by GAL4 derivatives. Nature *345,* 361–364.

Ceribelli, M., Dolfini, D., Merico, D., Gatta, R., Vigano`, A. M., Pavesi, G., and Mantovani, R. (2008). The histone-like NF-Y is a bifunctional transcription factor. Molecular and cellular biology *28*, 2047–2058.

Chae, H.D., Yun, J., Bang, Y.J., and Shin, D.Y. (2004). Cdk2-dependent phosphorylation of the NF-Y transcription factor is essential for the expression of the cell cycle-regulatory genes and cell cycle G1/S and G2/M transitions. Oncogene *23*, 4084–4088.

Chan, Q.K., Lam, H.M., Ng, C.F., Lee, A.Y., Chan, E.S., Ng, H.K., Ho, S.M., and Lau, K.M. (2010). Activation of GPR30 inhibits the growth of prostate cancer cells through sustained activation of Erk1/2, c-jun/c-fosdependent upregulation of p21, and induction of G(2) cell-cycle arrest. Cell Death Differ *17*, 1511–1523.

Chang, Z. F., and Liu, C. J. (1994). 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. *26*, 17893–17898.

Chassanidis, C., Kalamaras, A., Phylactides, M., Pourfarzad, F., Likousi, S., Maroulis, V., Papadakis, M.N., Vamvakopoulos, N.K., Aleporou-Marinou, V., Patrinos, G.P., and Kollia, P. (2009). The Hellenic type of nondeletional hereditary persistence of fetal hemoglobin results from a novel mutation (g.-109G>T) in the HBG2 gene promoter. Ann Hematol *88*, 549–555.

Chatelin, L., Volovitch, M., Joliot, H., Perez, F., and Prochiantz, A. (1996). Transcription factor hoxa-5 is taken up by cells in culture and conveyed to their nuclei. Mech Dev *55*, 111–117.

Chen, X.W., Mo, Q.H., Li, Q., Zeng, R., and Xu, X.M. (2007). A novel mutation of -73(A–>T) in the CCAAT box of the beta-globin gene identified in a patient with the mild beta-thalassemia intermedia. Ann Hematol *86*, 653–657.

Chen, L., and Widom, J. (2005). Mechanism of transcriptional silencing in yeast. Cell *120*, 37–48.
Chen, M., Zakrewsky, M., Gupta, V., Anselmo, A. C., Slee, D. H., Muraski, J. A. and Mitragotri, S. (2014). Topical delivery of siRNA into skin using SPACE-peptide carriers. Journal of Controlled Release *179*, 33-41.

Baldwin, A.S., Carthew, R.W., and Sharp, P.A. (1988). Human CCAAT-binding proteins have heterologous subunits. Cell *8,* 11-24.

Clayton, A.H., Atcliffe, B.W., Howlett, G.J. and Sawyer, W.H. (2006). Conformation and orientation of penetratin in phospholipid membranes. J. Pept. Sci. *12*, 233–238.

Conaway, R.C., and Conaway, J.W. (2011). Function and regulation of the Mediator complex. Curr. Opin. Genet. Dev. *21*, 225–230.

Conner, S.D. and Schmid, S.L. (2003). Regulated portals of entry into the cell. Nature *422*, 37–44

Das, C., Lucia, M.S., Hansen, K.C.,and Tyler, J.K. (2009). CBP/p300-mediated acetylation of histone H3 on lysine 56. Nature *459*, 113-117.

Delvoye, N.L., Destroismaisons, N.M., and Wall, LA. (1993). Activation of the beta-globin promoter by the locus control region correlates with binding of a novel factor to the CAAT box in murine erythroleukemia cells but not in K562 cells. Mol Cell Biol *13,* 6969-83.

Di Agostino, S., Strano, S., Emiliozzi, V., Zerbini, V., Mottolese, M., Sacchi, A., Blandinoand, G., and Piaggio, G. (2006). Gain of function of mutant p53: The mutant p53/NF-Y protein complex reveals an aberrant transcriptional mechanism of cell cycle regulation. Cancer cell *10*, 191–202.

Ding, N., Zhou, H., Esteve, P.O., Chin, H.G., Kim, S., Xu, X., Joseph, S.M.,Friez, M.J., Schwartz, C.E., Pradhan, S., and Boyer, T.G. (2008). Mediator links epigenetic silencing of neuronal gene expression with x-linked mental retardation. Mol. Cell *31*, 347–359.

Dolfini, D., Zambelli, F., Pavesi, G., and Mantovani, R. (2009). A perspective of promoter architecture from the CCAAT box. Cell Cycle *8,* 4127-4137.

Dolfini, D., Minuzzo, M.,Pavesi, G., Mantovani, R. (2012). The Short Isoform of NF-YA Belongs to the Embryonic Stem Cell Transcription Factor Circuitry. STEM CELLS *30*, 2450–2459.

Domashenko, A.D., Danet-Desnoyers, G., Aron, A. et al. (2012). TAT-mediated transduction of NF-Ya peptide induces the ex vivo proliferation and engraftment potential of human hematopoietic progenitor cells. Blood *116*, 2676–2683.

Dorn, A., Bollekens, J., Staub, A., Benoist, C., and Mathis, D. (1987). A multiplicity of CCAAT box-binding proteins. Cell *50*, 863–872.

Dou, Y., Milne, T.A., Tackett, A.J., Smith, E.R., Fukuda, A., Wysocka, J., Alli, C.D., Chait, B.T., Hess, J.L., and Roeder, R.G. (2005). Physical association and coordinate function of the H3 K4 methyltransferase MLL1 and the H4 K16 acetyltransferase MOF. Cell *121,* 873-885.

Eberharter, A., and Becker, P. ATP-dependent nucleosome remodelling: factors and functions. Journal of Cell Science *117*, 3707-3711.

Farina, A., Manni, I., Fontemaggi, G., Tiainen, M., Cenciarelli, C., Bellorini, M., Mantovani, R., Sacchi, A., and Piaggio, G. (1999). Down-regulation ofcyclin B1 gene transcription in terminally differentiated skeletal muscle cells is associated with loss of functional CCAAT-binding NF-Y complex. Oncogene *18*, 2818–2827.

Favaro, M., Toledo, d. M., Alves, R., Santos, C., Beloti, L., Janissen, R., de la Torre, L., Souza, A., and Azzoni, A. (2014). Development of a non-viral gene delivery vector based on the dynein light chain Rp3 and the TAT peptide. Journal of biotechnology *173,* 10-18.

Feng, S. and Holland, E.C. (1988). HIV-1 tat trans-activation requires the loop sequence within tar. Nature *334*, 165–167.

Forsberg EC, Passegué E, Prohaska SS, Wagers AJ, Koeva M, Stuart JM, and Weissman IL. (2010). Molecular signatures of quiescent, mobilized and leukemia-initiating hematopoietic stem cells. PLoS ONE 5:e8785.

Framson, P., and Bornstein P. (1993). A serum response element and a binding site for NF-Y mediate the serum response of the human thrombospondin 1 gene. J Biol Chem *5,* 4989-96.

Frankel, A. D. and C. O. Pabo (1988). Cellular uptake of the tat protein from human immunodeficiency virus. Cell *55*, 1189-1193.

Frontini, M., Imbriano, C., diSilvio, A., Bell, B., Bogni, A., Romier, C., Moras, D., Tora, L., Davidson, I., and Mantovani, R. (2002). NF-Y Recruitment of TFIID, Multiple Interactions with Histone Fold TAFIIs. J Biol Chem *277*, 5841–5848.

Gardiner-Garden, M., and Frommer, M. (1987). CpG islands in vertebrate genomes. J. Mol. Biol. *196*, 261–82.

Goodarzi, H., Elemento, O., and Tavazoie, S. (2009). Revealing global regulatory perturbations across human cancers. Mol Cell *36*, 900–911.

Grskovic, M., Chaivorapol, C., Gaspar-Maia,A., Li, H., and Ramalho-Santos, M. (2007). Systematic identifi- cation of cis-regulatory sequences active in mouse and human embryonic stem cells. PLoS Genet *3*, 1524-1540.

Gurtner, A., Manni, I., Fuschi, P., Mantovani, R., Guadagni, F., Sacchi, A., and Piaggio, G. (2003). Requirement for down-regulation of the CCAAT binding activity of the NF-Y transcription factor during skeletal muscle differentiation. Mol Biol Cell *14*, 2706–2715.

H. Izumi, Molander, C., Penn, L.Z., Ishisaki, A., Kohno, K., and Funa, K. (2001). Mechanism for the transcriptional repression by c-Myc on PDGF (β)-receptor. J Cell Sci *114*, 1533-1544.

Hamanaka, R, Smith, M.R., and O'Connor, P.M. (1995). Polo-like kinase is a cell cycle-regulated kinase activated during mitosis. J Biol Chem *270,* 21086-21091.

Han, K., Jeon, M.J., Kim, K.A., Park, J., and Choi, S.Y. (2000). Efficient intracellular delivery of GFP by homeodomains of Drosophila Fushi-tarazu and Engrailed proteins. Mol Cells *10*, 728–732.

Han, M., Toli, J., and Abdellatif, M. (2011). MicroRNAs in the cardiovascular system. Curr. Opin. Cardiol. *26*, 181–189.

Harreither, E., Rydberg, H.A., Åmand, H.L., Jadhav, V., Fliedl, L., Benda, C., Esteban, M.A., Pei, D., Borth, N., Grillari-Voglauer, R., Hommerding, O., Edenhofer, F., Nordén, B., and Grillari, J. (2014). Characterization of a novel cell penetrating peptide derived from human Oct4. Cell Regeneration 3:2.

He, Z., Li, J.J., Zhen, C.H., Feng, L.Y., Ding, X.Y. (2006). Effect of leukemia inhibitory factor on embryonic stem cell differentiation: implications for supporting neuronal differentiation. Acta Pharmacol Sin. 27, 80-90.

Hodawadekar, S.C., and Marmorstein, R. (2007). Chemistry of acetyl transfer by histone modifying enzymes: structure, mechanism and implications for effector design. Oncogene *26*, 5528-5540.

Hu, Q. and Maity, S.N. (2000) Stable expression of a dominant negative mutant of CCAAT binding factor/NF-Y in mouse fibroblast cells resulting in retardation of cell growth and inhibition of transcription of various cellular genes.J. Biol. Chem *275*, 4435–4444.

Hu, Q., Lu, J.F., Luo, R., Sen, S. and Maity, S.N. (2006) Inhibition of CBF/NF-Y mediated transcription activation arrests cells at G2/M phase and suppresses expression of genes activated at G2/M phase of the cell cycle. Nucleic Acids Res *34, 6*272–6285.

Huang, S., Ling, J.J., Yang, S., Li, X.J., and Li, S. (2011). Neuronal expression of TATA box-binding protein containing expanded polyglutamine in knock-in mice reduces chaperone protein response by impairing the function of nuclear factor-Y transcription factor. Brain *134*, 1943–1958.

Huang, W., Zhao, S., Ammanamanchi, S., Brattain, M., Venkatasubbarao, K. and Freeman, J.W. (2005). Trichostatin A induces transforming growth factor beta type II receptor promoter activity and acetylation of Sp1 by recruitment of PCAF/p300 to a Sp1.NF-Y complex. J Biol Chem *280*, 10047-10054.

Hughes, R., Kristiansen, M., Lassot, I., Desagher, S., Mantovani, R. and Ham, J. (2011). NF-Y is essential for expression of the proapoptotic bim gene in sympathetic neurons. Cell Death and Differentiation *18*, 937–947.

Ivanov, A.I. (2008). Pharmacological inhibition of endocytic pathways: is it specific enough to be useful? Methods Mol. Biol. *440*, 15–33.

Jackson, S.M., Ericsson, J., Osbornei, T.F., and Edwards, P.A. (1995). NF-Y has a novel role in steroldependent transcription of two cholesterogenic genes. J Biol Chem *270*, 21445–21448.

Jones, P.A., Laird, P.W. (1999). Cancer epigenetics comes of age. Nat. Genet. *21*, 163–167.

Jürchott, K., Kuban, R.J., Krech, T., Blüthgen, N., Stein, U., Walther, W., Friese, C., Kielbasa, S.M., Ungethüm, U., Lund, P., Knösel, T., Kemmner, W., Morkel, M., Fritzmann, J., Schlag, P.M., Birchmeier, W., Krueger, T., Sperling, S., Sers, C., Royer, H.D., Herzel, H., and Schäfer, R. (2010). Identification of Y-box binding protein 1 as a core regulator of MEK/ERK pathwaydependent gene signatures in colorectal cancer cells. PLoS Genet 6:e1001231.

Joliot, A., Pernelle, C., Deagostini-Bazin, H., and Prochiantz, A. (1991). Antennapedia homeobox peptide regulates neural morphogenesis. Proc Natl Acad Sci U S A *88*, 1864-1868.

Kagey, M.H., Newman, J.J., Bilodeau, S., Zhan, Y., Orlando, D.A., van Berkum, N.L., Ebmeier C.C., Goossens, J., Rahl, P.B., Levine, S.S., Taatjes, D.J., Dekker, J., and Young, R.A. (2010). Mediator and cohesin connect gene expression and chromatin architecture.Nature *467*, 430–435.

Kahle, J., Baake, M., Doenecke, D., and Albig, W. (2005). Subunits of the heterotrimeric transcription factor NF-Y are imported into the nucleus by distinct pathways involving importin β and importin 13. Molecular and cellular biology *25*, 5339–5354.

Kanazawa, T., Morisaki, K., Suzuki,S., and Takashima, Y. (2014). Prolongation of life in rats with malignant glioma by intranasal siRNA/drug codelivery to the brain with Cell-Penetrating Peptide-modified micelles. Molecular Pharmaceutics *11*, 1471-1478.

Kaufmann, R., Straussberg, R., Mandel, H., Fattal-Valevski, A., Ben-Zeev, B.,Naamati, A., Shaag, A., Zenvirt, S., Konen, O., Mimouni-Bloch, A., et al. (2010).Infantile cerebral and cerebellar atrophy is associated with a mutation in the MED17 subunit of the transcription preinitiation mediator complex. Am. J.Hum. Genet. *87*, 667–670.

Kim, J.H., Lee, J.N., and Paik, Y.K. (2001). Cholesterol Biosynthesis from LanosterolA concerted role for sp1 and nf-y-binding sites for sterol-mediated regulation of rat 7-dehydrocholesterol reductase gene expression. J Biol Chem *276*, 18153–18160.

Kong, K.A., Gadi, J., Park, H.W., Bok, J., and Kim, M.H. (2008). The third helix of the murine Hoxc8 homeodomain facilitates protein transduction in mammalian cells. Biochem Biophys Res Commun *377*, 161–164.

Kristiansen, M., and Ham, J. (2014). Programmed cell death during neuronal development: the sympathetic neuron model. Cell Death and Differentiation *21*, 1025–1035.

Krivega, I., and Dean, A. (2012). Enhancer and promoter interactions-long distance calls. Curr Opin Genet Dev *22*, 79–85.

Lagor, W.R., de Groh, E.D., and Ness, G.C. (2005). Diabetes alters the occupancy of the hepatic 3-hydroxy-3-methylglutaryl-CoA reductase promoter. J Biol Chem *280*, 36601–36608.

Landsberger, N., and Wolffe, A.P. (1995). Role of chromatin and Xenopus laevis heat shock transcription factor in regulation of transcription from the X. laevis hsp70 promoter in vivo. Mol Cell Biol *15*, 6013–6024.

Larivie` re, L., Seizl, M., and Cramer, P. (2012). A structural perspective on Mediator function. Curr. Opin. Cell Biol. *24*, 305–313.

Lee, M.R., Kim, S.H., Cho, H.J., Lee, K.Y., Moon, A.R., Jeong, H.G., Lee, J.S., Hyun, J.W., Chung, M.H., and You, H.J. (2004). Transcription factors NF-YA regulate the induction of human OGG1 following DNA-alkylating agent methylmethane sulfonate (MMS) treatment. J Biol Chem *279*, 9857–9866.

Lettice, L.A., Heaney, S.J., Purdie, L.A., Li, L., de Beer, P., Oostra, B.A., Goode, D., Elgar, G., Hill, R.E., and de Graaff, E. (2003). A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. Hum. Mol. Genet. *12,* 1725,1735.

Li, Q., Peterson, K.R., Fang, X., and Stamatoyannopoulos, G. (2002). Locus control regions. Blood*100*, 3077–3086.

Li, X.Y., Mantovani, R., Hooft van Huijsduijnen, R., Andre, I., Benoist, C. and Mathis, D. (1992). Evolutionary variation of the CCAAT-binding transcription factor NF-Y. Nucleic Acids Res *20*, 1087-1091.

Liberati, C., Cera, M.R., Secco, P., Santoro, C., Mantovani, R., Ottolenghi, S., and Ronchi, A. (2001). Cooperation and competition between the binding of COUP-TFII and NF-Y on human epsilon- and gammaglobin gene promoters. J Biol Chem *276*, 41700–41709.

Lim, C.Y., Santoso, B., Boulay, T., Dong, E., Ohler, U., and Kadonaga, J.T. (2004). The MTE, a new core promoter element for transcription by RNA polymerase II. *Genes Dev*. *18*, 1606–1617.

Lin, C.Y., Love´ n, J., Rahl, P.B., Paranal, R.M., Burge, C.B., Bradner, J.E., Lee,T.I., and Young, R.A. (2012). Transcriptional amplification in tumor cells with elevated c-Myc. Cell *151*, 56–67.

Lin, Y.C., Chen, Y.N., Lin, K.F., Wang, F.F., Chou, T.Y. and ChenY-C, M.Y. (2014). 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 and Disease *5*, e987.

Lin, Y.S., Carey, M., Ptashne, M., and Green, M.R. (1990). How different eukaryotic transcriptional activators can cooperate promiscuously. Nature *345*, 359–61.

Lu, J., Li, G., Lan, M.S., Zhang, S., Fan, W., Wang, H., and Lu, D. (2007). Pax4 paired domain mediates direct protein transduction into mammalian cells. Endocrinology *148*, 5558–5565.

Lu, Y.H., Dallner, O.S., Birsoy, K., Fayzikhodjaeva, G., and Friedman, J.M. (2015). Nuclear Factor-Y is an adipogenic factor that regulates leptin gene expression. Molecular metabolism *4,* 392-405.

Luo, R., Klumpp, S.A., Finegold, M.J. and Maity, S.N. Inactivation of CBF/NF-Y in postnatal liver causes hepatocellular degeneration, lipid deposition, and endoplasmic reticulum stress. Scientific reports 1 : 136.

Maity, S.N., and de Crombrugghe, B. (1998). Role of the CCAAT-binding protein CBF/NF-Y in transcription. Trends Biochem Sci. *23*, 174-178.

Malecki, M.T. (2005). Genetics of type 2 diabetes mellitus. Diabetes Res. Clin.Pract. *68*, 10–21.

Malik, S., and Roeder, R.G. (2010). The metazoan Mediator co-activator complex as an integrative hub for transcriptional regulation. Nat. Rev. Genet. *11*, 761–772.

Manni, I., Caretti, G., Artuso, S., Gurtner, A., Emiliozzi, V., Sacchi, A., Mantovani, R., and Piaggio, G. (2008). Posttranslational regulation of NF-YA modulates NF-Y transcriptional activity. Molecular Biology of the Cell *19*, 5203–5213.

Mantovani, R. (1998). A survey of 178 NF-Y binding CCAAT boxes. Nucleic Acids Research *26*, 1135–1143

Mantovani, R., Pessara, U., Tronche, F., Li, X.Y., Knapp, A.M., Pasquali, J.L., Benoist, C., and Mathis, D. (1992). Monoclonal antibodies to NF-Y define its function in MHC class II and albumin gene transcription. EMBO J *11,* 3315–3322.

Maurano, M.T., Humbert, R., Rynes, E., Thurman, R.E., Haugen, E., Wang, H., Reynolds, A.P., Sandstrom, R., Qu, H., Brody, J., Shafer, A., Neri, F., Lee, K., Kutyavin, T., Stehling-Sun, S., Johnson, A.K., Canfield, T.K., Giste, E., Diegel, M., Bates, D., Hansen, R.S., Neph, S., Sabo, P.J., Heimfeld, S., Raubitschek, A., Ziegler, S., Cotsapas, C., Sotoodehnia, N., Glass, I., Sunyaev, S.R., Kaul, R., and Stamatoyannopoulos, J.A. (2012). Systematic localization of common disease-associated variation in regulatory DNA. Science *337*, 1190–1195.

Milletti, F. (2012). Cell-penetrating peptides: classes, origin, and current landscape. Drug Discov. Today *17*, 850–860.

Nardini, M., Gnesutta, N., Donati, G., Gatta, R., Forni, C., Fossati, A.,Vonrhein, C., Moras, D., Romier, C., Bolognesi, M., and Mantovani, R. (2013). Sequence-specific transcription factor NF-Y displays histone-like DNA binding and H2B-like ubiquitination. Cell *152*, 132–143.

Ng, S.S., Yue, W.W., Oppermann, U., and Klose, R.J. (2009). Dynamic protein methylation in chromatin biology. Cell Mol Life Sci *66*, 407-422.

Nie, Z., Hu, G., Wei, G., Cui, K., Yamane, A., Resch, W., Wang, R., Green, D.R.,Tessarollo, L., Casellas, R., et al. (2012). c-Myc is a universal amplifier of expressed genes in lymphocytes and embryonic stem cells. Cell *151*, 68–79.

Niimi, T., Hayashi, Y., Futaki, S. and Sekiguchi, K. (2004). SOX7 and SOX17 regulate the parietal endoderm-specific enhancer activity of mouse laminin alpha1 gene. J Biol Chem *279*, 38055-38061.

Orange, J.S. and May, M.J. (2008) Cell penetrating peptide inhibitors of nuclear factor-kappa B. Cell. Mol. Life Sci. *65*, 3564–3591.

Ott, M., Dorr, A., Hetzer-Egger, C., Kaehlcke, K., Schnolzer, M., Henklein, P., Cole, P., Zhou, M.M., and Verdin, E. (2004). Tat acetylation: a regulatory switch between early and late phases in HIV transcription elongation. Novartis Found Symp *259*, 182–193.

Papageorgiou, N., Tousoulis, D., Androulakis, E., Siasos, G., Briasoulis, A.,Vogiatzi, G., Kampoli, A.M., Tsiamis, E., Tentolouris, C., and Stefanadis, C. (2012). The role of microRNAs in cardiovascular disease. Curr. Med. Chem.*19*, 2605–2610.

Park, J., J. Ryu, K. A. Kim, H. J. Lee, J. H. Bahn, K. Han, E. Y. Choi, K. S. Lee, H. Y. Kwon and Choi, S. Y. (2002). 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 *83*, 1173-1181.

Parthun, M.R. (2007). Hat1: the emerging cellular roles of a type B histone acetyltransferase. Oncogene *26*, 5319-5328.

Pooga, M., U. Soomets, M. Hallbrink, A. Valkna, K. Saar, K. Rezaei, U. Kahl, X.J. and Jing-Ptashne, M., and Gann, A. (1997). Transcriptional activation by recruitment. Nature *386*, 569–577.

Qian, Y. Q., Billeter, M., Otting, G., Muller, M., Gehring, W. J., and Wuthrich, K. (1989) Cell *59*, 573-580.

Ragin, A. D., Morgan, R.A., and Chmielewski, J. (2002). Cellular Import Mediated by Nuclear Localization Signal Peptide Sequences. Chemistry & Biology *8*, 943–948.

Raymondjean, M., Cereghini, S., and Yaniv, M. (1988). Several distinct "CCAAT" box binding proteins coexist in eukaryotic cells. Proc Natl Acad Sci USA *85,* 757-761.

Reith, W., Siegrist, C.A., Durand, B., Barras, E. and Mach, B. (1994). Function of major histocompatibility complex class II promoters requires cooperative binding between factors RFX and NF-Y. Proc Natl Acad Sci U S A *91*, 554-558.

Rhodes, D.R., Kalyana-Sundaram, S., Mahavisno, V., Barrette, T.R., Ghosh, D., and Chinnaiyan, A.M. (2005). Mining for regulatory programs in the cancer transcriptome. Nat Genet *37*, 579–583.

Rodríguez-Paredes, M., and Esteller, M. (2011). Cancer epigenetics reaches mainstream oncology. Nature Medicine *17*, 330–339.

Roeder, G. E., Parish, J. L., Stern, P. L. and Gaston, K. (2004). Herpes simplex virus VP22–human papillomavirus E2 fusion proteins produced in mammalian or bacterial cells enter mammalian cells and induce apoptotic cell death. Biotechnology and applied biochemistry *40*, 157-165.

Romier, C., Cocchiarella, F., Mantovani, R., and Moras, D. (2003). The NF-YB/NF-YC structure gives insight into DNA binding and transcription regulation by CCAAT factor NF-Y. J Biol Chem. *10*, 1336-1345.

Rothbard, J.B., Jessop, T.C., Lewis, R.S., Murray, B.A. and Wender, P.A. (2004). Role of membrane potential and hydrogen bonding in the mechanism of translocation of guanidinium-rich peptides into cells. J. Am. Chem. Soc. *126*, 9506–9507.

Salvatore, G., Nappi, T.C., Salerno, P., Jiang, Y., Garbi, C., Ugolini, C., Miccoli, P., Basolo, F., Castellone, M.D., Cirafici, A.M., Melillo, R.M., Fusco, A., Bittner, M.L., and Santoro, M. (2007). A cell proliferation and chromosomal instability signature in anaplastic thyroid carcinoma. Cancer Res *67*, 10148–10158.

Scafoglio, C., Ambrosino, C., Cicatiello, L., Altucci, L., Ardovino, M., Bontempo, P., Medici, N., Molinari, A.M., Nebbioso, A., Facchiano, A., Calogero, R.A., Elkon, R., Menini, N., Ponzone, R., Biglia, N., Sismondi, P., De Bortoli, M., and Weisz, A. (2006). Comparative gene expression profiling reveals partially overlapping but distinct genomic actions of different antiestrogens in human breast cancer cells. J Cell Biochem *98*, 1163–1184.

Schmitt, S., Prestel, M., and Paro, R. (2005). Intergenic transcription through a polycomb group response element counteracts silencing. Genes Dev. *19*, 697–708.

Schöler, H.R., Ruppert, S., Suzuki, N., Chowdhury, K., and Gruss, P. (1990). New type of POU domain in germ line-specific protein Oct-4. Nature *344*, 435–439.

Sinha, S., Adler, A.S., Field, Y., Chang, H.Y., and Segal, E. (2008). Systematic functional characterization of cis-regulatory motifs in human core promoters. Genome Res *18*, 477–488.

Shin, M. J., Kim, D. W., Lee, Y. P., Ahn, E. H., Jo, H. S., Kim, D.S., Kwon,O.S., Kang, T.C., Cho, Y.J. and Park, J. (2014). Tat-glyoxalase protein inhibits against ischemic neuronal cell damage and ameliorates ischemic injury. Free Radical Biology and Medicine *67,* 195-210.

Spaeth, J.M., Kim, N.H., and Boyer, T.G. (2011). Mediator and human disease. Semin. Cell Dev. Biol. *22*, 776–787.

Srinivasan, L., and Atchison, M.L. (2004). YY1 DNA binding and PcG recruitment requires CtBP. Genes Dev. *18*, 2596–2601.

Steward, M.M., Lee, J.S., O'Donovan, A., Wyatt, M., Bernstein, B.E., and Shilatifard, A. (2006). Molecular regulation of H3K4 trimethylation by ASH2L, a shared subunit of MLL complexes. Nat Struct Mol Biol *13,* 852-854.

Sunyakumthorn, P.,Boonsaen, T.,Boonsaeng, V., Wallace, J.C., and Jitrapakdee, S. (2005). Involvement of specific proteins (Sp1/Sp3) and nuclear factor Y in basal transcription of the distal promoter of the rat pyruvate carboxylase gene in *b*-cells. Biochemical and Biophysical Research Communications *329*, 188–196.

Taatjes, D.J. (2010). The human Mediator complex: a versatile, genome-wide regulator of transcription. Trends Biochem. Sci. *35*, 315–322.

Teran-Garcia, M., Rufo, C., Nakamura, M.T., Osborne, T.F., and Clarke, S.D. (2002). NF-Y Involvement in the polyunsaturated fat inhibition of fatty acid synthase gene transcription. Biochemical and Biophysical Research Communications *290,* 1295–1299.

Thomassen, M., Tan, Q., and Kruse, T.A. (2008). Gene expression meta-analysis identifies metastatic pathways and transcription factors in breast cancer. BMC Cancer 8:394.

Tjeertes, J.V., Miller, K.M., and Jackson, S.P. (2009). Screen for DNA-damage-responsive histone modifications identifies H3K9Ac and H3K56Ac in human cells. EMBO J *28*, 1878-1889.

Tsurusaki, Y., Okamoto, N., Ohashi, H., Kosho, T., Imai, Y., Hibi-Ko, Y., Kaname, T., Naritomi, K., Kawame, H., Wakui, K., et al. (2012). Mutations affecting components of the SWI/SNF complex cause Coffin-Siris syndrome. Nat. Genet. *44,* 376–378.

Vermeulen, K., Berneman, Z.N., and Van Bockstaele, D.R. (2003). Cell cycle and apoptosis. Cell Prolif *36*, 165–175.

Versteege, I., Sévenet, N., Lange, J., Rousseau-Merck,M.F., Ambros, P., Handgretinger, R., Aurias, A., and Delattre, O. (1998). Truncating mutations of hSNF5/INI1 in aggressive paediatric cancer. Nature *394,* 203–206.

Von Richter, O., Pitarque, M., Rodríguez-Antona, C., Testa, A., Mantovani, R., Oscarson, M., Ingelman-Sundberg, M. (2004). Polymorphic NF-Y dependent regulation of human nicotine C-oxidase (CYP2A6). Pharmacogenetics *14*, 369–379.

Wender, P.A., Mitchell, D.J., Pattabiraman, K., Pelkey, E.T., Steinman, L. and Rothbard, J.B. (2000). The design, synthesis, and evaluation of molecules that enable or enhance cellular uptake: peptoid molecular transporters. Proc. Natl. Acad. Sci. USA *97*, 13003–13008.

Xhemalce, B., Dawson, M. A. and Bannister, A. J. (2011). Histone Modifications. Reviews in Cell Biology and Molecular Medicine.

Xia Hao, X. (1998). Cell penetrating constructs regulate galanin receptor levels and modify

Yamanaka, K., Mizuarai, S., Eguchi, T., Itadani, H., Hirai, H., and Kotani, H. (2009). Expression levels of NF-Y target genes changed by CDKN1B correlate with clinical prognosis in multiple cancers. Genomics *94*,219–227.

Yamanaka, T., Tosaki, A., Kurosawa, M., Matsumoto, G., Koike, M., Uchiyama, Y., Maity, S.N., Shimogori, T., Hattori, N., and Nukina, N. NF-Y inactivation causes atypical neurodegeneration characterized by ubiquitin and p62 accumulation and endoplasmic reticulum disorganization. Nature communications *5*, 3354.

Yamanaka, T., Miyazaki, H., Oyama, F., Kurosawa, M., Washizu, C., Doi, H., and Nukina, N. (2008). Mutant Huntingtin reduces HSP70 expression through the sequestration of NF-Y transcription factor. EMBO J *27*, 827–839.

Yang, C., Bolotin, E., Jiang, T., Sladek, F.M., and Martinez, E. (2007). Prevalence of the initiator over the TATA box in human and yeast genes and identification of DNA motifs enriched in human TATA-less core promoters. Gene *389*, 52–65.

Yang, X.J., and Seto, E. (2008). The Rpd3/Hda1 family of lysine deacetylases: from bacteria and yeast to mice and men. Nat Rev Mol Cell Biol *9*, 206-218.

Yun, J., Chae, H.D., Choi, T.S., Kim, E.H., Bang, Y.J., Chung, J., Choi, K.S., Mantovani, R., and Shin, D.Y. (2003). Cdk2-dependent phosphorylation of the NF-Y transcription factor and its involvement in the p53-p21 signaling pathway. J Biol Chem *278*, 36966–36972.

Zaro, J.L. and Shen, W.C. (2005). Evidence that membrane transduction of oligoarginine does not require vesicle formation. Exp. Cell Res. *307*, 164–173.

Zhu, J., Zhang, Y., Joe, G.J., Pompetti, R., and Emerson, S.G. (*2005).* NF-Ya activates multiple hematopoietic stem cell (HSC) regulatory genes and promotes HSC self-*renewal.* PNAS *102,* 11728-11733.

Zhu, X., Wang, Y., Pi, W., Liu, H., Wickrema, A. and Tuan, D. (2012). NF-Y recruits both transcription activator and repressor to modulate tissue- and developmental stage-specific expression of human gamma-globin gene. PLoS One *7*, e47175.

Zwicker, J., Gross, C., Lucibello, F.C., Truss, M., Ehiert, F., Engeland, K., and Muller, R. (1995). Cell cycle regulation of cdc25C transcription is mediated by the periodic repression of the glutamine rich activators NF-Y and Spi1. Nucleic Acids Research *23*, 3822-3830.

PART II

Accepted paper:

Paolo Benatti[1], Maria Luisa Chiaramonte[2], Mariangela Lorenzo[2], John A Hartley[3], Daniel Hochhauser[3], Nerina Gnesutta[2], Roberto Mantovani[2], Carol Imbriano[1] and Diletta Dolfini[2]*.

NF-Y activates genes of metabolic pathways altered in cancer cells. *Oncotarget*

Oncotarget

MANUSCRIPT HOME | AUTHOR INSTRUCTIONS | REVIEWER INSTRUCTIONS | LOGOUT | JOURNAL HOME

Detailed Status Information

Manuscript #	006396R1
Current Revision #	1
Other Version	006396
Submission Date	2015-10-22 11:06:18
Current Stage	
Title	NF-Y activates genes of metabolic pathways altered in cancer cells.
Manuscript Type	Research Paper
Corresponding Author	Diletta Dolfini (Università degli Studi di Milano)
Contributing Authors	Paolo Benatti , Maria Chiaramonte , Mariangela Lorenzo , John Hartley , Daniel Hochhauser , Nerina Gnesutta , Roberto Mantovani , Carol Imbriano
Abstract	The trimeric transcription factor NF-Y binds to the CCAAT box, an element enriched in promoters of genes overexpressed in tumors. Previous studies on the NF-Y regulome identified the general term metabolism as significantly enriched. We dissect here in detail the targeting of metabolic genes by integrating analysis of NF-Y genomic binding and profilings after inactivation of NF-Y subunits in different cell types. NF-Y controls de novo biosynthetic pathways of lipids, teaming up with the master SREBPs regulators. It activates glycolytic genes, but, surprisingly, is neutral or represses mitochondrial respiratory genes. NF-Y targets the SOCG (Serine, One Carbon, Glycine) and Glutamine pathways, as well as genes involved in the biosynthesis of polyamines and purines. Specific cancer-driving nodes are generally under NF-Y control. Altogether, these data delineate a coherent strategy to promote expression of metabolic genes fuelling anaerobic energy production and other anabolic pathways commonly altered in cancer cells.
Keywords	transcription, cancer metabolism, NF-Y, Glycolysis, SOCG pathway
Conflict of Interest	**No**, there is no conflict of interest that I should disclose, having read the above statement.
Decision	Accept / 2015-11-15

NF-Y activates genes of metabolic pathways altered in cancer cells.

Paolo Benatti[1], Maria Luisa Chiaramonte[2], Mariangela Lorenzo[2], John A Hartley[3], Daniel Hochhauser[3], Nerina Gnesutta[2], Roberto Mantovani[2], Carol Imbriano[1] and Diletta Dolfini[2]*.

1) Dipartimento di Scienze della Vita, Università di Modena e Reggio Emilia, Via Campi 213/D, 41125 Modena, Italy;

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

3) Cancer Research UK Drug-DNA Interactions Research Group, UCL Cancer Institute, Paul O'Gorman Building, University College London, 72 Huntley Street, London WC1E 6BT, UK.

*To whom correspondence should be addressed.

e-mail: diletta.dolfini@unimi.it;

Tel: 39-02-50315004

Fax: 39-02-50315044

Keywords: transcription, cancer metabolism, NF-Y, Glycolysis, SOCG pathway

ABSTRACT

The trimeric transcription factor NF-Y binds to the CCAAT box, an element enriched in promoters of genes overexpressed in tumors. Previous studies on the NF-Y regulome identified the general term *metabolism* as significantly enriched. We dissect here in detail the targeting of metabolic genes by integrating analysis of NF-Y genomic binding and profilings after inactivation of NF-Y subunits in different cell types. NF-Y controls *de novo* biosynthetic pathways of lipids, teaming up with the master SREBPs regulators. It activates glycolytic genes, but, surprisingly, is neutral or represses mitochondrial respiratory genes. NF-Y targets the SOCG (Serine, One Carbon, Glycine) and Glutamine pathways, as well as genes involved in the biosynthesis of polyamines and purines. Specific cancer-driving nodes are generally under NF-Y control. Altogether, these data delineate a coherent strategy to promote expression of metabolic genes fuelling anaerobic energy production and other anabolic pathways commonly altered in cancer cells.

INTRODUCTION

The CCAAT box is an important element present in promoters and enhancers of eukaryotic genes. It is bound by the evolutionarily conserved NF-Y (also named CBF), a trimer formed by NF-YA, NF-YB and NF-YC. NF-YB and NF-YC have histone-like structures, which, upon hetero-dimerization, present a complex surface for NF-YA association, in turn providing sequence-specific CCAAT recognition [1]. The use of genome-wide assays has generalized the concept previously derived by *in vitro* experiments that NF-Y is the primary CCAAT binding protein [2]. Yeast *S. cerevisiae* produces energy and ethanol through glycolysis and fermentation when grown in medium containing glucose; when challenged with non-fermentable carbon sources, yeast cells switch to oxygen-fueled metabolism, by activation of nuclear genes of the mitochondrial respiratory complexes. All these genes contain a CCAAT sequence in their regulatory UAS (Upstream Activating Sequences) sequences and are dependent upon the NF-Y yeast homologue HAP2/3/4/5 [3].

In mammals, the NF-Y regulome is apparently more complex, but it is becoming intelligible, thanks to converging sets of data: (i) the precise biochemical characterization of the target sequence led to the definition of a DNA-binding matrix with high information content, characterized in hundreds mammalian promoters, highlighting a strong positional bias [4]. (ii) Genome-wide experiments confirmed and further extended these observations to enhancers and other genomic regions [5-11]. (iii) Profiling analysis of genes whose expression is affected by functional inactivation of one of the subunits [11-13] found that in addition to a positive role on transcriptional units, NF-Y is also part of repressive mechanisms of transcription.

A common theme in the analysis of the NF-Y regulome in mammalian cells is the presence of the term *metabolism* at the top of Gene Ontology categorizations; functional dissection of individual promoters of metabolic genes indeed indicated the importance of NF-Y for high level of expression. There has been a renewed interest in the transcriptional control of such genes, since specific metabolic pathways are found altered in cancer cells; a vast array of biochemical, genetic and pharmacological data highlight the importance of the expression levels of single genes for "metabolic reprogramming", an hallmark of the development and progression of tumors [14-16]. This is well exemplified by the recent computational evaluation of expression levels of genes of the SOCG -Serine One Carbon Glycine- pathway across cancer samples, indicating that collective overexpression of genes is predictive of the increased flux of metabolites observed in tumors [17]. Interestingly, analysis of large sets of expression profilings comparing tumors and normal tissues indicate that the NF-Y matrix is enriched in promoters of genes overexpressed in cancer cells [18]; however, it was not determined whether these genes belong to specific pathways. Finally, compelling genetic experiments have recently established that NF-YC, with TAF12 and RAD54L, is a driver oncogene of choroid plexus carcinomas [19].

For these reasons, we decided to take a closer look at the metabolic pathways influenced by the transcriptional activity of NF-Y. We analysed available genomic data and performed additional gene expression experiments after inactivation of NF-Y subunits, to rationalize its role in the regulation of metabolic genes. The results point to specific pathways, and within them specific nodes, which are under tight NF-Y control.

RESULTS AND DISCUSSION

Experimental strategy.

Top rank GO terms of NF-Y-regulated genes include metabolic pathways [11-13]. For this reason, we focused specifically on metabolic genes with the following strategy.

(i) We analyzed the results of Affymetrix gene expression profilings of Hela-S3 cells inactivated of NF-YA by shRNA interference [11] and performed additional profilings in epithelial HCT116 and H322 cells under identical conditions of NF-YA inactivation. Fig. S1 shows the levels of NF-YA, as assessed by Western blot analysis; the complete list of genes up- and down-regulated in HCT116 and H322, considering a threshold of 1.3 fold difference, is in Fig. S2. Moreover, we analyzed the recently reported profilings data of mES cells functionally inactivated of individual and of all three subunits of NF-Y [12].

(ii) To validate profilings, we analyzed the expression of selected genes of the affected pathways by qRT-PCR after inactivation of the NF-YB subunit in Hela cells with two different shRNAs.

(iii) To match function with location, we analyzed ChIP-Seq data of NF-YA and NF-YB in Hela-S3, GM12878 and K562 derived from ENCODE [9, 11], as well as the individual binding of the three subunits in mouse ES cells [12]. Specifically, we verified binding to the *bona fide* CCAAT matrix in the prototypical promoter position, between -60 and -100 from the TSS [4], as well as in distal locations classified as enhancers by virtue of the presence of appropriate epigenetic marks (H3K27ac, H3K4me1).

The lack of NF-Y functional dependence in the presence of promoter binding might be due to an effective CCAAT-independence of the gene -and activity of compensatory TFs- or to the long half-life of the specific mRNA: the latter is a particularly relevant point, as

we consider here housekeeping genes involved in basic metabolic functions, whose mRNAs tend to be relatively stable.

Overall density of NF-Y regulated genes in metabolic pathways.

The profilings and ChIP-Seq data of Hela-S3 cells were first analyzed within the KEGG global metabolic chart. Genes whose expression is decreased after NF-YA inactivation are in red and those increased are in green (Fig. 1A); genes with core promoters bound by NF-YA are in blue (Fig. 1B). The charts of gene expression profilings of human H322 and HCT116 cells after NF-YA, and of mouse ES cells after NF-YA, NF-YB and NF-YC inactivation, are shown in Figs. S3 and S4. Genes bound by NF-YB in Hela-S3, showing marginal differences with NF-YA, are in Fig. S5. Similarly, NF-YA and NF-YB binding in K562 and GM12878, as well as NF-YA, NF-YB and NF-YC in mES cells are shown in Figs. S6-8. A bird's eye view of the results collectively indicates high densities in areas of lipids, carbohydrates and nucleotides metabolisms. Other locations -glycan, vitamins, terpenoids and xenobiotics- are mostly depleted, with only selected genes positive. Aminoacids pathways show density in specific areas. We calculated the statistical significance of the regulated pathways in profilings of Hela-S3 and mES cells using KOBAS 2.0 [20], a software providing a better definition for the different metabolic sub-pathways. This led to the identification of globally enriched terms (Fig. S9). As previously reported [11], the Hela-S3 pathways at the top of the list are related to *cell-cycle* and *mitosis*, but specific metabolic terms are enriched and they are indeed predominant in mES cells. Focusing specifically on metabolic terms (Table 1), we note that anabolic genes are less expressed after NF-Y removal, whereas catabolic ones tend to be increased; among the formers, *aminoacids* -notably Ala, Asp, Glu; Ser, Gly; Gln-

lipids and *nucleic acids* are enriched. As for carbohydrate, *carbon metabolism* -mostly glycolysis- is enriched. We hereafter further examine the role of NF-Y in such pathways.

Lipids metabolism.

NF-Y activates genes of the two main branches of lipids metabolism, cholesterol and fatty acids (Fig. 2A): inactivation leads to decreased expression of most cholesterol genes, with the exception of HMGCS2, GGPS1, IDI1/2, HSD17B7, SC5D (Fig. 2A). A similar situation is observed for the fatty acids branch. With the exception of HMGCS2 and PMVK, all promoters are bound *in vivo* in most cell types, including those negatively regulated in mES cells. Rate limiting enzymes of cholesterol -HMGCS1, HMGCR, Squalene Synthase (SQLE), DHCR - and fatty acids pathways -Acetyl COA Carboxylase (ACACA), Fatty Acids Synthase (FASN), Steatoryl CoA Desaturase (SCD)- were previously reported to be under NF-Y positive control [21-36]. Acyl-CoA lyase (ACLY) and Acetyl-CoA synthase of the short chains (ACSS2) are also controlled.
Sustained proliferation rates of cells require a high level of *de novo* biosynthesis of lipids, which are incorporated into newly formed cell membranes. Indeed, these pathways are crucial for cancer cells, and many genes are overexpressed [37]. What is particularly relevant is that the genes more sensitive to NF-Y-inactivation are also those implicated in cancer development. On the cholesterol side, the mevalonate genes are overexpressed in many cancers, including HMGCS1, indeed a cancer-driving gene [38], and the target of statins, a promising “new” class of anti-cancer drugs [39]. In the fatty acids branch, NF-Y targets FASN, SCD1 and GPAM, frequently overexpressed in tumors; FASN is a driver of neoplastic transformation [40]. ACSS2 is essential for transformation of acetate to Acetyl-CoA, thereafter used for lipids synthesis, as well as for histone acetylations [41,42]; expression of ACSS2 is high in many types of tumors and KO mice are protected

in mouse models of liver carginogenesis. The parallel strong NF-Y positivity of the less studied ACSM3 and ACSL3 invites evaluation of their expression levels across cancer samples, and role in tumorigenesis. Finally, ACLY is at the crossroad of lipids and glucose metabolism, mediating the production of Acetyl-CoA by transformation of cytoplasmic citrate to oxaloacetate. ACLY deregulation is generally observed in cancer, often correlating with tumor stage and prognosis [43] and silencing blocks proliferation of cancer cells by different mechanisms, impairing lipid biosynthesis, glycolytic enzymes, citrate accumulation and histone acetylation.

SREBP-1 and SREBP-2 are the master TFs of cholesterol and fatty acids biosynthesis [44]; the SREBP-1 gene itself is controlled by NF-Y [45,46]. Genome-wide experiments reported a significant overlap between NF-YA and SREBP-1 sites [7], and enrichment of NF-Y sites in SREBP-1 peaks in mouse liver cells [47]. ChIP-Seq data of SREBP-1/2 in HepG2 and GM12878 were recently reported by the ENCODE consortium: inspection of binding to lipidogenic genes indicates a perfect correlation with NF-Y in cholesterol genes, and near perfect in fatty acids genes (SREBP1/2 positive promoters are marked in grey in Fig. 2A). We thus decided to analyze the global genomic overlap of NF-Y and SREBP-1/2: to do so, we used the PScan-ChIP software [48], which allows the identification of Transcription Factor Binding Site (TFBS) present in the JASPAR or TRANSFAC databases, as enriched in close proximity -<75 bps- of genomic peaks of a given TF. The statistical enrichment of positively correlating TFBSs is measured as "global" or "local", depending on the robustness of the overlap (Fig. S10). The results of such analysis for SREBP-1 and SREBP-2 are shown in Fig. 2B: the NF-Y consensus present in JASPAR -termed NFYA or NFYB- is significantly enriched in SREBP-1 and SREBP-2 peaks in HepG2 cells treated with pravastatin, but not in untreated GM12878 (Fig. S10); similarly, SREBP-1 peaks are enriched in CCAAT boxes in HepG2 cells

treated with insulin. For SREBPs in pravastatin treated HepG2, the program signals "global" enrichments -p values of 10^{-20}- indicating that the CCAAT box is a "primary" binding site. By comparison, the SREBF motif has similar p values, indicating that the majority of SREBP-1 genomic sites in HepG2 have both a canonical SRE and a CCAAT box. No other TFBS is significantly enriched. The data of SREBP-1 peaks of insulin-treated HepG2 are very similar, except that RFX motifs also emerge, with similar frequencies (Fig. 2B). To validate these results, we used the MEME tools for *de novo* motif discovery, to inspect SREBP-1/2 genomic peaks: we indeed retrieved logos identical to the NF-Y binding site -the ATTGG reverse of CCAAT- in addition to the expected SRE sequence, and, in insulin-treated HepG2, the RFX logo (Fig. 2B). Thus, two types of analysis indicate that NF-Y and SREBPs share a vast set of genomic locations, extending beyond lipidogenic genes, at least in liver cells. Intriguingly, this is not observed in untreated GM12878 B cells, possibly suggesting treatment- or cell-type specific regulation.

Mechanistically, NF-Y and SREBPs are believed to serve different functions in transcriptional activation: the former as a "pioneer" TF, which binds to promoters and predispose a positive chromatin environment [11,12,49,50]; SREBPs as "activator" TFs promoting transcriptional elongation. In fact, nuclear translocation and DNA-binding of SREBPs are visible under specific conditions: typically, this requires relocation from the ER into the nucleus, following addition of PTMs and proteolysis, according to specific stimuli [44,47].

Respiratory genes, Glycolysis and the TCA cycle.

The NF-Y homologue HAP2/3/5 is the master transcriptional activator of all genes of the respiratory chain in yeast, following the shift from fermentable (glucose) to non-

fermentable (lactate) carbon sources [3]. The promoters of two human respiratory chain genes -ATP Synthase β/ATP5B and CYC1- were originally shown to rely on CCAAT boxes [51-54], and it could be inferred that NF-Y globally controls mammalian respiratory genes as well. The mitochondrial electron transport chain is composed of five complexes (I-V), each containing multiple subunits encoded by nuclear and mitochondrial genes. Surprisingly, we notice differential densities, both in genomic binding and transcriptional activation, in the areas of oxygen-fueled (low) and carbohydrate (high) metabolisms (Fig. 1); this is matched by KOBAS analysis, with terms such as *aerobic respiration* and *oxidative phosphorylation* being up-regulated upon NF-YA inactivation, hence normally repressed (Table 1). Further analysis of the genes of the 5 respiratory complexes indicates the following: CYC1 (complex III) is -modestly- down-regulated in NF-Y-inactivated cells, but not bound by NF-Y; ATP5B (complex V) is bound and regulated; in general, however, only a handful genes -NDUFA2, NDUFS8, ATP5E, ATP5G1 and SDHB- are directly activated in cancer cells, eight genes in mES cells (Fig. 3A). A larger set of genes is repressed by NF-Y, either through direct binding, or indirectly. We verified the expression analysis by inactivation of NF-YB through shRNA interference of Hela cells: of seven representative genes analyzed by qRT-PCR, either bound or not by NF-Y *in vivo*, four showed negligible changes, ATP6AP was modestly increased and CYC1 was modestly decreased, in accordance with the profilings data (Fig. 3B). Experiments using a second shRNA directed against NF-YB yielded similar results (Fig. S11). In summary, NF-Y activation of respiratory genes is globally an exception, and expression of most genes are either not controlled, or repressed.

On the other hand, glycolytic genes are densely populated of NF-Y-activated units (Figs. 1 and 4A). The sensitivity of such genes to NF-Y inactivation is very high in mES cells;

PKFB2 and SLC16A7 are the only genes showing up-regulation. Critical genes such as PFKFB, PGK, GAPDH, PKM, LDH are bound and heavily rely on NF-Y activity. Note that Aldolase B, one of the first NF-Y promoters dissected [55,56], is negative, possibly because it is mostly expressed in liver cells, not assayed here. Regulation is similar in cancer cells, with differences in SLC2A3, SLC16A7, PFKFB2/3 (in HCT116 and H322), PKM and HKII. Note that HKII and PKM promoters were shown to be NF-Y-dependent [57-59]. We verified the expression of some of these genes in NF-YB-inactivated Hela cells by qRT-PCR: most are decreased, with the exception of HKI, HKII and PKM (Fig. 4B and Fig. S11). We also checked the protein levels of several glycolytic enzymes, and found HKI, HKII, GAPDH and LDHA to be decreased, whereas PFKP was unchanged (Fig. 4C). These data are in agreement with the mRNA analysis, with two exceptions: HKII, whose protein is decreased and mRNA is increased after NF-YA inactivation, and PKM, showing no change at the protein level. The differential effects could be due to the long half-life of the proteins, which would require assessment at longer times after NF-Y removal, or to the previously reported differences in sets of genes regulated by NF-YA and NF-YB inactivation [13].

Downstream of pyruvate, Acetyl-CoA enters the TCA cycle, which is functionally linked to mitochondrial oxidative phosphorylation: at the level of binding, the density of NF-Y sites in promoters is high, but the regulation is apparently complex and, intriguingly, somewhat dissimilar in cancer and mES cells (Fig. S12A). The Pyruvate Dehydrogenase -PDH- complex, leading to Acetyl-CoA production by oxidation of pyruvate, illustrates such complexity: PDHA1 is affected, but not the DLD and DLAT subunits, which are, if anything, repressed by NF-Y; at the same time, the regulatory PD Kinase -PDK1/4- which inhibits the activity of the PDH complex, is robustly activated by NF-Y, as shown

before [60]. Parallel pathways feed into oxalacetate from (i) Phosphoenol-pyruvate *via* PCK (PEPCK), positively controlled by NF-Y, notably PCK2; (ii) Pyruvate Carboxylase (PC), which is bound -promoter and enhancer- and negatively regulated (Fig. S12A). The control of PC by NF-Y is also complex [61], including positive effects upon removal of single repressive CCAAT boxes [62]. Note the PC route is crucial for anaplerosis, that is, replenishment of the pools of the TCA cycle intermediates [63]. Production of citrate from Acetyl-CoA and oxaloacetate, and further to isocitrate, is mediated by Citrate synthase and Aconitase, whose promoters are bound *in vivo*, but modestly affected. Thereafter, enzymes are not uniformly regulated: Succinate Dehydrogenases (SDH) and Malate Dehydrogenase (MDH) are repressed by NF-Y in cancer cells (Fig. S12A). For IDH1/2, OGDH and FH, there is less significant regulation in the different cancer cell types analyzed, and a positive –indirect- effect on IDH2 in mES cells. We assayed some of the TCA genes by qRT-PCR after NF-YB inactivation in Hela cells: consistent with the profilings, PDHB, PDK1, ACLY but not PDHA1, are decreased, whereas OGDH, IDH2 and DLD are unchanged (Fig. S12B). In the second set of experiments using shNF-YB-2, the data are similar, except that OGDH is decreased, and PDHB is not, as in the profilings, (Fig. S11). The protein levels of PDHA were evaluated in Western Blots and found to be unchanged (Fig. S12C). Note that NF-Y appears to promote expression of enzymes converting α-ketoglutarate and malate out of the TCA cycle (See below). In summary, NF-Y does bind to most promoters of TCA cycle genes, but its role is less relevant, and possibly cell-type specific; in general, it does not uniformly favor the progression of Acetyl-CoA derived from glucose, lipids and aminoacids catabolism through the oxidative production of energy, particularly in cancer cells. Again, this is in striking contrast to the results obtained in yeast, where the CIT1, IDH, FH and MH genes

are robustly activated by the HAP2/3/4/5 complex during activation of the respiratory response [64].

A major and unexpected finding of our analysis is that the control of energy metabolism by NF-Y changed during evolution, from promoting mitochondrial energy production, in *fungi*, to activating genes of anaerobic pathways, in mammals. The preference for anaerobic energy production by glycolysis in the presence of oxygen is a hallmark of cancer cells, a phenomenon known as the Warburg effect [65]. This requires a high activity of glycolytic enzymes; accordingly, the overall levels of many of them are elevated in many types of tumors, with respect to normal cells, mainly because of increased mRNA expression [66].

In general, NF-Y appears to have a coherent strategy of promoting glucose, but not oxygen utilization. Among glycolytic enzymes, GAPDH produces Glyceraldeyde-3P, which enters the biosynthetic pathway of Serine biosynthesis (See below). PKM promotes the final, rate-limiting step of glycolysis to generate pyruvate and ATP and is important in cancer cells [67]. Lactate Dehydrogenases are key to drive regeneration of NAD^+ and continuous ATP production from glycolysis. Lactate Dehydrogenase is produced by two genes -LDHA and LDHB- whose relative expression varies depending upon the cellular status: in cancer cells, it is shifted toward LDHA, generating the LDH5 enzyme, most efficient in driving the Pyruvate-to-Lactate conversion required to keep the glycolytic flux flowing [68]. It is thus noteworthy that NF-Y is more important for LDHA expression in cancer cells, and for LDHB in mESCs. At the end of glycolysis, NF-Y does not appear to promote a flux of Acetyl-CoA entering the TCA cycle: this is exemplified by the modest impact on PDH, on the associated subunits, and activation of its repressor PDK. At the

same time, NF-Y promotes exit of α-ketoglutarate and malate from the TCA cycle to enter alternate pathways, in cancer cells at least (See below).

Long time considered as housekeeping genes subject to modest levels of regulation, the promoters of these genes recently regained interest, but information about the DNA elements and TFs acting on many of them remains scarcely abundant. HIF-1α, MYC, STAT3 and p53 are TFs associated with the regulation of glycolytic enzymes [69-74]. The formers are generally activators, while p53 is inhibitory, at least in the wt configuration. HIF-1α is activated by anaerobic stress and constitutively active in many cancer cells, targeting essentially all glycolytic genes; MYC was shown to interact with NF-Y directly [75] and the peaks of MYC and NF-Y overlap significantly in ENCODE data [11]. Interestingly, p53 was shown to promote gluconeogenesis [76], while oncogenic Gain-of-Function mutations of p53, reported to impact on expression of NF-Y genes [77,78], mediate metabolic reprogramming and increased glycolysis [79]: it will be interesting to evaluate the effects of mutp53 on the metabolic genes targeted by NF-Y identified here.

The SOCG and Glutamine pathways.

We noticed high densities of NF-Y targets in specific pathways of Serine/Glycine, Glutamine and Methionine, all enriched in KOBAS analysis (Table I). The SOCG pathway is central for *de novo* production of Serine and Glycine, and for additional metabolisms -folates, nucleotides- crucially required for growing cells [80]. NF-Y binding is present on all promoters of the pathway in at least one cell type with the exception of the SLC19A1 transporter, MTHFD2L and AMT (Fig. 5A). Serine synthesis stems from a branch of glycolysis, through Glyceraldeyde 3P being sequentially metabolized by three enzymes: PHGDH, PSAT and PSHP; the promoters are all bound, and the mRNAs

decreased upon NF-YA removal. For the PHGDH promoter, genetic evidence of the importance of the CCAAT box was reported [81]. Conversion to Glycine is then obtained through the activity of SHMT, present in two isoforms, one located in the cytoplasm -SHMT1- the other -SHMT2- in mitochondria: NF-YA regulates the former, but apparently not the latter. qRT-PCR analysis of Hela cells inactivated with the two shRNAs against NF-YB confirmed the control of PHGDH, PSAT1, PSPH, whereas SHMT1 was, if anything, positively affected (Fig. 5B and Fig. S11). Interestingly, NF-Y also impacts on the mitochondrial branch, by activation of two key enzymes, MTHFDL1 and MTHFD2. Control of MTHFR and DHFR signals a further role on folates metabolism, which is linked to Methionine synthesis and nucleotides production (See below). In summary, the data indicate that NF-Y plays a crucial role in the activation of SOCG genes.

Glutamine is another pathway dense with NF-Y targets, at the level of promoter recognition and function (Fig. 5A). Binding of NF-Y is absent only in SLC3A2, CAD, SLC38A5 and SLC38A3; the latter gene, GLUD1 and GPT are unchanged in expression after NF-YA-inactivation in cancer cells. Most genes are activated by NF-Y through direct promoter binding [82]. Specifically, GLS and GLS2 are differentially regulated in Hela and mES cells: the former is bound and activated in Hela (but repressed in H322), but not regulated in mES cells. The reciprocal is true for GLS2. These enzymes are crucial for the conversion of Glutamine to Glutamate, which can then be transformed to α-ketoglutarate by GLUD1, not regulated by NF-Y in cancer cells and modestly repressed in mES cells. This branch is a second leg of anaplerotic mechanisms filling the TCA cycle with metabolites from the Glutamine pathway: the positivity of GOT1/2, but not GLUD1, is yet another indication that NF-Y is not favoring anaplerosis *per se*. GOT1/2, in fact, convert Aspartate from Glutamine into oxaloacetate, then converted into malate [83]:

thereafter, the Malic enzymes ME1 ME2, both regulated by NF-Y (Fig. 5A), mediate pyruvate production and exit from TCA. In summary, NF-Y appears to promote a shift of Glutamine catabolism from the TCA cycle into alternative biosynthetic pathways, matching the modest role exerted on PDH (Fig. S12A). The negative regulation of SDH (Fig. 3A and S12A) and the activation of genes deviating metabolites from the TCA cycle are a further indication that NF-Y is not promoting oxygen-mediated production of energy.

The Glutamine and SOCG pathways are altered in cancer cells. As to the former, tumors usually become addicted to high levels of Glutamine. The key GLS2 is activated by p53 [84,85], and regulation is shared by p63 and p73, with a different outcome depending from the isoforms [86,87]. Genes of the SOCG pathways are overexpressed in cancer [88], the global alteration directly impinges on an altered flux of metabolites [17]. PHGDH and SHMT2 are predictive of survival outcomes in breast cancer, PSPH of hepatocellular carcinomas [89,90]. SHMT2 is a cancer driver gene [91,92]; PHGDH overexpression, resulting from genomic amplification, is essential for growth of certain human breast cancers [93] and melanomas [94]; PSAT1 is overexpressed in colon cancer [95]. In general, there are very few data on the mechanisms of transcriptional regulation of SOCG pathway genes: p53 is a repressor of PHGDH [96], whereas cMYC overexpression leads to increased biosynthesis of serine [90]. The identification of NF-Y/CCAAT as a pivotal element should facilitate the identification of neighboring elements and TFs involved in regulation of these units.

Polyamine and purine metabolism.

The biosynthetic pathways of purines and polyamines have high density of NF-Y-controlled genes (Fig. 1 and Table I). In the polyamine pathway, which is extremely

important for cancer cells (97), there are two rate-limiting steps, mediated by ODC1, from the Urea cycle, and AMD1, from the Methionine salvage pathway. AMD1 is directly controlled by NF-Y in cancer and mES cells, but NF-Y inactivation also leads to an indirect decrease of ODC1 (Fig. 6A). Additional regulated genes are ADI1, AGMAT, GOT1, SMOX, SMS and SRM. ADI1, AMD1, GOT1 and SRM were tested in Hela cells inactivated of NF-YB: GOT 1 was unaffected, ADI1 was substantially decreased, AMD1 and SRM decreased with one of the shRNAs used (Fig. 6B and Figure S11).
AMD1 and ODC1 are deregulated in a number of cancers, notably prostate adenocarcinomas and MYCN-amplified neuroblastomas: in the latter system, ODC1 is required for tumor formation and overexpression predicts patient survival [98]. ODC1 pharmacological inhibition leads to normalization of several pathways -NMYC levels, LIN28/Let-7 expression and glycolysis- altered in neuroblastomas, and it is currently tested in clinical trials [99]. Interestingly, ODC1 and AMD1 are independently involved in mESC self-renewal, in assays of LIF-deprived cells, and can increase reprogramming of differentiated fibroblasts by OCT4, SOX2 and KLF4 [100]. NF-Y is also involved in mESC self-renewal [12,101,102]: a positive role on expression of the three TFs was reported, as an important part of the cohort of NF-Y targets involved in stem cells maintenance.
Many genes of the purine pathway are under NF-Y control (Fig. S13); two targets are *bona fide* tumor suppressors: MTAP, frequently deleted in many types of cancers along with the neighboring CDKN2A [103], and Adenylate Kinase 2 AK2 [104]. MTAP shows differential regulation by NF-Y, repressed in cancer cells and activated in mES cells. As for the synthesis of dNTPs required for the DNA biosynthetic pathways (DNA replication and repair), the key step is the reduction from ribonucleotides to desossiribonucleotides, performed by the rate-limiting enzyme Ribonucleotide Reductase (RnR), an

heterotetramer composed of two large (RRM1) and two small (RRM2) subunits. While RRM1 is abundant and constant throughout the cell-cycle, the limiting RRM2 is transcribed only in S-phase [105]. RRM1 is regulated by NF-Y in Hela cells, but the dependence is higher for RRM2 (Fig. S13), whose promoter contains three functionally crucial CCAAT boxes [106], conserved and important also in Zebrafish [107]. Recently, RRM2 has attracted considerable therapeutic interest, because its targeting leads cancer cells to senesce [108,109]. It is important to note that cells inactivated of NF-YA are crippled in S-phase progression, develop signs of DNA-damage, with subsequent triggering of an apoptotic response [12,13]: it is possible that a decrease in the dNTPs pools, due to a decrease in the conversion from NTPs as a consequence of the RRM2 drop, might be at least partially responsible for this behavior.

Conclusions and perspectives.

The detailed analysis of metabolic genes under the control of NF-Y described here and summarized in Fig. 7 opens several future avenues of investigations. (i) A thorough phylogenetic analysis of the conservation of CCAAT boxes in metabolic genes targeted by NF-Y should be performed; specifically, genes involved in energy production should be examined, to identify when the switch from aerobic -in yeast- to anaerobic -in mammals- energy production has occurred during evolution. (ii) Further studies should be undertaken on the many NF-Y metabolic targets whose role in cancer development is unknown. Systematic investigation of their expression levels in large cohorts of gene expression data of cancer cells could point at additional regulatory nodes to be further examined experimentally by genetic and biochemical means. (iii) A few TFs known to impact on transcription of metabolic genes emerged, and some have been associated to NF-Y activity; in this regard, the SREBP analysis performed here is illustrative of what

can be done, given the availability of large datasets of TFs genomic sites; the identification of major NF-Y partners in the regulation of metabolic genes will shed light on the arrangement of promoter sites and reciprocal interplay. Such catalogue will help our understanding of regulatory circuits of cancer cells.

ACKNOWLEDGEMENTS

We thank V. Santoro, K. Kiakos, F. Vena, G. Liccardi and R. Vuononvirta at UCL Cancer Institute.

FUNDING

PB was a recipient of an EMBO Short Term fellowship and a recipient of a Fondazione Italiana Ricerca Cancro (*FIRC*) fellowship. The work is supported by AIRC grants 14130 to RM and 14210 to CI, and by the Cancer Research UK Programme Grant (C2259/A16569) to DH.

CONFLICT OF INTEREST STATEMENT

The authors declare that there are no conflicts of interest

MATERIALS AND METHODS

Lentiviral knockdown and gene expression arrays.

Scrambled control (shSC), NF-YA (shNF-YA) and two NF-YB (shNF-YB-1 -gctatgtctactttaggcttt-; shNF-YB-2 -ccaaagaatgtgttcaagaatc-) shRNAs were cloned into pLKO.1 vector (Sigma Aldrich), and viral production and transduction were carried out as previously described [13]. H322 and HCT116 cells were transduced with shSC or shNF-YA viral supernatants, in triplicate, and cells collected after 72 hrs of incubation. Hela cells were infected with shSC or shNF-YB or shNF-YA and collected at 72 hrs after infection. In the experiments shown in Figure S11, transient DNA transfections with shSC and shNF-YB-2 were performed in triplicate with Lipofectamine 2000 (Invitrogen 11668027). Cells were collected after 72 hours. Knockdown efficiency was assayed by PCR on cDNAs and by Western Blots on whole cell protein extracts using anti-NF-YA, anti NF-YB and anti-Actin antibodies. Total RNA was prepared by Trizol extraction and retrotranscribed with Iscript cDNA Synthesis kit (BIORAD 170-8890). For arrays, RNA was prepared according to Affymetrix standard protocol and hybridized to Hu-Gene 2.0 expression arrays.

Gene expression analyses.

For HeLa cells, raw data were retrieved from GSE40215 [11]; for mouse embryonic stem cells, raw data were retrieved from GSE56840 [12]. For gene expression analysis of HCT116 and H322 using the Affymetrix platform, biological triplicates of control and shNF-YA-treated cells were independently processed: normalization (rma), quality controlled, probe set filtered, identification of differentially expressed probe sets and annotation of those probe sets to gene symbols were performed using Bioconductor packages (Affy and Limma). We defined upregulated and downregulated genes when the fold change is above 1.3 and FDR

<0.05. Gene Ontologies and pathways analyses were performed using the KOBAS 2.0 tool with default settings [20]. Raw data have been deposited in GEO Repository under the accession number GSE70543.

Immunoblots.

For Western Blot analysis, NF-Y-inactivated and control cells were lysed in lysis buffer (50mM Tris–HCl pH 8.0, 120mM NaCl, 1% Triton X100, 20%SDS, 1mM EDTA, protease and phosphatase inhibitors). Equivalent amount of extracts were run on SDS-PAGE, transferred to nitrocellulose membrane (Whatman) and immunoblotted with the following antibodies: anti-Actin (sc-1616, Santa Cruz) anti-NF-YA (Mab1a), anti-NF-YB (Pab001, GeneSpin), anti-Glycolytic enzymes (Glycolysis 2 kit, Cell Signaling 8337S), anti-Vinculin (SAB4200080, Sigma Aldrich).

qRT-PCR.

The *Primer3 0.4.0.* program was used for primers design using default parameters. See Fig. S14 for primer sequences. qPCR was performed with a Biorad myIQ instrument: values are normalized over an internal control ribosomal gene used as normalizator -RPS20- and are represented as fold-enrichment over control sample shSC. Data are presented as mean ± standard error of fold change of 3 biological replicates run in qPCR triplicates. Statistical significance are assessed with one sample t-test and indicated with asterisk when $p<0.05$.

REFERENCES

1. Nardini M, Gnesutta N, Donati G, Gatta R, Forni C, Fossati A, Vonrhein C, Moras D, Romier C, Bolognesi M, Mantovani R. Sequence-specific transcription factor NF-Y displays histone-like DNA binding and H2B-like ubiquitination. Cell 2013; 152:132-143.

2. Dolfini D, Mantovani R. Targeting the Y/CCAAT box in cancer: YB-1 (YBX1) or NF-Y? Cell Death Differ 2013; 20:676-85.

3. Buschlen S, Amillet JM, Guiard B, Fournier A, Marcireau C, Bolotin-Fukuhara M. The *S. Cerevisiae* HAP complex, a key regulator of mitochondrial function, coordinates nuclear and mitochondrial gene expression. Comp Funct Genomics 2003; 4:37-46.

4. Dolfini D, Zambelli F, Pavesi G, Mantovani R. A perspective of promoter architecture from the CCAAT box. Cell Cycle 2009; 8: 4127-37.

5. Testa A, Donati G, Yan P, Romani F, Huang TH, Viganò MA, Mantovani R. Chromatin immunoprecipitation (ChIP) on chip experiments uncover a widespread distribution of NF-Y binding CCAAT sites outside of core promoters. J Biol Chem 2005; 280:13606-15.

6. Ceribelli M, Dolfini D, Merico D, Gatta R, Viganò AM, Pavesi G, Mantovani R. The histone-like NF-Y is a bifunctional transcription factor. Mol Cell Biol 2008; 28:2047-58.

7. Reed BD, Charos AE, Szekely AM, Weissman SM, Snyder M. (2008). Genome-wide occupancy of SREBP1 and its partners NFY and SP1 reveals novel functional roles and combinatorial regulation of distinct classes of genes. PLoS Genet. 4:e1000133.

8. Tiwari VK, Stadler MB, Wirbelauer C, Paro R, Schubeler D, Beisel C. A chromatin-modifying function of JNK during stem cell differentiation. Nat Genet 2011; 44:94-100.

9. Wang J, Zhuang J, Iyer S, Lin X, Whitfield TW, Greven MC, Pierce BG, Dong X, Kundaje A, Cheng Y, Rando OJ, Birney E, Myers RM, et al. Sequence features and chromatin structure around the genomic regions bound by 119 human transcription factors. Genome Res 2012; 22:1798-812.

10. Wang J, Zhuang J, Iyer S, Lin XY, Greven MC, Kim BH, Moore J, Pierce BG, Dong X, Virgil D, Birney E, Hung JH, Weng Z. Factorbook.org: a Wiki-based database for transcription factor-binding data generated by the ENCODE consortium. Nucleic Acids Res 2013; 41:D171-6.

11. Fleming JD, Pavesi G, Benatti P, Imbriano C, Mantovani R, Struhl K. NF-Y co-associates with FOS at promoters, enhancers, repetitive elements, and inactive chromatin regions, and is stereo-positioned with growth-controlling transcription factors. Genome Res 2013; 23:1195-209.

12. Oldfield AJ, Yang P, Conway AE, Cinghu S, Freudenberg JM, Yellaboina S, Jothi R. Histone-fold domain protein NF-Y promotes chromatin accessibility for cell type-specific master transcription factors. Mol Cell 2014; 55:708-22.

13. Benatti P, Dolfini D, Viganò 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.

14. DeBerardinis RJ, Lum JJ, Hatzivassiliou G, Thompson CB. The biology of cancer: metabolic reprogramming fuels cell growth and proliferation. Cell Metab 2008; 7:11–20.

15. Kroemer G, Pouyssegur J. Tumor cell metabolism: cancer's Achilles' heel. Cancer Cell 2008; 13:472–482.
16. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. Cell 2011; 144: 646–674.
17. Mehrmohamadi M, Liu X, Shestov AA, Locasale JW. Characterization of the usage of the serine metabolic network in human cancer. Cell Rep 2014; 9:1507-19.
18. Goodarzi H, Elemento O, Tavazoie S. Revealing global regulatory perturbations across human cancers. Mol Cell 2009; 36:900-11.
19. Tong Y, Merino D, Nimmervoll B, Gupta K, Wang YD, Finkelstein D, Dalton J, Ellison DW, Ma X, Zhang J, Malkin D, Gilbertson RJ. Cross-Species Genomics Identifies TAF12, NFYC, and RAD54L as Choroid Plexus Carcinoma Oncogenes. Cancer Cell 2015; 27:712-27.
20. Xie C, Mao X, Huang J, Ding Y, Wu J, Dong S, Kong L, Gao G, Li CY, Wei L. KOBAS 2.0: a web server for annotation and identification of enriched pathways and diseases. Nucleic Acids Res 2011; 39:W316-22.
21. Jackson SM, Ericsson J, Osborne TF, Edwards PA. NF-Y has a novel role in sterol-dependent transcription of two cholesterogenic genes. J Biol Chem 1995; 270:21445-8.
22. Jackson SM, Ericsson J, Mantovani R, Edwards PA. Synergistic activation of transcription by nuclear factor Y and sterol regulatory element binding protein. J Lipid Res 1998; 39:767-76.
23. Inoue J, Sato R, Maeda M. Multiple DNA elements for sterol regulatory element-binding protein and NF-Y are responsible for sterol-regulated transcription of the genes for human 3-hydroxy-3-methylglutaryl coenzyme A synthase and squalene synthase. J Biochem 1998; 123:1191-8.

24. Bennett MK, Osborne TF. Nutrient regulation of gene expression by the sterol regulatory element binding proteins: increased recruitment of gene-specific coregulatory factors and selective hyperacetylation of histone H3 in vivo. Proc Natl Acad Sci USA 2000; 97:6340-4.

25. Sato R, Okamoto A, Inoue J, Miyamoto W, Sakai Y, Emoto N, Shimano H, Maeda M. Transcriptional regulation of the ATP citrate-lyase gene by sterol regulatory element-binding proteins. J Biol Chem 2000; 275:12497-502.

26. Misawa K, Horiba T, Arimura N, Hirano Y, Inoue J, Emoto N, Shimano H, Shimizu M, Sato R. Sterol regulatory element-binding protein-2 interacts with hepatocyte nuclear factor-4 to enhance sterol isomerase gene expression in hepatocytes. J Biol Chem 2003; 278:36176-82.

27. Schiavoni G, Bennati AM, Castelli M, Fazia MA, Beccari T, Servillo G, Roberti R. Activation of TM7SF2 promoter by SREBP-2 depends on a new sterol regulatory element, a GC-box, and an inverted CCAAT-box. Biochim Biophys Acta 2010; 1801:587-92.

28. Kim JH, Lee JN, Paik YK. Cholesterol biosynthesis from lanosterol. A concerted role for Sp1 and NF-Y-binding sites for sterol-mediated regulation of rat 7-dehydrocholesterol reductase gene expression. J Biol Chem 2001; 276:18153-60.

29. Zerenturk EJ, Sharpe LJ, Brown AJ. Sterols regulate 3β-hydroxysterol Δ24-reductase (DHCR24) via dual sterol regulatory elements: cooperative induction of key enzymes in lipid synthesis by Sterol Regulatory Element Binding Proteins. Biochim Biophys Acta 2012; 1821:1350-60.

30. Prabhu AV, Sharpe LJ, Brown AJ. The sterol-based transcriptional control of human 7-dehydrocholesterol reductase (DHCR7): Evidence of a cooperative regulatory program in cholesterol synthesis. Biochim Biophys Acta 2014; 1842:1431-39.

31. Shi X, Metges CC, Seyfert HM. Interaction of C/EBP-beta and NF-Y factors constrains activity levels of the nutritionally controlled promoter IA expressing the acetyl-CoA carboxylase-alpha gene in cattle. BMC Mol Biol. 2012; 13:21.
32. Xiong S, Chirala SS, Wakil SJ. Sterol regulation of human fatty acid synthase promoter I requires nuclear factor-Y- and Sp-1-binding sites. Proc Natl Acad Sci USA 2000; 97:3948-53.
33. Teran-Garcia M, Rufo C, Nakamura MT, Osborne TF, Clarke SD. NF-Y involvement in the polyunsaturated fat inhibition of fatty acid synthase gene transcription. Biochem Biophys Res Commun 2002; 290:1295-99.
34. Lefevre P, Tripon E, Plumelet C, Douaire M, Diot C. Effects of polyunsaturated fatty acids and clofibrate on chicken stearoyl-coA desaturase 1 gene expression. Biochem Biophys Res Commun 2001; 280:25-31.
35. Schweizer M, Roder K, Zhang L, Wolf SS. Transcription factors acting on the promoter of the rat fatty acid synthase gene. Biochem Soc Trans 2002; 30:1070-72.
36. Mauvoisin D, Rocque G, Arfa O, Radenne A, Boissier P, Mounier C. Role of the PI3-kinase/mTor pathway in the regulation of the stearoyl CoA desaturase (SCD1) gene expression by insulin in liver. J Cell Commun Signal 2007; 1:113-25.
37. Currie E, Schulze A, Zechner R, Walther TC, Farese RV Jr. Cellular fatty acid metabolism and cancer. Cell Metab 2013; 18:153-61.
38. Clendening JW, Pandyra A, Boutros PC, El Ghamrasni S, Khosravi F, Trentin GA, Martirosyan A, Hakem A, Hakem R, Jurisica I, Penn LZ. Dysregulation of the mevalonate pathway promotes transformation. Proc Natl Acad Sci USA 2010; 107:15051-6.
39. Clendening JW, Penn LZ. Targeting tumor cell metabolism with statins. Oncogene 2012; 31:4967-78.

40. Menendez JA, Lupu R. Fatty acid synthase and the lipogenic phenotype in cancer pathogenesis. Nat Rev Cancer 2007; 7:763-77.

41. Comerford SA, Huang Z, Du X, Wang Y, Cai L, Witkiewicz AK, Walters H, Tantawy MN, Fu A, Manning HC, Horton JD, Hammer RE, McKnight SL, Tu BP. Acetate dependence of tumors. Cell 2014; 159:1591-602.

42. Mashimo T, Pichumani K, Vemireddy V, Hatanpaa KJ, Singh DK, Sirasanagandla S, Nannepaga S, Piccirillo SG, Kovacs Z, Foong C, Huang Z, Barnett S, Mickey BE, et al. Acetate is a bioenergetic substrate for human glioblastoma and brain metastases. Cell 2014; 159:1603-14.

43. Zaidi N, Swinnen JV, Smans K. ATP-citrate lyase: a key player in cancer metabolism. Cancer Res 2012; 72:3709-14.

44. Jeon TI, Osborne TF. SREBPs: metabolic integrators in physiology and metabolism. Trends Endocrinol Metab 2012; 23:65-72.

45. Amemiya-Kudo M, Shimano H, Yoshikawa T, Yahagi N, Hasty AH, Okazaki H, Tamura Y, Shionoiri F, Iizuka Y, Ohashi K, Osuga J, Harada K, Gotoda T, et al. Promoter analysis of the mouse sterol regulatory element-binding protein-1c gene. J Biol Chem 2000; 275:31078-85.

46. Deng X, Cagen LM, Wilcox HG, Park EA, Raghow R, Elam MB. Regulation of the rat SREBP-1c promoter in primary rat hepatocytes. Biochem Biophys Res Commun. 2002; 290:256-62.

47. Gilardi F, Migliavacca E, Naldi A, Baruchet M, Canella D, Le Martelot G, Guex N, Desvergne B; CycliX Consortium. Genome-wide analysis of SREBP1 activity around the clock reveals its combined dependency on nutrient and circadian signals. PLoS Genet 2014; 10:e1004155.

48. Zambelli F, Pesole G, Pavesi G. PscanChIP: Finding over-represented transcription factor-binding site motifs and their correlations in sequences from ChIP-Seq experiments. Nucleic Acids Res 2013; 41:W535-43.

49. Li Q, Herrler M, Landsberger N, Kaludov N, Ogryzko VV, Nakatani Y, Wolffe AP. Xenopus NF-Y pre-sets chromatin to potentiate p300 and acetylation-responsive transcription from the Xenopus hsp70 promoter in vivo. EMBO J 1998; 17:6300-15.

50. Sherwood RI, Hashimoto T, O'Donnell CW, Lewis S, Barkal AA, van Hoff JP, Karun V, Jaakkola T, Gifford DK. Discovery of directional and nondirectional pioneer transcription factors by modeling DNase profile magnitude and shape. Nat Biotechnol 2014; 32:171-78.

51. Ohta S, Tomura H, Matsuda K, Kagawa Y. Gene structure of the human mitochondrial adenosine triphosphate synthase beta subunit. J Biol Chem 1988; 263:11257-62.

52. Neckelmann N, Warner CK, Chung A, Kudoh J, Minoshima S, Fukuyama R, Maekawa M, Shimizu Y, Shimizu N, Liu JD, et al. The human ATP synthase beta subunit gene: sequence analysis, chromosome assignment, and differential expression. Genomics 1989; 5:829-43.

53. Suzuki H, Hosokawa Y, Toda H, Nishikimi M, Ozawa T. Common protein-binding sites in the 5'-flanking regions of human genes for cytochrome c1 and ubiquinone-binding protein. J Biol Chem 1990; 265:8159-63.

54. Dyer MR, Walker JE. Sequences of members of the human gene family for the c subunit of mitochondrial ATP synthase. Biochem J 1993; 293:51-64.

55. Tsutsumi K, Ito K, Ishikawa K. Developmental appearance of transcription factors that regulate liver-specific expression of the aldolase B gene. Mol Cell Biol 1989; 9:4923-31.

56. Raymondjean M, Pichard AL, Gregori C, Ginot F, Kahn A. Interplay of an original combination of factors: C/EBP, NFY, HNF3, and HNF1 in the rat aldolase B gene promoter. Nucleic Acids Res 1991; 19:6145-53.

57. Yamada K, Tanaka T, Miyamoto K, Noguchi T. Sp family members and nuclear factor-Y cooperatively stimulate transcription from the rat pyruvate kinase M gene distal promoter region via their direct interactions. J Biol Chem 2000; 275:18129-37.

58. Osawa H, Robey RB, Printz RL, Granner DK. Identification and characterization of basal and cyclic AMP response elements in the promoter of the rat hexokinase II gene. J Biol Chem. 1996; 271:17296-303.

59. Lee MG, Pedersen PL. Glucose metabolism in cancer: importance of transcription factor-DNA interactions within a short segment of the proximal region og the type II hexokinase promoter. J Biol Chem 2003; 278:41047-58.

60. Kwon HS, Huang B, Ho Jeoung N, Wu P, Steussy CN, Harris RA. Retinoic acids and trichostatin A (TSA), a histone deacetylase inhibitor, induce human pyruvate dehydrogenase kinase 4 (PDK4) gene expression. Biochim Biophys Acta 2006; 1759:141-51.

61. Thonpho A, Rojvirat P, Jitrapakdee S, MacDonald MJ. Characterization of the distal promoter of the human pyruvate carboxylase gene in pancreatic beta cells. PLoS One 2013; 8:e55139.

62. Sunyakumthorn P, Boonsaen T, Boonsaeng V, Wallace JC, Jitrapakdee S. Involvement of specific proteins (Sp1/Sp3) and nuclear factor Y in basal transcription of the distal promoter of the rat pyruvate carboxylase gene in beta-cells. Biochem Biophys Res Commun 2005; 329:188-96.

63. Mullen AR, DeBerardinis RJ. Genetically-defined metabolic reprogramming in cancer. Trends Endocrinol Metab 2012; 23:552-59.

64. Fendt SM, Sauer U. Transcriptional regulation of respiration in yeast metabolizing differently repressive carbon substrates. BMC Syst Biol 2010; 4:12

65. Warburg O. On the origin of cancer cells. Science 1956; 123:309-14.

66. Altenberg B, Greulich KO. Genes of glycolysis are ubiquitously overexpressed in 24 cancer classes. Genomics 2004; 84:1014-20.

67. Luo W, Semenza GL. Emerging roles of PKM2 in cell metabolism and cancer progression. Trends Endocrinol Metab 2012; 23:560-66.

68. Augoff K, Hryniewicz-Jankowska A, Tabola R. Lactate dehydrogenase 5: An old friend and a new hope in the war on cancer. Cancer Lett 2014; 358:1-7.

69. Semenza GL. HIF-1 mediates metabolic responses to intratumoral hypoxia and oncogenic mutations. J Clin Invest 2013; 123:3664-71.

70. Maddocks OD, Vousden KH. Metabolic regulation by p53. J Mol Med 2011; 89:237-45.

71. Wahlström T, Arsenian Henriksson M. Impact of MYC in regulation of tumor cell metabolism. Biochim Biophys Acta 2014; 1849:563-69.

72. Soga T. Cancer metabolism: key players in metabolic reprogramming. Cancer Sci 2013; 104:275-81.

73. Chen JQ, Russo J. Dysregulation of glucose transport, glycolysis, TCA cycle and glutaminolysis by oncogenes and tumor suppressors in cancer cells. Biochim Biophys Acta 2012; 1826:370-84.

74. Demaria M, Poli V. PKM2, STAT3 and HIF-1α: The Warburg's vicious circle. JAKSTAT 2012; 1:194-6.

75. Izumi H, Molander C, Penn LZ, Ishisaki A, Kohno K, Funa K. Mechanism for the transcriptional repression by c-Myc on PDGF beta-receptor. J Cell Sci 2001; 114:1533-44.

76. Goldstein I, Yizhak K, Madar S, Goldfinger N, Ruppin E, Rotter V. p53 promotes the expression of gluconeogenesis-related genes and enhances hepatic glucose production. Cancer Metab 2013; 1:9.
77. Di Agostino S, Strano S, Emiliozzi V, Zerbini V, Mottolese M, Sacchi A, Blandino G, Piaggio G. Gain of function of mutant p53: the mutant p53/NF-Y protein complex reveals an aberrant transcriptional mechanism of cell cycle regulation. Cancer Cell 2006; 10:191-202.
78. Weissmueller S, Manchado E, Saborowski M, Morris JP 4th, Wagenblast E, Davis CA, Moon SH, Pfister NT, Tschaharganeh DF, Kitzing T, Aust D, Markert EK, Wu J, et al. Mutant p53 drives pancreatic cancer metastasis through cell-autonomous PDGF receptor β signaling. Cell 2014; 157:382-94.
79. Zhang C, Liu J, Liang Y, Wu R, Zhao Y, Hong X, Lin M, Yu H, Liu L, Levine AJ, Hu W, Feng Z. Tumour-associated mutant p53 drives the Warburg effect. Nat Commun 2013; 4:2935.
80. Locasale JW. Serine, glycine and one-carbon units: cancer metabolism in full circle. Nat Rev Cancer 2013; 13:572-83.
81. Jun DY, Park HS, Lee JY, Baek JY, Park HK, Fukui K, Kim YH. Positive regulation of promoter activity of human 3-phosphoglycerate dehydrogenase (PHGDH) gene is mediated by transcription factors Sp1 and NF-Y. Gene 2008; 414:106-14.
82. Pérez-Gómez C, Matés JM, Gómez-Fabre PM, del Castillo-Olivares A, Alonso FJ, Márquez J. Genomic organization and transcriptional analysis of the human l-glutaminase gene. Biochem J 2003; 370:771-84.
83. Son J, Lyssiotis CA, Ying H, Wang X, Hua S, Ligorio M, Perera RM, Ferrone CR, Mullarky E, Shyh-Chang N, Kang Y, Fleming JB, Bardeesy N, et al. Glutamine

supports pancreatic cancer growth through a KRAS-regulated metabolic pathway. Nature 2013; 496:101-5.

84. Suzuki S, Tanaka T, Poyurovsky MV, Nagano H, Mayama T, Ohkubo S, Lokshin M, Hosokawa H, Nakayama T, Suzuki Y, Sugano S, Sato E, Nagao T, et al. Phosphate-activated glutaminase (GLS2), a p53-inducible regulator of glutamine metabolism and reactive oxygen species. Proc Natl Acad Sci USA 2010; 107:7461-66.
85. Hu W, Zhang C, Wu R, Sun Y, Levine A, Feng Z. Glutaminase 2, a novel p53 target gene regulating energy metabolism and antioxidant function. Proc Natl Acad Sci USA 2010; 107:7455-60.
86. Giacobbe A, Bongiorno-Borbone L, Bernassola F, Terrinoni A, Markert EK, Levine AJ, Feng Z, Agostini M, Zolla L, Agrò AF, Notterman DA, Melino G, Peschiaroli A. p63 regulates glutaminase 2 expression. Cell Cycle 2013; 12:1395-405.
87. Amelio I, Markert EK, Rufini A, Antonov AV, Sayan BS, Tucci P, Agostini M, Mineo TC, Levine AJ, Melino G. p73 regulates serine biosynthesis in cancer. Oncogene 2014; 33:5039-46.
88. Amelio I, Cutruzzolá F, Antonov A, Agostini M, Melino G. Serine and glycine metabolism in cancer. Trends Biochem Sci 2014; 39:191-98.
89. Antonov A, Agostini M, Morello M, Minieri M, Melino G, Amelio I. Bioinformatics analysis of the serine and glycine pathway in cancer cells. Oncotarget 2014; 5:11004-13.
90. Sun L, Song L, Wan Q, Wu G, Li X, Wang Y, Wang J, Liu Z, Zhong X, He X, Shen S, Pan X, Li A, Wang Y, Gao P, Tang H, Zhang H. cMyc-mediated activation of serine biosynthesis pathway is critical for cancer progression under nutrient deprivation conditions. Cell Res 2015; 25:429-44.

91. Lee GY, Haverty PM, Li L, Kljavin NM, Bourgon R, Lee J, Stern H, Modrusan Z, Seshagiri S, Zhang Z, Davis D, Stokoe D, Settleman J, de Sauvage FJ, Neve RM. Comparative oncogenomics identifies PSMB4 and SHMT2 as potential cancer driver genes. Cancer Res 2014; 74:3114-26.
92. Kim D, Fiske BP, Birsoy K, Freinkman E, Kami K, Possemato RL, Chudnovsky Y, Pacold ME, Chen WW, Cantor JR, Shelton LM, Gui DY, Kwon M, et al. SHMT2 drives glioma cell survival in ischaemia but imposes a dependence on glycine clearance. Nature 2015; 520:363-67.
93. Possemato R, Marks KM, Shaul YD, Pacold ME, Kim D, Birsoy K, Sethumadhavan S, Woo HK, Jang HG, Jha AK, Chen WW, Barrett FG, Stransky N, et al. Functional genomics reveal that the serine synthesis pathway is essential in breast cancer. Nature 2011; 476:346-50.
94. Locasale JW, Grassian AR, Melman T, Lyssiotis CA, Mattaini KR, Bass AJ, Heffron G, Metallo CM, Muranen T, Sharfi H, Sasaki AT, Anastasiou D, Mullarky E, et al. Phosphoglycerate dehydrogenase diverts glycolytic flux and contributes to oncogenesis. Nat Genet 2011; 43:869-74.
95. Vié N, Copois V, Bascoul-Mollevi C, Denis V, Bec N, Robert B, Fraslon C, Conseiller E, Molina F, Larroque C, Martineau P, Del Rio M, Gongora C. Overexpression of phosphoserine aminotransferase PSAT1 stimulates cell growth and increases chemoresistance of colon cancer cells. Mol Cancer 2008; 7:14.
96. Ou Y, Wang SJ, Jiang L, Zheng B, Gu W. p53 Protein-mediated regulation of phosphoglycerate dehydrogenase (PHGDH) is crucial for the apoptotic response upon serine starvation. J Biol Chem 2015; 290:457-66.
97. Casero RA Jr, Marton LJ. Targeting polyamine metabolism and function in cancer and other hyperproliferative diseases. Nat Rev Drug Discov 2007; 6:373-90.

98. Hogarty MD, Norris MD, Davis K, Liu X, Evageliou NF, Hayes CS, Pawel B, Guo R, Zhao H, Sekyere E, Keating J, Thomas W, Cheng NC, et al. ODC1 is a critical determinant of MYCN oncogenesis and a therapeutic target in neuroblastoma Cancer Res 2008; 68:9735-45.

99. Lozier AM, Rich ME, Grawe AP, Peck AS, Zhao P, Chang AT, Bond JP, Sholler GS. Targeting ornithine decarboxylase reverses the LIN28/Let-7 axis and inhibits glycolytic metabolism in neuroblastoma. Oncotarget 2015; 6:196-206.

100. Zhao T, Goh KJ, Ng HH, Vardy LA. A role for polyamine regulators in ESC self-renewal. Cell Cycle 2012; 11:4517-23.

101. Grskovic M, Chaivorapol C, Gaspar-Maia A, Li H, Ramalho-Santos M. Systematic identification of cis-regulatory sequences active in mouse and human embryonic stem cells.PLoS Genet 2007; 3:e145.

102. Dolfini D, Minuzzo M, Pavesi G, Mantovani R. Stem Cells. The short isoform of NF-YA belongs to the embryonic stem cell transcription factor circuitry. 2012; 30:2450-9.

103. Bertino JR, Waud WR, Parker WB, Lubin M. Targeting tumors that lack methylthioadenosine phosphorylase (MTAP) activity: current strategies. Cancer Biol Ther 2011; 11:627-32.

104. Kim H, Lee HJ, Oh Y, Choi SG, Hong SH, Kim HJ, Lee SY, Choi JW, Su Hwang D, Kim KS, Kim HJ, Zhang J, Youn HJ, et al. The DUSP26 phosphatase activator adenylate kinase 2 regulates FADD phosphorylation and cell growth. Nat Commun 2014; 5:3351.

105. Filatov D, Thelander L. Role of a proximal NF-Y binding promoter element in S phase-specific expression of mouse ribonucleotide reductase R2 gene. J Biol Chem 1995; 270:25239-43.

106. Liu X, Zhou B, Xue L, Qiu W, Shih J, Zheng S, Yen Y. Nuclear factor Y regulation and promoter transactivation of human ribonucleotide reductase subunit M2 gene in a Gemcitabine resistant KB clone. Biochem Pharmacol 2004; 67:1499-511.

107. Shang H, Li Q, Feng G, Cui Z.Identification and characterization of alternative promoters, transcripts and protein isoforms of zebrafish R2 gene. PLoS One 2011; 6:e24089.

108. Aird KM, Zhang G, Li H, Tu Z, Bitler BG, Garipov A, Wu H, Wei Z, Wagner SN, Herlyn M, Zhang R. Suppression of nucleotide metabolism underlies the establishment and maintenance of oncogene-induced senescence. Cell Rep 2013; 3:1252-65.

109. Aird KM, Li H, Xin F, Konstantinopoulos PA, Zhang R. Identification of ribonucleotide reductase M2 as a potential target for pro-senescence therapy in epithelial ovarian cancer. Cell Cycle 2014; 13:199-207.

FIGURE LEGENDS

Figure 1. NF-Y targets in metabolic pathways.

Global map of metabolic pathways targeted by NF-YA. A) Metabolic genes upregulated (in green) and downregulated (in red) after silencing of NF-YA in HeLa cells [11]. B) Metabolic genes with core promoters bound by NF-YA are indicated in blue according to ENCODE ChIP-Seq data. The maps are constructed with the KEGG Mapper v 2.5 tool.

Figure 2. NF-Y activates genes of lipid metabolism.

A) Genes involved in cholesterol and fatty acids metabolisms are shown. The heatmap represents the log2 fold change of relative expression derived from profiling analysis (See colour scale) in the indicated cell lines after inactivation of NF-Y subunits: H322 (This manuscript, Fig S2), HCT116 (This manuscript, Fig. S2), HeLaS3 [11] and mESC [12]. The presence of NF-Y binding in ENCODE datasets is indicated with P (core promoter binding), E (external enhancer), Ei (enhancer in gene body) or R (repetitive sequence within 5 kb from the TSS). *In vivo* binding of SREB-1/2 according to ChIP-Seq experiments is indicated by a grey background. B) Analyses of over-represented motifs in SREBPs peaks in HepG2 ChIP-Seq data analyzed by ENCODE. The matrices were derived with Pscan-ChIP (Left panel) and MEME (Right panel), and the relative p-values are shown.

Figure 3. NF-Y and expression of respiratory genes.

A) Genes of oxidative phosphorylation chain complexes are shown with the relative expression levels after inactivation of NF-Y in different cell lines, and the relative

presence of NF-Y binding in P (promoter), E (external enhancer), Ei (enhancer in gene body) or R (repetitive sequence within 5 kb from TSS). B) qRT-PCR evaluation of expression levels of respiratory genes after inactivation of NF-YB in Hela cells. The average -/+ SD of three biological replicates is represented. (*p<0.05)

Figure 4. Glycolytic enzymes are regulated by NF-Y.

A) Genes encoding for glycolytic enzymes are shown with their expression levels after inactivation of NF-Y in different cell lines, and the presence of NF-Y binding is indicated as in Fig. 3. B) qRT-PCR evaluation of expression levels of selected genes after inactivation of NF-YB in Hela cells. The average -/+ SD of three biological replicates is represented. (*p<0.05). C) Western blot analysis of protein levels of NF-Y targets in Hela cells inactivated of NF-YA (Left panel), or NF-YB (Right panel).

Figure 5. Activation of SOCG genes by NF-Y.

A) Genes of the SOCG (Serine, One Carbon, Glycine) pathway are shown with the expression levels after inactivation of NF-Y, and the presence of NF-Y binding as in Fig. 3. B) qRT-PCR evaluation of expression levels of PHGDH, PSAT1, PSPH, SHMT1 after inactivation of NF-YB in Hela cells. The average -/+ SD of three biological replicates is represented (*p<0.05).

Figure 6. Activation of genes of the polyamine metabolism by NF-Y.

Genes of the polyamine pathway are shown with the expression levels after inactivation of NF-Y, and the presence of *in vivo* NF-Y binding as in Figs 3-5.

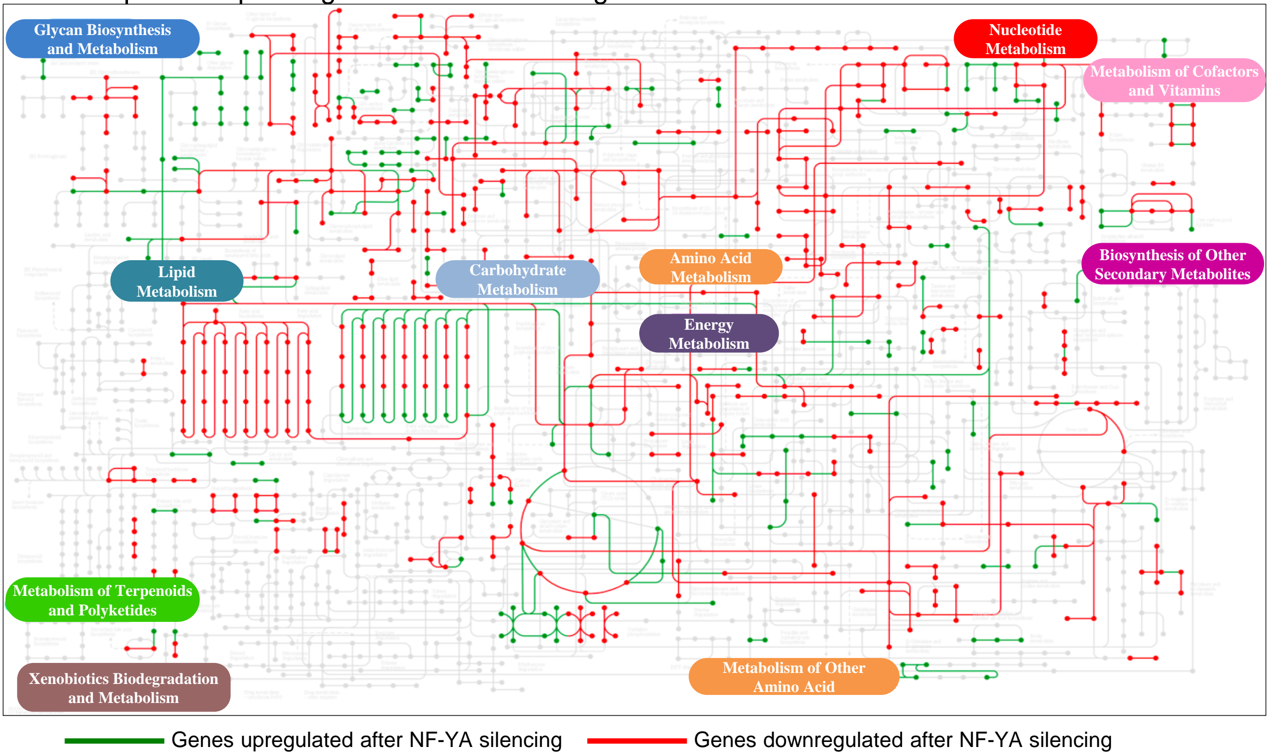
Figure 7. Schematic representation of metabolic pathways regulated by NF-Y**.**

Table 1. Metabolic terms enriched in genes whose expression is changed upon inactivation of NF-YA in Hela-S3 cells (Left panel) and in mouse ES cells (mESC) (Right panel). Pathways and Gene Ontology analyses were performed with Kobas 2.0 and metabolic terms extracted from the list. The full list of terms enriched in differentially expressed genes is shown in Fig. S9.

	HeLa-S3			mESC			
	CATEGORY	*ID*	*P-Value*	*CATEGORY*	*ID*	*P-Value*	
DOWNREGULATED	Alpha-amino acid biosynthetic process	GO:1901607	3.73E-04	Carboxylic acid metabolic process	GO:0019752	7.86E-16	DOWNREGULATED
	Cellular amino acid biosynthetic process	GO:0008652	7.16E-04	Metabolism	REACT_188937	3.20E-13	
	Alanine, aspartate and glutamate metabolism	hsa00250	7.93E-04	Small molecule metabolic process	GO:0044281	2.51E-12	
	Serine glycine biosynthesis	P02776	1.48E-03	Metabolic pathways	mmu01100	5.73E-11	
	Glutamine metabolic process	GO:0006541	2.86E-03	Cellular amino acid metabolic process	GO:0006520	4.98E-10	
	Amino acid synthesis and interconversion (transamination)	REACT_238	3.39E-03	Small molecule biosynthetic process	GO:0044283	6.69E-10	
	Metabolism of polyamines	REACT_14820	3.39E-03	Carbon metabolism	mmu01200	8.45E-08	
	Glutamine family amino acid metabolic process	GO:0009064	4.02E-03	Oxidation-reduction process	GO:0055114	1.65E-07	
	Polyamine metabolic process	GO:0006595	4.48E-03	Biosynthesis of amino acids	mmu01230	2.76E-07	
	Glycemic traits (pregnancy)	NHGRI GWAS	4.49E-03	Metabolism of amino acids and derivatives	REACT_232845	1.63E-06	
	L-serine metabolic process	GO:0006563	7.32E-03	Amino acid synthesis and interconversion (transamination)	REACT_239550	4.43E-06	
	Purine nucleobase biosynthetic process	GO:0009113	1.01E-02	Glycolysis / Gluconeogenesis	mmu00010	6.44E-06	
	Nucleotide catabolic process	GO:0009166	1.05E-02	Glycine, serine and threonine metabolism	mmu00260	1.87E-05	
	Fatty acid biosynthesis	hsa00061	1.94E-02	Glucose metabolism	REACT_261601	3.03E-05	
	Purine metabolism	hsa00230	2.92E-02	Fatty acid biosynthetic process	GO:0006633	3.70E-05	
				Cholesterol metabolic process	GO:0008203	6.50E-05	
UPREGULATED	Cellular catabolic process	GO:0044248	2.24E-06	Lipid metabolic process	GO:0006629	6.74E-05	
	Carboxylic acid catabolic process	GO:0046395	7.94E-05	Fructose and mannose metabolism	mmu00051	8.88E-05	
	Fatty acid catabolic process	GO:0009062	1.03E-04	Pyruvate metabolic process	GO:0006090	9.33E-05	
	Lipid oxidation	GO:003440	1.14E-04	Valine, leucine and isoleucine degradation	mmu00280	1.49E-04	
	Autophagy	GO:0006914	1.84E-04	Aspartate family amino acid metabolic process	GO:0009066	3.12E-04	
	Monocarboxylic acid catabolic process	GO:0072329	2.42E-04	Glycogen storage diseases	REACT_203949	4.37E-04	
	Branched-chain amino acid catabolism	REACT_197	1.12E-03	Metabolism of carbohydrates	REACT_248571	4.37E-04	
	Fatty acid beta-oxidation	GO:0006635	1.17E-03	Arginine and proline metabolism	mmu00330	6.05E-04	
	Macroautophagy	GO:0016236	1.93E-03	Activation of gene expression by SREBF (SREBP)	REACT_198969	8.19E-04	
	Protein catabolic process	GO:0030163	2.60E-03				
	Valine, leucine and isoleucine degradation	hsa00280	2.73E-03	Regulation of cellular macromolecule biosynthetic process	GO:2000112	1.10E-05	UPREGULATED
	Macromolecule catabolic process	GO:0009057	2.84E-03	Regulation of cellular biosynthetic process	GO:0010556	3.43E-05	
	Cellular response to starvation	GO:0009267	3.01E-03	Regulation of macromolecule metabolic process	GO:0006355	4.52E-05	
	Ribonucleoside bisphosphate metabolic process	GO:0033875	7.23E-03	Regulation of biosynthetic process	mmu05220	5.70E-05	
	Purine nucleoside bisphosphate metabolic process	GO:0034032	7.23E-03	modification-dependent macromolecule catabolic process	mmu04068	1.59E-04	
	Nucleotide metabolic process	GO:0009117	7.31E-03	MyD88:Mal cascade initiated on plasma membrane	GO:0031326	1.69E-04	
	Aerobic respiration	GO:0009060	1.64E-02	Toll Like Receptor TLR6:TLR2 Cascade	GO:0016070	1.72E-04	
	Nucleotide-sugar metabolic process	GO:0009225	1.67E-02	Ubiquitin-dependent protein catabolic process	GO:0060255	2.14E-04	
	Glycosphingolipid metabolism	REACT_116105	1.68E-02	Modification-dependent protein catabolic process	GO:0009889	2.25E-04	
	Metabolism of lipids and lipoproteins	REACT_22258	2.15E-02	Clathrin derived vesicle budding	GO:1903507	3.17E-04	
	Citrate cycle (TCA cycle)	hsa00020	5.24E-02	Negative regulation of macromolecule metabolic process	REACT_206529	3.33E-04	
	Fatty acid, triacylglycerol, and ketone body metabolism	REACT_22279	6.20E-02	Regulation of lipid metabolism by PPARalpha	REACT_263004	3.58E-04	
	Cytochrome c-mediated apoptotic response	REACT_831	6.91E-02	Signaling by PDGF	GO:0006511	3.78E-04	
	Oxidative phosphorylation	hsa00190	7.55E-02	Hippo signaling pathway	REACT_198602	7.65E-04	

A

Gene expression profiling after NF-YA silencing in HeLa-S3 cells



B

NF-YA binding in HeLa-S3 cells

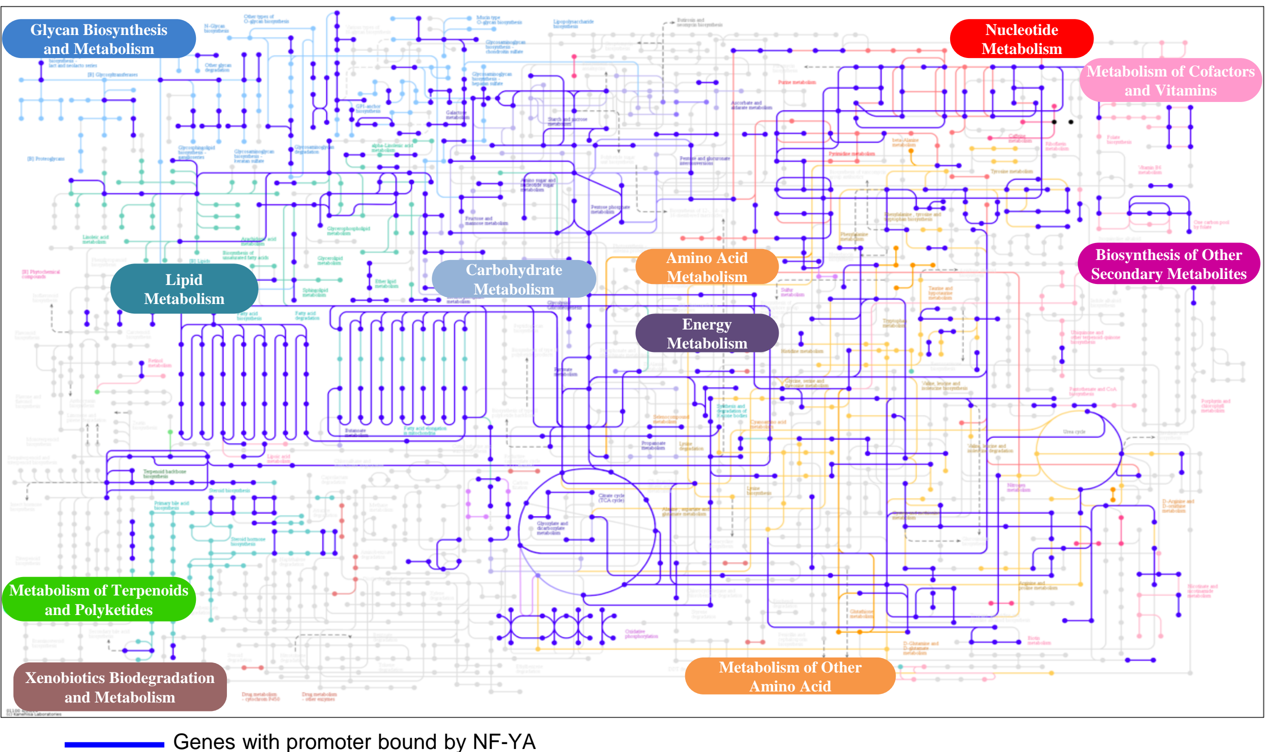
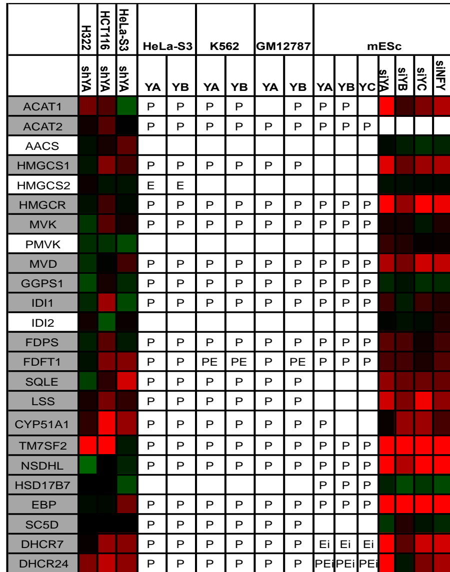
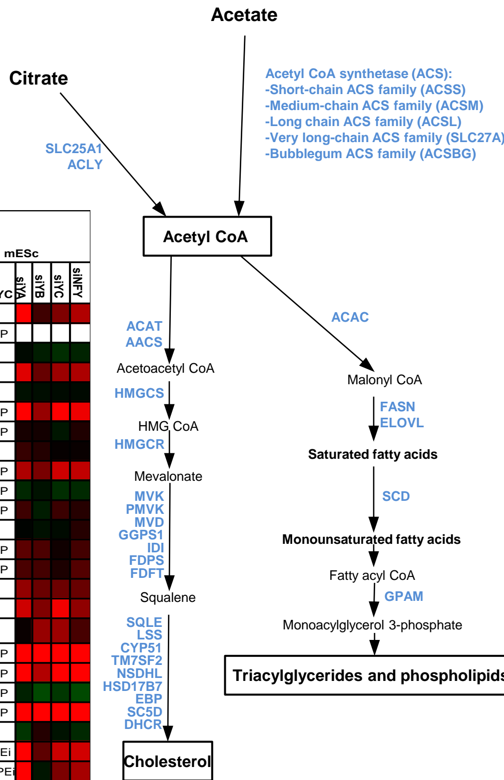
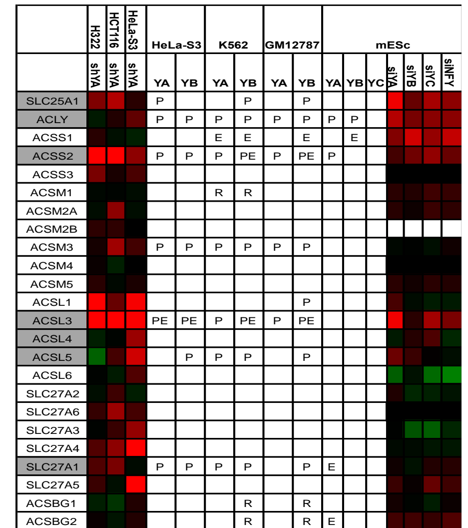


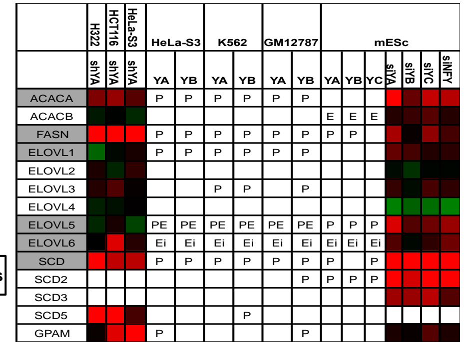
Figure 1

A









B

TF + cell line	NAME	JASPAR-ID	L P-VALUE	G P-VALUE	MOTIF	TF	P-VALUE
	PSCAN-ChIP results				**MEME results**		
SREBP1 HepG2+pravastatin	SREBF1	MA0595.1	5.87E-16	3.36E-23		SREBF1	1,7E-13
	SREBF2	MA0596.1	2.51E-13	9.57E-19			
	NFYA	MA0060.1	7.34E-07	3.51E-18		NFYA	5,3E-05
	NFYB	MA0502.1	1.61E-04	1.43E-14			
SREBP2 HepG2+pravastatin	NFYA	MA0060.1	1.29E-12	7.01E-26		NFYA	3.2E-12
	NFYB	MA0502.1	3.71E-10	4.62E-22			
	SREBF1	MA0595.1	3.96E-17	1.18E-20		SREBF1	3.0E-11
	SREBF2	MA0596.1	1.31E-18	2.47E-18			
SREBP1 HepG2+insulin	NFYA	MA0060.1	6.79E-70	8.42E-153		RFX2	5.1E-114
	RFX2	MA0600.1	2.17E-102	1.87E-129			
	Rfx1	MA0509.1	4.49E-102	3.01E-123		NF-YA	2.9E-112
	SREBF1	MA0595.1	2.72E-70	4.70E-122			
	RFX5	MA0510.1	1.01E-60	3.65E-128		SREBF1	6.0E-72
	NFYB	MA0502.1	3.43E-59	1.90E-118			
	SREBF2	MA0596.1	7.44E-61	2.79E-51		RFX3	3.1E-60

Figure 2

A

COMPLEX I

	H322	HCT116	HeLa-S3	HELA		K562		GM12787		mESc						
	shYA	shYA	shYA	YA	YB	YA	YB	YA	YB	YA	YB	YC	siYA	siYB	siYC	siNFY
NDUFA1									P							
NDUFA10																
NDUFA11				P	P	P	P	P	P	P						
NDUFA12																
NDUFA13																
NDUFA2				P	P	P	P	P	P	P	P	P				
NDUFA3																
NDUFA4																
NDUFA4L2				Ei					Ei							
NDUFA5																
NDUFA6				P	P	P	P	P	P	P		P				
NDUFA7				P												
NDUFA8				P					P							
NDUFA9																
NDUFAB1																
NDUFB1																
NDUFB10							E		E							
NDUFB11				P	P	P	P	P	P							
NDUFB2																
NDUFB3				P	P	P	P	P	P	P		P				
NDUFB4																
NDUFB5				P	P	P	P	P	P							
NDUFB6																
NDUFB7																
NDUFB8																
NDUFB9																
NDUFC1							P		P							
NDUFS1																
NDUFS2																
NDUFS3																
NDUFS4					E		E		E			E				
NDUFS5										P						
NDUFS6																
NDUFS7																
NDUFS8				P	P	P	P	P	P	P	P	P				
NDUFV1				P					P	P	P					
NDUFV2																
NDUFV3				P	P	P	P	P	P	P	P					

COMPLEX III

	H322	HCT116	HeLa-S3	HELA		K562		GM12787		mESc						
	shYA	shYA	shYA	YA	YB	YA	YB	YA	YB	YA	YB	YC	siYA	siYB	siYC	siNFY
CYC1																
UQCRQ										P						
UQCRFS1																
UQCRC2				P	P	P	P	P	P							
UQCRC1																
UQCRB																
UQCR11																
UQCR10				P	P		P		P							

COMPLEX V

	H322	HCT116	HeLa-S3	HELA		K562		GM12787		mESc						
	shYA	shYA	shYA	YA	YB	YA	YB	YA	YB	YA	YB	YC	siYA	siYB	siYC	siNFY
ATP12A																
ATP4A																
ATP4B																
ATP5A1																
ATP5B				P	P	P	P	P	P	P	P	P				
ATP5C1																
ATP5D																
ATP5E				P	P	P	P	P	P	P	P	P				
ATP5F1				P	P	P	P	P	P							
ATP5G1				P	P	P	P	P	P	P	P	P				
ATP5G2				PE	E		E	P	E	E	E	E				
ATP5G3																
ATP5H				P	P	P	P	P	P	P	P	P				
ATP5I																
ATP5J																
ATP5J2																
ATP5L																
ATP5O																
ATP6																
ATP6AP1				P	P	P	P	P	P	P	P	P				
ATP6V0A1										P						
ATP6V0A2				P	P	P	P	P	P							
ATP6V0A4																
ATP6V0B																
ATP6V0C																
ATP6V0D1				P					P							
ATP6V0D2																
ATP6V0E1																
ATP6V0E2							P		P	P						
ATP6V1A																
ATP6V1B1																
ATP6V1B2																
ATP6V1C1																
ATP6V1C2																
ATP6V1D										P						
ATP6V1E1				P	P	P	P	P	P	P	P	P				
ATP6V1E2																
ATP6V1F																
ATP6V1G1				P	P	P	P	P	P							
ATP6V1G2																
ATP6V1G3																
ATP6V1H																
PPA1																
PPA2																
LHPP																
TCIRG1																

COMPLEX II

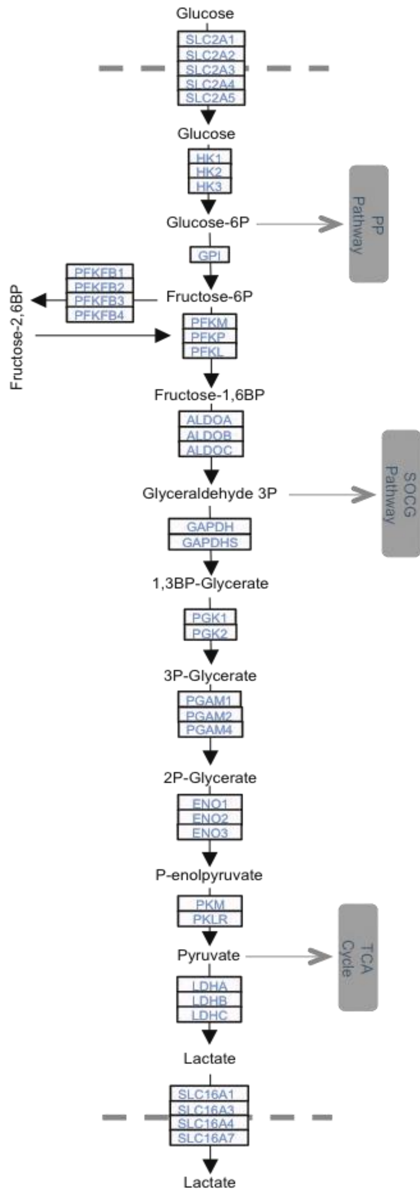
	H322	HCT116	HeLa-S3	HELA		K562		GM12787		mESc						
	shYA	shYA	shYA	YA	YB	YA	YB	YA	YB	YA	YB	YC	siYA	siYB	siYC	siNFY
SDHA																
SDHB				P					P							
SDHC																
SDHD																

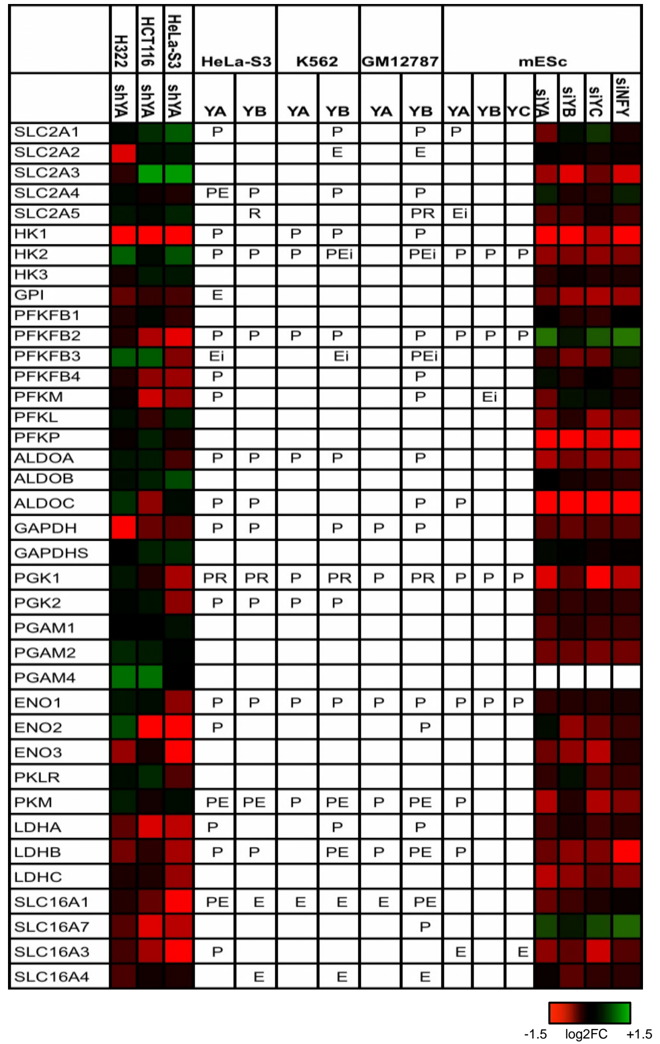
COMPLEX IV

	H322	HCT116	HeLa-S3	HELA		K562		GM12787		mESc						
	shYA	shYA	shYA	YA	YB	YA	YB	YA	YB	YA	YB	YC	siYA	siYB	siYC	siNFY
COX10																
COX11																
COX15																
COX17									P							
COX4I1																
COX4I2																
COX5A									P							
COX5B									P							
COX6A1				P	P		P		P	P		P				
COX6A2																
COX6B1				P	P	P	P	P	P	P						
COX6B2							P			P	P					
COX6C																
COX7A1										P	P					
COX7A2																
COX7A2L				P						E	E					
COX7B																
COX7B2							Ei		Ei							
COX7C																
COX8A				P	P	P	P	P	P	P	P					
COX8C																

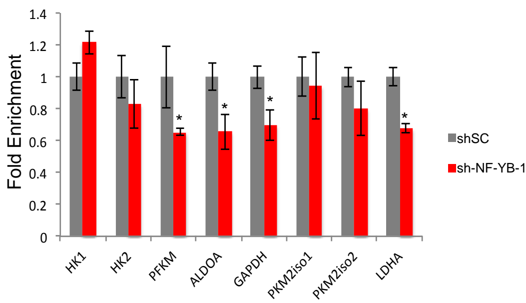
B

Figure 3





	H322	HCT116	HeLa-S3	HeLa-S3		K562		GM12787		mESc						
	shYA	shYA	shYA	YA	YB	YA	YB	YA	YB	YA	YB	YC	siYA	siYB	siYC	siNFY
SLC2A1				P			P		P	P						
SLC2A2							E		E							
SLC2A3																
SLC2A4				PE	P		P		P							
SLC2A5					R				PR	Ei						
HK1				P		P	P		P							
HK2				P	P	P	PEi		PEi	P	P	P				
HK3																
GPI				E												
PFKFB1																
PFKFB2				P	P	P	P		P	P	P	P				
PFKFB3				Ei			Ei		PEi							
PFKFB4				P					P							
PFKM				P					P		Ei					
PFKL																
PFKP																
ALDOA				P	P	P	P		P							
ALDOB																
ALDOC				P	P				P	P						
GAPDH				P	P		P	P	P							
GAPDHS																
PGK1				PR	PR	P	PR	P	PR	P	P	P				
PGK2				P	P	P	P									
PGAM1																
PGAM2																
PGAM4																
ENO1				P	P	P	P	P	P	P	P	P				
ENO2				P					P							
ENO3																
PKLR																
PKM				PE	PE	P	PE	P	PE	P						
LDHA				P			P		P							
LDHB				P	P		PE	P	PE	P						
LDHC																
SLC16A1				PE	E	E	E	E	PE							
SLC16A7									P							
SLC16A3				P						E		E				
SLC16A4					E		E		E							



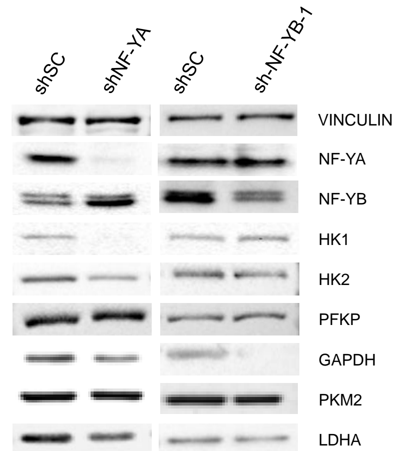
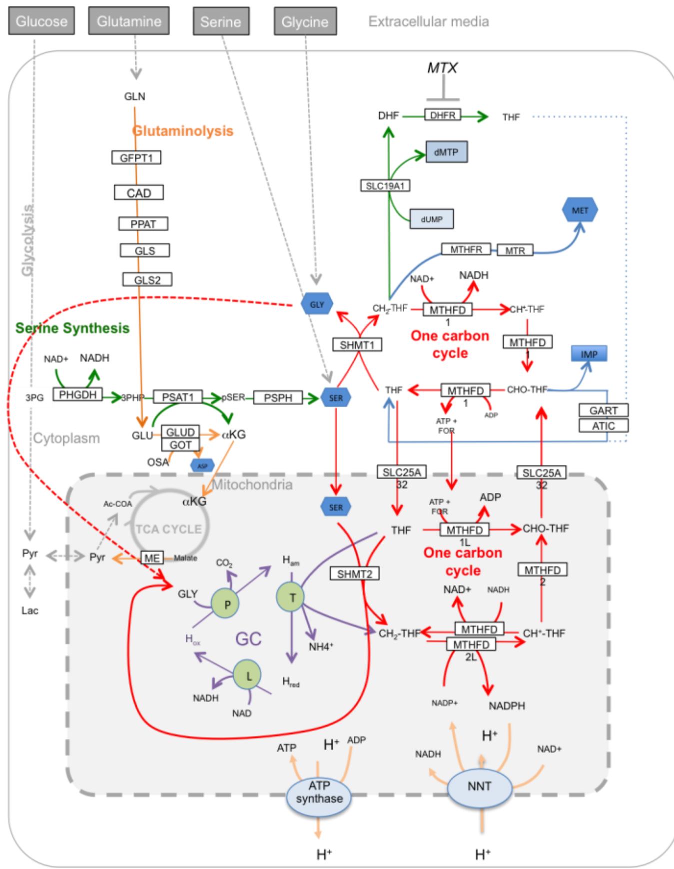


Figure 4

A



SOCG PATHWAY

	H322	HCT116	HeLa-S3	HELA		K562		GM12787		mESc						
	shYA	shYA	shYA	YA	YB	YA	YB	YA	YB	YA	YB	YC	siYA	siYB	siYC	siNFY
PHGDH				PE	PE		P	P	PE							
PSAT1				P	P	P	PE	P	PE	P	P	P				
PSPH				P	P	P	P	P	P	P	P	P				
SHMT1				P	P	P	P	P	P	P	P	P				
SHMT2								P								
MTHFD1				P	P	P	PE	P	PE	P	P	P				
MTHFD1L				E	E	E	E	E	E							
MTHFD2				P	P	P	P	P	P	P	P	P				
MTHFD2L																
MTHFR				P	PE	P	PE	P	PE	P						
DHFR							P		P	P						
SLC19A1																
SLC25A32							P		P	P						
MTR				P	P	P	P	P	P	P		P				
GART				P	P	P	P	P	P							
ATIC										P						
GCSH				P	P	P	P		P	P						
DLD				P	P	P	P	P	P	P						
GLDC									PEi	P	P	P				
AMT																

GLUTAMINE PATHWAY

	H322	HCT116	HeLa-S3	HELA		K562		GM12787		mESc						
	shYA	shYA	shYA	YA	YB	YA	YB	YA	YB	YA	YB	YC	siYA	siYB	siYC	siNFY
SLC1A5				P	P	P	P	P	P	P	P	P				
SLC3A2																
GFPT1				P	P	P	P	P	P							
PFAS									P	E	E	E				
PPAT				P	P		P		P							
CAD																
GLUD1				P	P		P		P	Ei	Ei					
GLS				P	P	P	P	P	P	P	P	P				
GLS2				P	P	P	P	P	P	P	P	P				
GOT1				P	P	P	P	P	P	P	P	P				
GOT2							P		P	P		P				
GPT				P	P		P		P							
ME1					Ei		Ei		Ei							
ME2				P	P	P	P	P	P	PE						
SLC38A1				P			Ei		Ei							
SLC38A2				P												
SLC38A3																
SLC38A5																

-1.5 log2FC +1.5

B

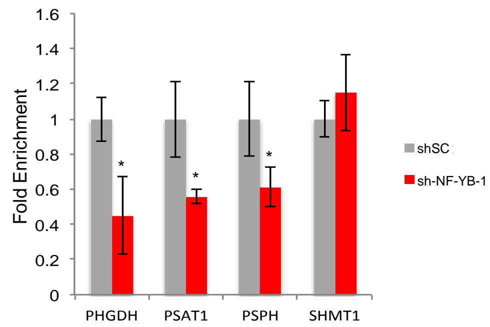
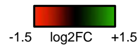


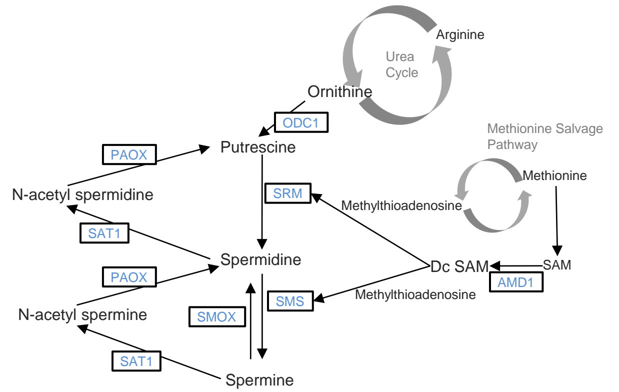
Figure 5

A

POLYAMINE PATHWAY

	H322 shYA	HCT116 shYA	HeLa-S3 shYA	HeLa-S3 YA	HeLa-S3 YB	K562 YA	K562 YB	GM12787 YA	GM12787 YB	mESc YA	mESc YB	mESc YC	mESc siYA	mESc siYB	mESc siYC	mESc siNFY
ARG1																
ADC						P	P		P	P						
ADI1				P			P		P							
AGMAT					E	E	E		E							
AMD1				P	P	P	P	P	P	P	P					
APIP					E	E	E		E	P						
ENOPH1				P			PE		PE	P						
GOT1				P	P	P	P	P	P	P	P	P				
MRI1																
MTAP				P	P				P	P		P				
ODC1																
PAOX																
SAT1				P	P		P		P							
SMOX																
SMS				P												
SRM				P	P	P	P	P	P	P	P					





B

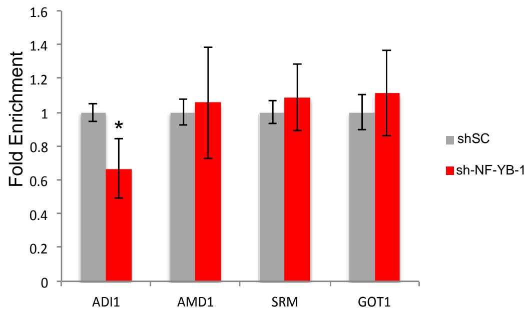


Figure 6

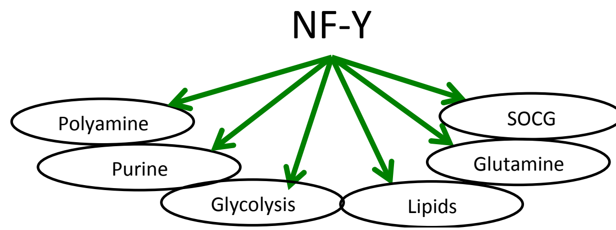
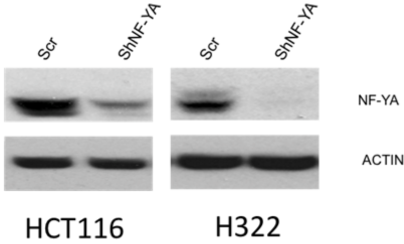


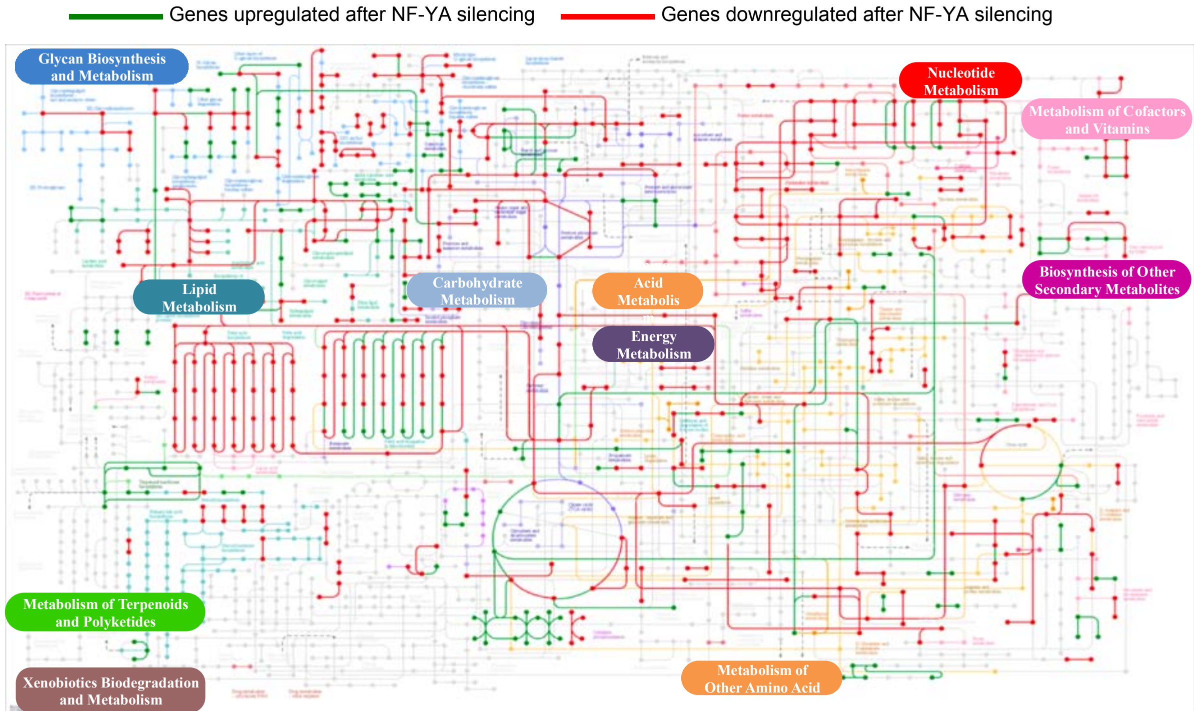
Figure 7

Fig.S1 Inactivation of NF-YA in HCT116 and H322 cell lines.

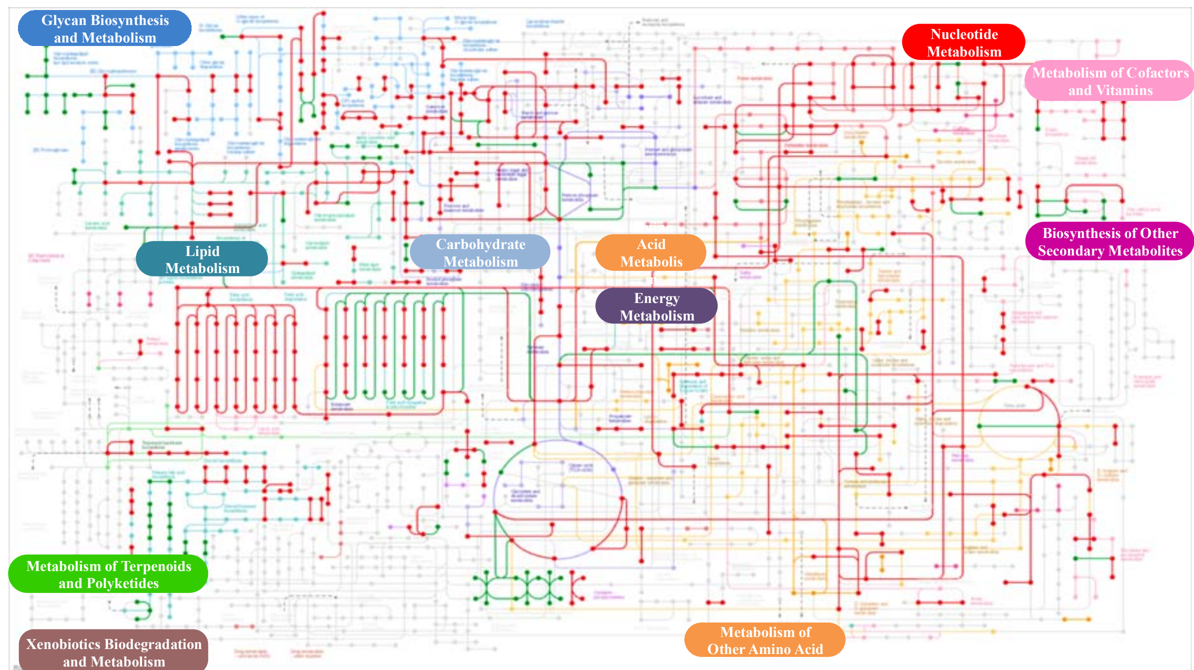


Cells were transduced with shSC or shNF-YA viruses and collected 72 hrs after treatment.

Fig.S3 Global Metabolic maps of genes deregulated after NF-YA inactivation in HCT116 and H322 cells

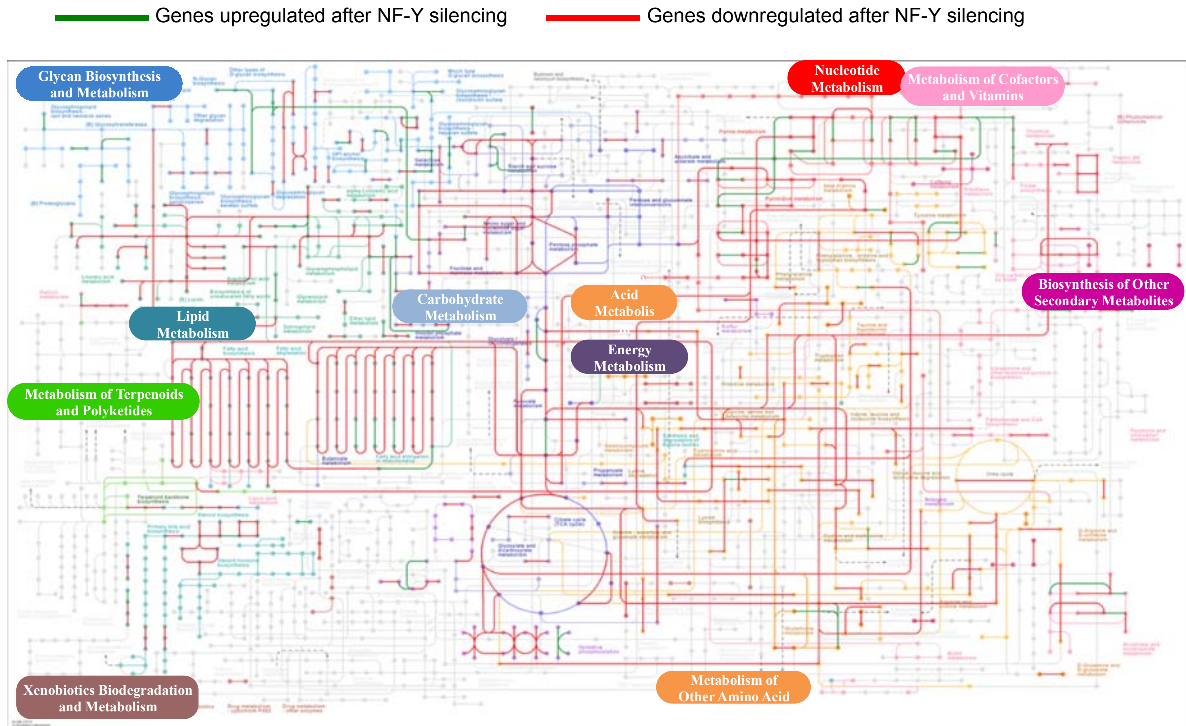


Gene expression profiling after NF-YA silencing in H322 cells

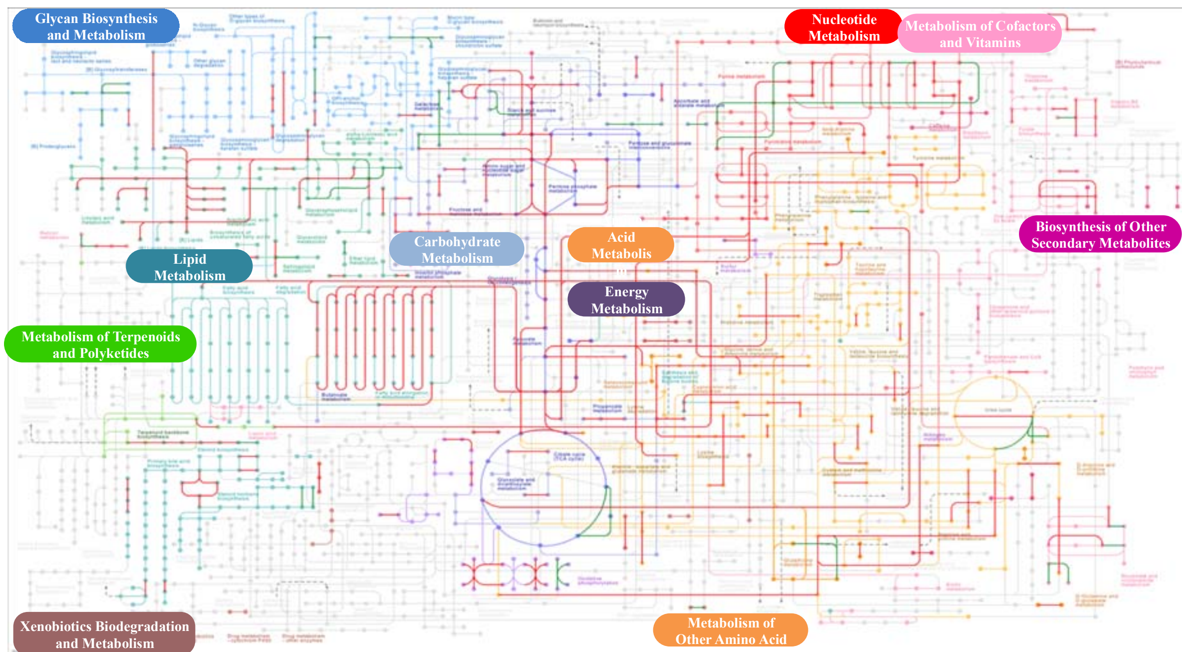


Gene expression profiling after NF-YA silencing in HCT116 cells

Fig.S4 Global Metabolic maps of genes deregulated after NF-YA, NF-YB, NF-YC and NF-Y (all subunits) inactivations in mouse embryonic stem cells.

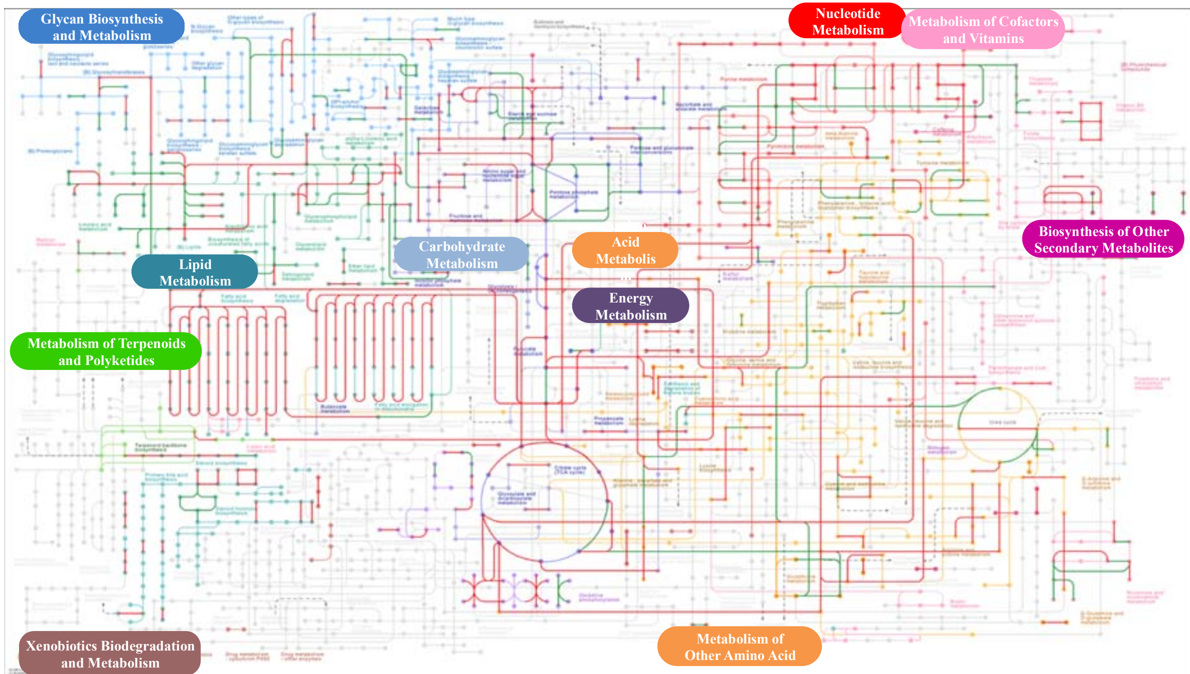


Gene expression profiling after NF-YA silencing in mES cells

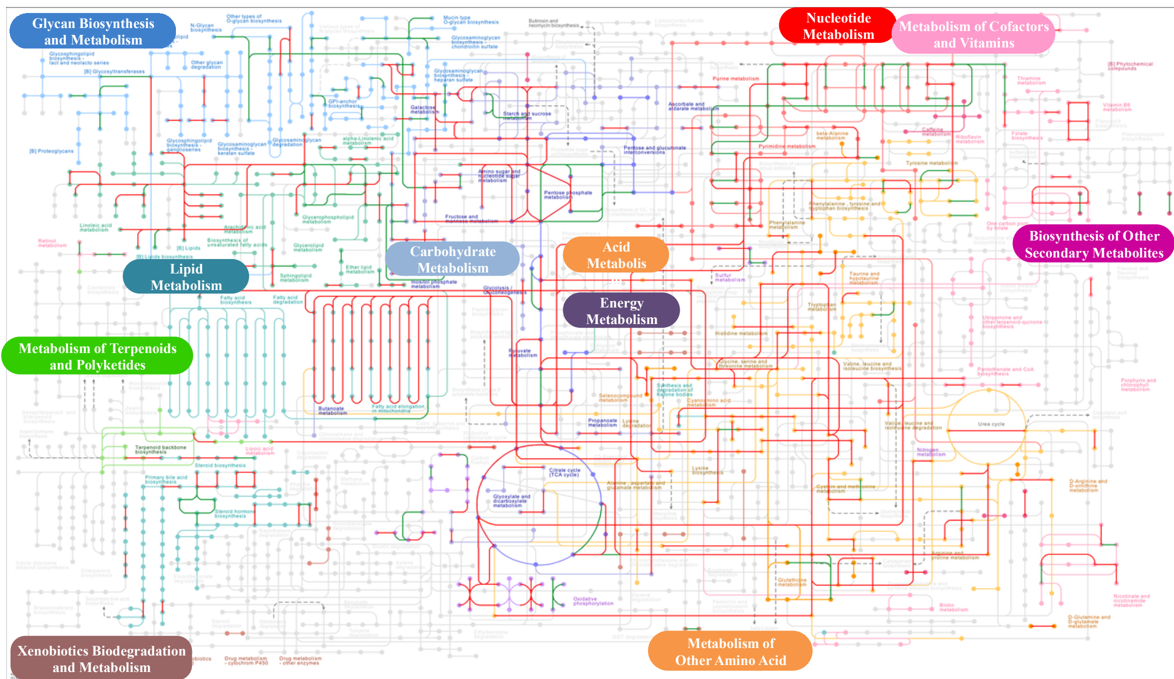


Gene expression profiling after NF-YB silencing in mES cells

Genes upregulated after NF-Y silencing   Genes downregulated after NF-Y silencing

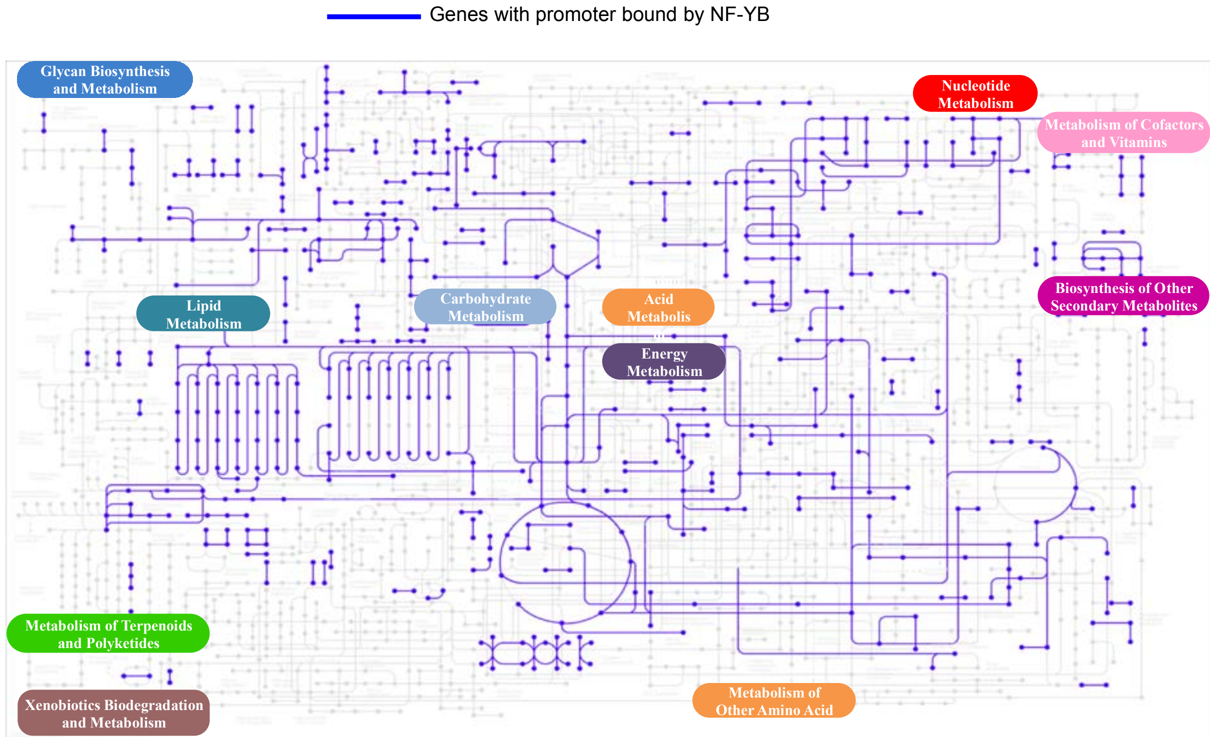


Gene expression profiling after NF-YC silencing in mES cells



Gene expression profiling after NF-Y silencing in mES cells

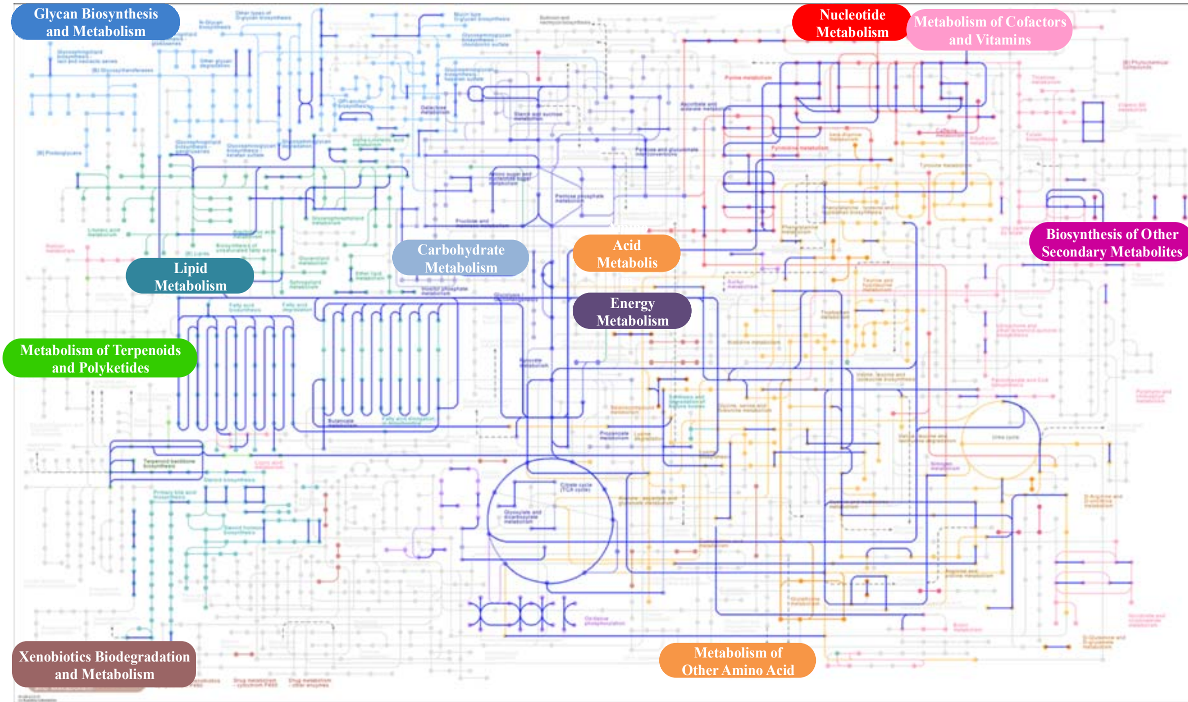
Fig.S5 Global Metabolic maps of genes bound in their promoters by NF-YB according to ENCODE data in HeLa cells.



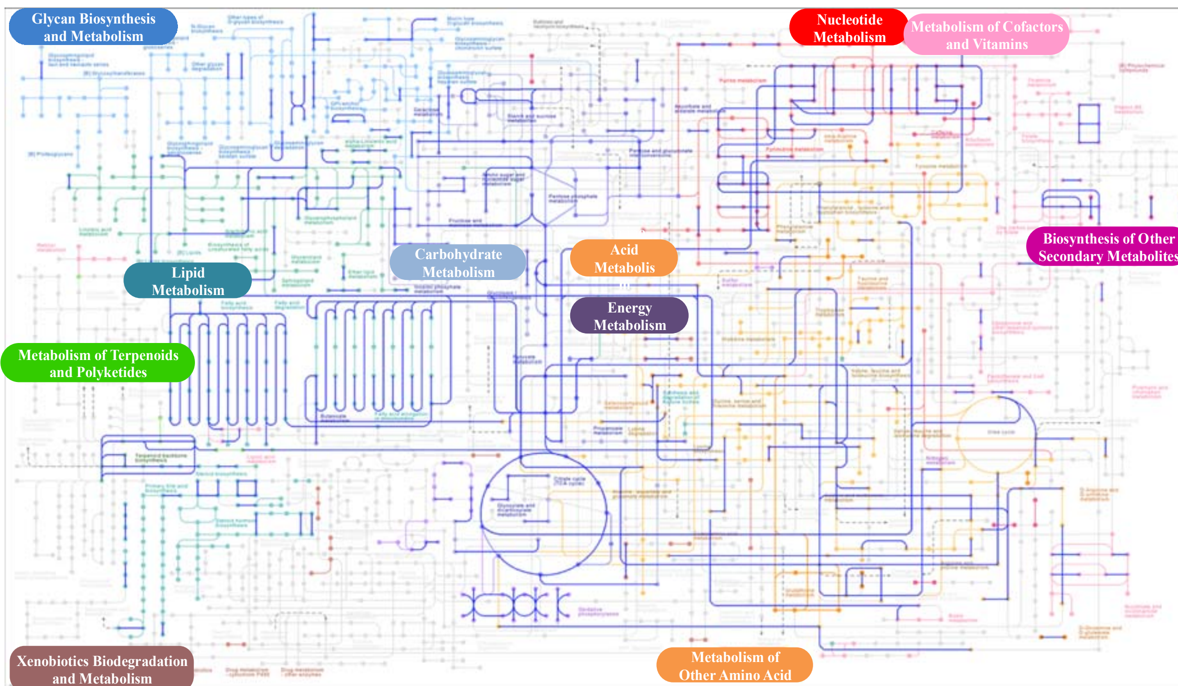
NF-YB binding in Hela cells

Fig.S6 Global Metabolic maps of genes bound in their promoters by NF-YA according to ENCODE data in K562 and GM12878 cells.

Genes with promoter bound by NF-YA



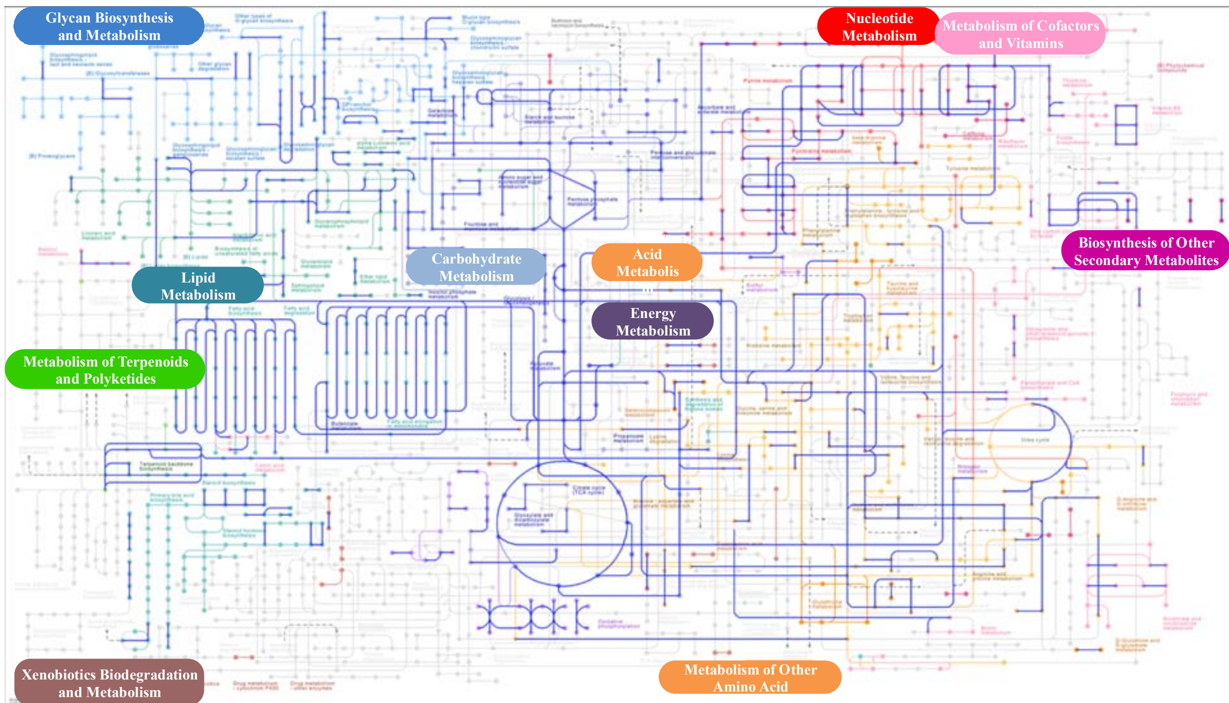
NF-YA binding in K562 cells



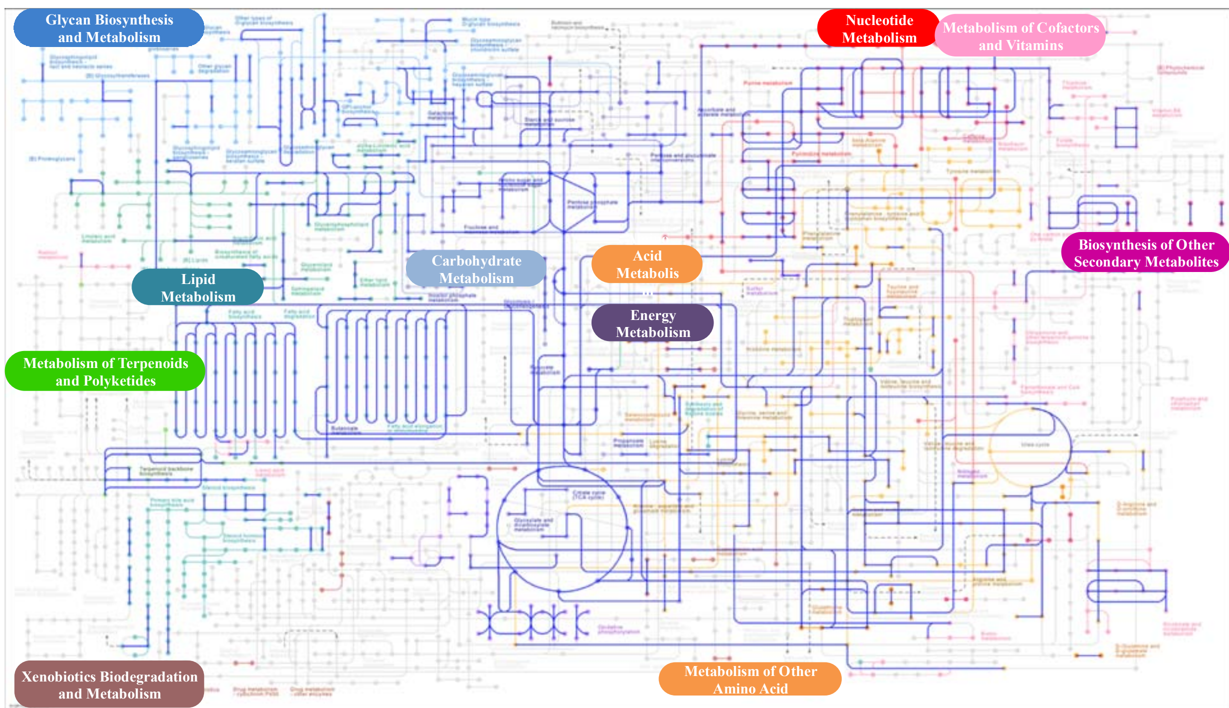
NF-YA binding in GM12878 cells

Fig.S7 Global Metabolic maps of genes bound in their promoters by NF-YB according to ENCODE data in K562 and GM12878 cells.

Genes with promoter bound by NF-YB



NF-YB binding in K562 cells



NF-YB binding in GM12878 cells

Fig.S8 Global Metabolic maps of genes bound in their promoters by NF-YA, NF-YB and NF-YC according to data published by Oldfield AJ et al..

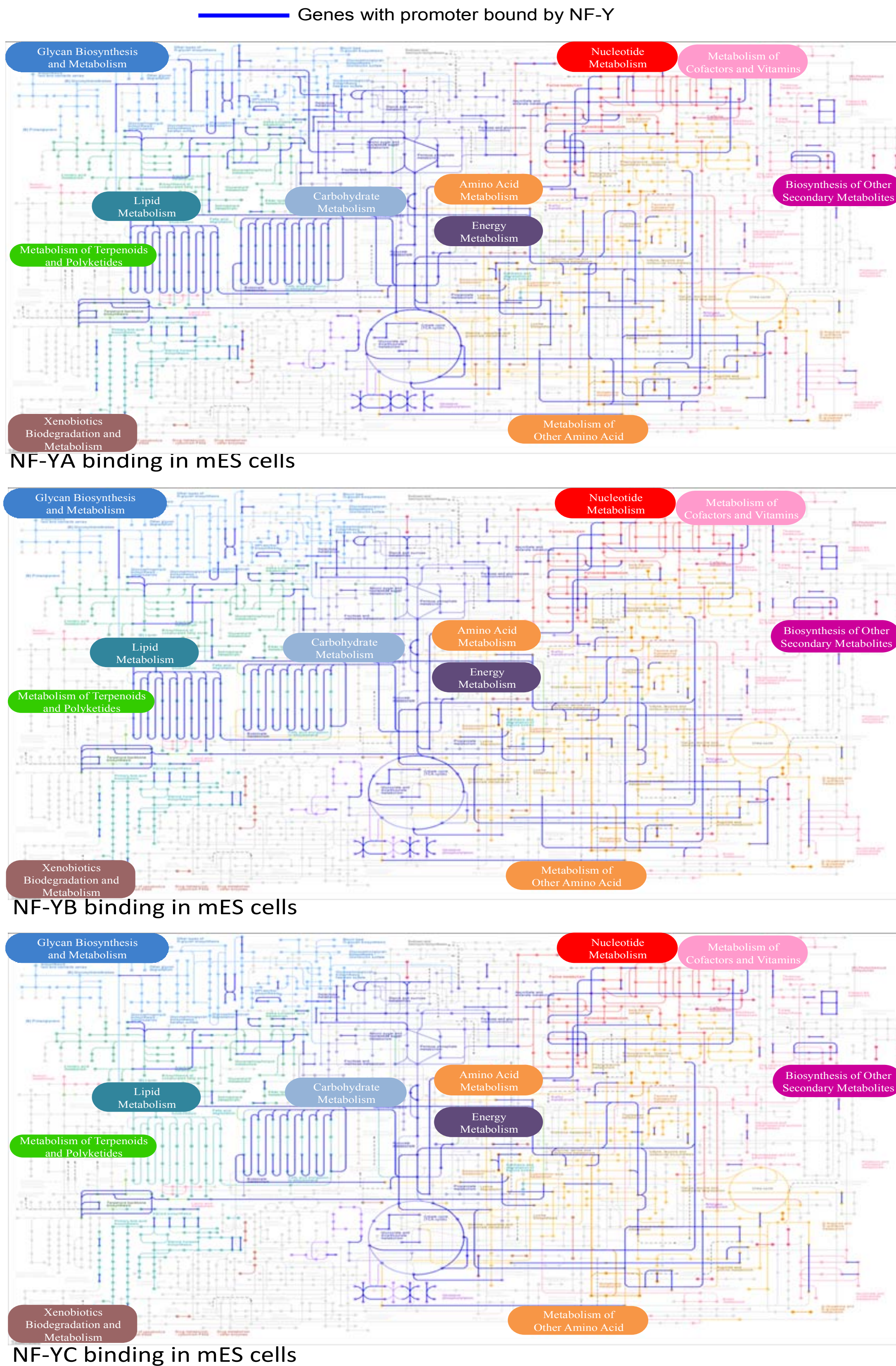
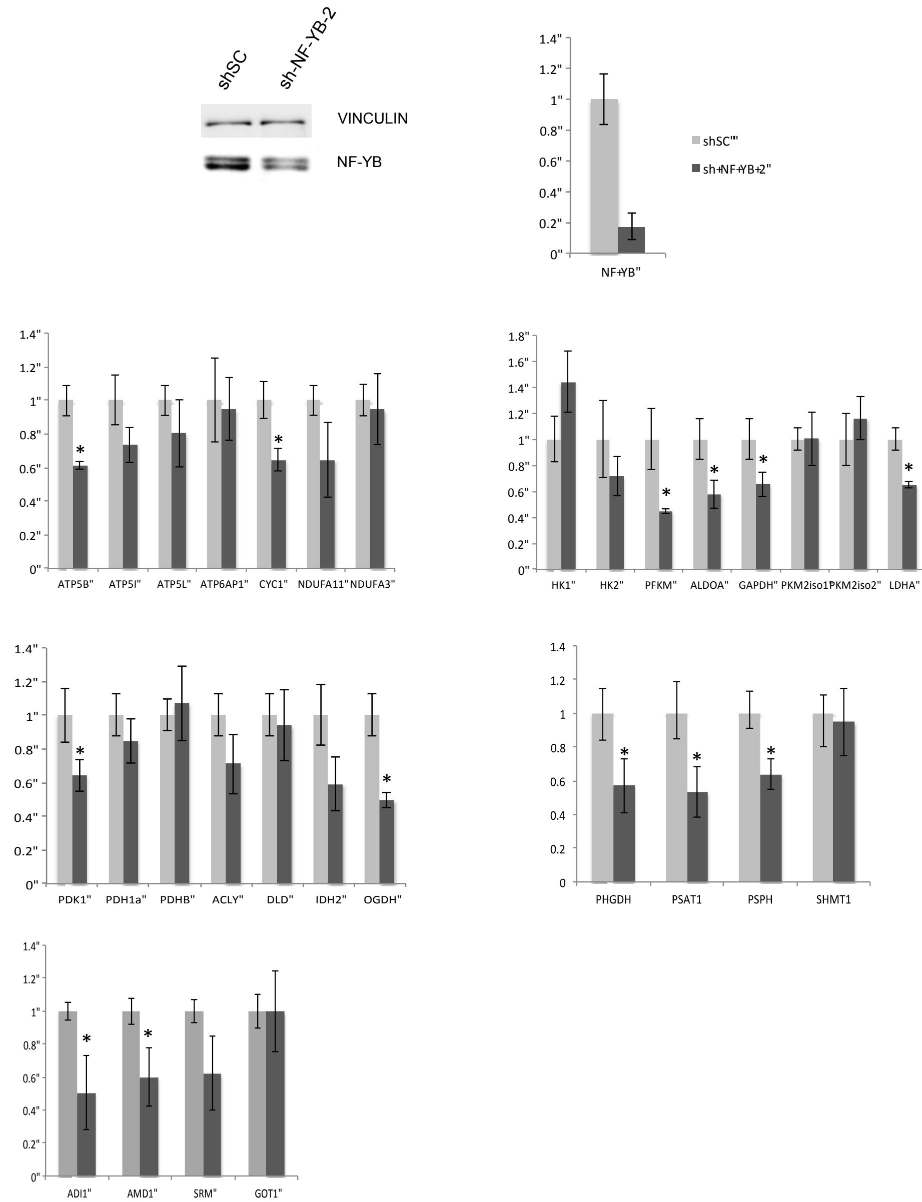


Fig.S11 Inactivation of NF-YB in Hela cells using sh-NF-YB-2 and analysis of expression of targets genes.



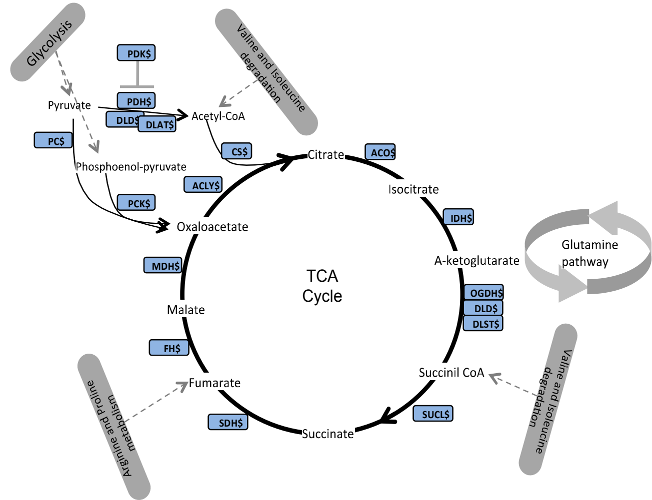
Western blot analysis of protein level of NF-YB in Hela cells inactivated of NF-YB using sh-NF-YB-2. qRT-PCR evaluation of expression levels of NF-YB and selected genes after inactivation of NF-YB in Hela cells. The average -/+ SD of three biological replicates is represented (*$p<0.05$).

Fig.S12 Regulation of genes of TCA cycle.

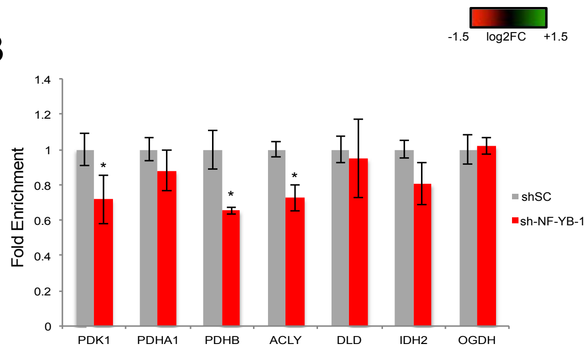
A

	H322	HCT116	HeLa-S3	HELA		K562		GM12787		mESc						
	shYA	shYA	shYA	YA	YB	YA	YB	YA	YB	YA	YB	YC	siYA	siYB	siYC	siNFY
PDK1						P	P			E	E	E				
PDK2				P	P	P	P	P	P	P	P	P				
PDK3				P	P		P		P							
PDK4				P	P		P		P	P	P	P				
PDHA1				P	P	P	P	P	P	P						
PDHA2																
PDHB				P					P	P	P	P				
DLAT									P							
PCK1																
PCK2				P	PEi	Ei	PEi	Ei	PEi	P						
PC				PE	PE	PE	PE	PE	PE	P	P	P				
ACLY				P	P	P	P	P	P	P	P					
CS				P	P	P	P	P	P	P	P					
ACO1																
ACO2				P	P	P	P	P	P							
IDH1																
IDH2				PEi	Ei	Ei	PEi		PEi							
IDH3A				P	P		P		P							
IDH3B									P							
IDH3G				P	P	P	P	P	P	P	P					
OGDH				P	P	P	P	P	P	P	P	P				
DLD				P	P	PE	PE	P	PE	P						
DLST																
SUCLA2																
SUCLG1					P		P		P							
SUCLG2				P	P	P	P	P	P							
SDHA																
SDHB				P					P							
SDHC																
SDHD																
FH				P	P	P	P	P	P	P	P					
MDH1				P	P	P	P	P	P	P						
MDH2																

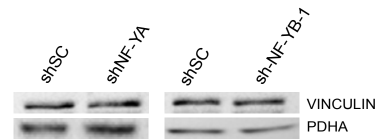
-1.5 log2FC +1.5



B

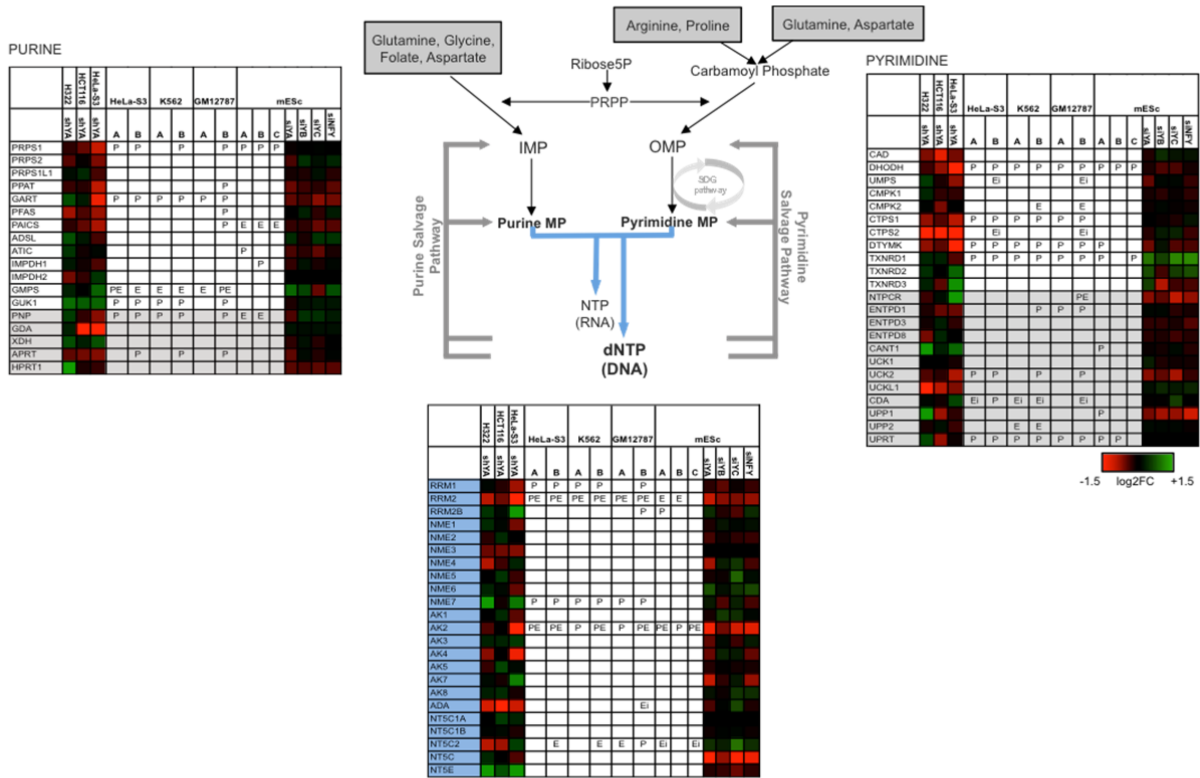


C



A) Expression levels of genes encoding TCA enzymes are shown after inactivation of NF-Y in different cell lines, and the presence of NF-Y binding is indicated as in Fig. 3. **B**) qRT-PCR validation of expression levels of selected genes after inactivation of NF-YB in Hela cells. The average -/+ SD of three biological replicates is represented (*p<0.05). **C**) Western blot analysis of protein levels of PDHA1 in Hela cells inactivated of NF-YA (Left panel) or NF-YB (Right panel).

FigS13 Activation of genes of purine and pyrimidine metabolisms by NF-Y.



Genes of the purine and pyrimidine metabolisms are shown with the expression levels after inactivation of NF-Y, and the presence of *in vivo* NF-Y binding.

PDK1	TCTCAGGACACCATCCGTTC
	GGGCCATTGTGGGAATGAC
PDH1a	GACTGTACGCCGAATGGAGT
	CCGGTAGGCTGTGATGAGAT
PDHB	AGGCATAGGGACATCAGCAC
	GGCCACAGTTTGGAGTAGGA
ACLY	CTGGGCGGTACAGTTTGGAG
	GGACCAGAAGGGAGTGACCA
DLD	CACACCCTGAAGTTGCTTGG
	CACCATGCCATCTGTGTCAG
IDH2	ACCTGAGCCACCATGTCATC
	CGCTGTACATGAGCACCAAG
OGDH	GCTGAGCCCTTTCTCAGTGG
	GGGGACTCTGGTAGGCAGTG
RPS20	GCGCCTCTTATCAAGTCAGC
	CGGAAAAACACCCGTGGAG
ATP5B	ACCTCAGCAACCTGGAATGG
	GTGGGGTGCAAAAGATCCTG
ATP5I	GGTCGCAGGGTCACTCACTT
	GGCAGAAGAGGAGAGGAGGA
ATP5L	CCTGCTGAGATCCCTAGAGCT
	GCCCCGCTTGCCTATAATC
ATP6AP1	ATACACAGCGGCCCTCACAG
	GGTCCTTGTACGCCACAGAG
CYC1	GTCACTGCGGGAAGGTCTCT
	GATGGTCGTGCTCTGGCTCA
NDUFA3	CAGTGCTGGTCGTGTCCTT
	TAGTTGTAGGGCGTGGCCTTG
NDUFA11	GCAGCCACCGAGGAAGTAGT
	CGCTGCCTACAGAGTCACAC
HK1	CGGACCACGGTTGGTGTC
	CTTGCCGCTGCCACTCTC
HK2	AGGAGAACAAAGGCGAGGAG
	CGCACGGTCTTATGTAGACG
PFKM	CTGGTGGATGGTGGAGATCA
	CACCCCCAATGACACAGAGA
ALDOA	TTGGGGGTGTCATCCTCTTC
	GGGTGGTAGTCTCGCCATTT
GAPDH	GCCTCAAGATCATCAGCAATGC
	CCACGATACCAAAGTTGTCATGG
PKM2 isoform 1	GGCTGAAGGCAGTGATGTGG

	GGCTTCCATGAGGTCTGTGG
PKM2 isoform 2	TACCAGCGACCCCACAGAAG
	GGAAGATGCCACGGTACAGG
LDHA	AGGCTGGGAGTTCACCCATT
	CAACCACCTGCTTGTGAACC
PHGDH	GCGTGTGGTGAACTGTGC
	AGGTGGGGACAGCTGATG
PSAT1	AGCAGGAAGGTGTGCTGACT
	GTAGGAGGCATCTGGGTTGA
PSPH	GCTGCGTCTCATCAAAACCT
	AGCTCAATATCCCAGCAACC
SHMT1	GCCCTTCCACCATCTGTG
	GACGGAGCTGGGCTACAA
ADI1	TCC CTC ACA TCG AAG TAC CC
	GGA ACT ACT CCT GGA TGG ACA
AMD1	AGC ACT CTG GCA ATC AAG AC
	CAG ACC TCC TAT GAT GAC CTG A
GOT1	GCT TCT AGT CGT GCC CTG AG
	CAC CCT CTC TAA CCC TGA GC
SRM	GGC AGC AGA GGA CAC CAT
	CAG GAT GCC TTC GAC GTG

The Fig. S2, S9 and S10 will be available on-line.

PART III

Manuscript in preparation:

The evolutionary conserved domain of NF-YA harbors two Cell Penetrating Peptides.

Report:

Preliminary analysis of recombinant NF-YA activity in mESCs and hints about the role of NF-Y in hiPSCs circuitry.

The evolutionary conserved domain of NF-YA harbors two Cell Penetrating Peptides.

Maria Luisa Chiaramonte[1], Mario Minuzzo[1], Nerina Gnesutta[1], Diletta Dolfini[1] and Roberto Mantovani[1]*

[1] Dipartimento di Bioscienze, Università degli Studi di Milano.

*Corresponding author

Roberto Mantovani,

Dipartimento di Bioscienze, Università degli Studi di Milano

Via Celoria 26, 20133, Milan, Italy

Tel. 02503 15005

e-mail: mantor@unimi.it

ABSTRACT

Cell Penetrating Peptides -CPPs- are short aminoacidic stretches that enable proteins to be internalized into cells. They are usually rich in basic residues, and often organized as alpha helices. NF-Y is a transcription factor trimer that binds to the widespread CCAAT box element. The NF-YA subunit possesses two α-helices rich in basic residues in the 56 aminoacids evolutionarily conserved part. Protein transfections with recombinant NF-YA showed efficient entry at nanomolar concentrations. Mutagenesis pinpointed two CPPs in the subunits interaction A1 and DNA-binding A2. The half-life of the transduced protein is shorter in transformed cells than in immortalized HaCaT and NIH3T3. In NIH3T3 cells, NF-YA affects transcription of CCAAT-dependent promoters inducing growth retardation, and an increase in the expression of pro-apoptotic genes. In conclusion, we describe novel CPPs empowering NF-YA with new and unexpected properties.

1 INTRODUCTION

Cell Penetrating Peptides (CPPs) are short aminoacids stretches which possess the capacity to mediate the internalization of the proteins, also functioning as carriers of additional materials, including nucleic acids and drugs (Dietz *et al.,* 2004; Mae *et al.,* 2006). They were originally discovered within sequences of proteins with a transcriptional role, the Drosophila Antennapedia and HIV TAT (Derossi *et al.,* 1994; Green *et al.,* 1988,). The peptides were extensively characterized: when fused to proteins, they provide a useful mean to efficiently introduce them into cells. Such protein transduction systems appear to be efficient for a number of proteins, with a wealth of possible biotechnological and therapeutic implications being currently explored (Roeder *et al*., 2004; Araki *et al*., 2010). Two common structural features of CPPs are the abundance of Arginines and the presence of amphipathic α-helices (Futaki *et al.,* 2001; Crombez *et al.,* 2008; Eiríksdóttir *et al.,* 2010).

One of the proteins in which TAT fusions proved valuable was the NF-YA subunit of the trimeric transcription factor NF-Y: it was demonstrated that a functional GST-TAT-NF-YA could be efficiently introduced into nuclei of hematopoietic cells by protein transfection (Domashenko *et al.*, 2010). Short term exposures of this protein, in the 10-50 nanomolar range, significantly affected transcription of NF-Y target genes, without inducing overt cellular toxicity. Indeed, it was shown that treatment of pools of cells enriched in hematopoietic stem cells –HSCs- led to an expansion of the stem precursors pool *ex vivo*, with a subsequent increase in bone marrow engraftment in irradiated nude mice. The functionality of recombinant TAT-NF-YA further assessed in mouse ES cells (Dolfini *et al.*, 2012), and embryonal carcinoma NT2 cells (Moijsin *et al.,* 2015), thus confirming that this strategy is potentially exploitable.

Functionally, NF-Y is a "pioneer" TF, responsible to set up a favorable chromatin environment for companion "activating" TFs to bind DNA and induce mRNA elongation by the RNA Pol II machinery (Kabe *et al.,* 2005). Gene deletion of NF-YA leads to lack of development due to early embryo failure, and conditional KO experiments established multiple roles in proliferating cells -liver and HSCs-, and in post-mitotic cells, adipocytes and cortical neurons (Bhattacharya *et al.,* 2003; Lu *et* al., 2015; Bungartz *et al.,* 2012; Yamanaka *et al*., 2014).

NF-YA is the limiting subunit of the NF-Y trimer, which comprises two Histone Fold Domain -HFD- subunits, NF-YB and NF-YC (Nardini *et al.,* 2013). The structure of the trimer bound to the target CCAAT sequence illustrates the unique modalities of DNA recognition: it is based on several non sequence-specific contacts of HFD subunits with the phosphate backbone *a-la* H2A/H2B, and specific CCAAT-binding in the minor groove by conserved structures of NF-YA: an amphipathic α-helix -A2- and the unique GxGGRF motif. The HFD interacting part of NF-YA is located in another

α-helix, A1. Overall, the conserved 56-aminoacids domain of NF-YA is extremely basic, and subject to post-translational modifications, including Lysines-mediated poly-ubiquitination and degradation, responsible for the relatively rapid turnover of the protein (Manni *et al.,* 2008).
The inspection of the NF-Y/CCAAT 3D structure, in particular of NF-YA, suggested us the presence of features of CPPs. We decided to verify this hypothesis.

2 Materials and Methods

2.1 Protein production and purification. The wild type long and short NF-YA in pEt3b, pEt3b-NF-YB and pEt32-NF-YC plasmids were described (Li *et al.,* 1992). The mutants YAl 31, 32, 33 and 34 of NF-YA were constructed by PCR, using mismatched oligos; the fragments were inserted using *Kpn*I-*Bam*HI sites in the pEt3b/NF-YAl backbone. pEt3b/NF-YA23 and pEt3b/NF-YA29, described in Mantovani *et al.,* 1994, were used to create mutants YAl 23-33, YAl 23-34, YAl 31-29 and YAl 32-29. The sequences of the mutants were checked by DNA sequencing.

NF-YA recombinant proteins were produced using *Escherichia coli* BL21(DE3): logarithmically growing cells were induced for 3 hours 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 (Li *et* al., 1992). The procedure was repetead twice. Inclusion bodies were solubilized in 8 M urea, 150 mM NaCl, 20 mM Tris-HCl, pH 7.5, 0.1 mM DTT, renatured by dialysis in urea-free buffer (150 mM NaCl, 20 mM Tris-HCl, pH 7.5, 0,1 mM DTT, 10% glycerol), centrifuged and stored at -80°C. Each recombinant NF-YA described (wt and mutants) was prepared three times. LPS was removed from protein preparations using Triton X114, as described (Liu *et al*., 1997). 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 minutes.

Recombinant NF-YB and NF-YC (37 kD isoform) were produced separately as described above, 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 100 mM KCl, 20 mM Tris-HCl, pH 7.8, 5 mM β-mercaptoethanol, 6 M GnCl. NF-YB and NF-YC were mixed in equimolar amounts and renaturated by slow dialysis in 10 mM Tri-HCl pH8, 300 mM KCl, 10% glicerol, 1 mM DTT and 3.5 M urea for 2 hours at room temperature and then in-urea free buffer (10 mM Tri-HCl pH8, 300 mM KCl, 10% glicerol, 1 mM DTT) and stored at -80°C. After renaturation, NF-YB/NF-YC were 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 Tri-HCl pH8, 300 mM KCl, 10% glicerol, 1 mM DTT, dialyzed over-night in the same buffer without Imidazole, and stored at -80°C.

2.2 Electrophoretic mobility shift assay. The probe used was a 31 bp double-stranded oligonucleotide containing the Hsp70 promoter CCAAT-Box, 5'-labelled with Cy5: (5'-

CTTCTGAGCCAATCACCGAGCTCGATGAGGC-3' Sigma). The annealing reaction was performed with an equimolar concentration of the complementary oligonucleotide (20 μM) in a final volume of 40 μl of TE. The dose-response experiments were performed with different concentrations of NF-YA and at 20 nM and 60 nM of DNA and NF-YB/NF-YC, respectively. Protein was diluted in dilution buffer (400 mM NaCl, 10 mM Tris-HCl, pH 8, 10% glycerol, 0,1 mg/mL BSA, 2.5 mM DTT). For incubation, binding buffer 5X (250 mM KCl, 25 mM $MgCl_2$, 60 mM Tris-HCl, pH 8, 2.5 mM EDTA, 60% glycerol, 25 mM DTT, 1 mg/mL BSA) was added to proteins and to labeled DNA. Reactions were incubated at 30°C for 30 minutes, protected from light, and loaded on the gel (TBE 0.25X, 4% acrylamide, 19:1 Bis ratio, 2.5% glycerol, 0.1% APS), in a dark room at 4°C; electrophoresis was carried out at constant 120 V for 3 hours. Detection was performed with the ChemiDoc^{TM} MP imaging system (Biorad). Each EMSA experiment was performed in triplicate.

2.3 Cell cultures and NF-YA protein transduction. Human HaCaT keratinocytes, human colon carcinoma HCT116, human cervical carcinoma HeLa, human osteosarcoma U2OS, human fibrosarcoma HT1080 and NIH3T3 mouse fibroblasts were cultured in Dulbecco's modified Eagle's medium (DMEM) containing 10% of fetal bovine serum (FBS), L-glutamine, 100 units/ml penicillin and 100 μg/ml streptomycin at 37°C in a humidified 5% CO_2 atmosphere. 25 x 10^4 cells/well in 6 well plates were washed with DMEM without serum before protein delivery. Exponentially growing cells, were washed with PBS, treated with the indicated amounts of recombinant NF-YA in serum-free medium for 30 minutes at 37°C in 5% CO_2. After 30', cells were washed with serum-free medium and further incubated with recombinant NF-YA for further 30'. Finally, cells were washed with serum-free medium and cultured in complete DMEM, and collected at different time points.
Transfections with wt and mutant NF-YA proteins were repeated 2-4-times.

2.4 Cell extracts and Western Blot Analysis. 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 mixture (Sigma) by incubation on ice for 30 minutes. NP40 was added to the final concentration of 0.65%. Cells were vortexed and centrifuged at 13,000 rpm for 5 minutes. Pelleted nuclei were resuspended in ice-cold Buffer C (1 M NaCl, 20 mM Hepes, pH 7.8, 1 mM EDTA, 1 mM EGTA, 1 mM dithiothreitol, protease inhibitors). Protein concentration was determined by Bradford (Bio-Rad Protein Assay). 10 μg of nuclear extracts were loaded on 10% SDS-polyacrylamide gel and analyzed by Western blot using the following primary antibodies: anti-NF-YA (Santa Cruz Biotechnology), anti-Vinculin (Santa Cruz Biotechnology), anti-Lamin B (Sigma), anti-LPS (Hycult biotech). Protein-antibody detection was performed using peroxidase-conjugate

secondary antibodies and ECL reagents (GeneSpin). Visualization of images was obtained with the ChemiDoc^{TM} MP imaging system (Biorad) and for quantification of the data we used the ImageLab software (Biorad).

2.5 RNA extraction and RT-PCR analysis. Total RNAs were extracted using Trizol reagent (Sigma) according to the manufacturer's instructions. 1 μg of RNA was retrotranscribed with SuperScript^{TM} reverse transcriptase (Biorad) and random primers. cDNA was diluted 1:20 with water and analyzed in triplicate in q-RT-PCR analysis. To normalize the data we used the housekeeping ribosomal gene RPS15. A comparative $2^{-\Delta\Delta Ct}$ was used to detect relative gene expression. q-PCR analysis was performed using the primers indicated in Supplementary Fig. 6

2.6 Cell counts by trypan blue exclusion. 80 x 10^4 NIH3T3 cells were plated in triplicate for each treatment in 60 cm plates and collected by trypsinization at different times. The growth curves were obtained by counting cells in Bürker chamber, using the trypan blue exclusion method. Duplicate biological replicates, each with triplicate samples, with two different proteins preparations, were performed.

3 RESULTS

3.1 Recombinant NF-YA enters cells efficiently.

We noticed that the evolutionarily conserved part of NF-YA contains two features of CPPs, the presence of α-helices -A1 and A2- and the abundance of Arg residues (Nardini *et al.,* 2013). We decided to test whether recombinant NF-YA would enter cells in the absence of any fused CPP. We produced in *E. coli* the two recombinant NF-YA isoforms, "long" and "short" -NF-YAl and NF-YAs- (Li *et al.*,1992; Mantovani *et al.,* 1992): the proteins were found in inclusion bodies, with a high degree of purity, as checked by SDS-PAGE (Fig. S1A). The recombinant proteins were solubilized in 8M urea and renatured by slowly removing the denaturing agent, as described (Mantovani, 1992). To get rid of contaminating LPS-like material present in the bacterial preparations, we used the Triton X114 precipitation protocol (Liu *et al.*, 1997): indeed, this procedure largely eliminated LPS from our preparations, as checked by Western blot analysis (Fig. S1B).

The recombinant proteins were then quantified and used in protein transfections of human colon carcinoma HCT116 cells. We tried different transfections procedures and finally set up a simplified system by adding the proteins twice at 30' intervals in serum-free medium.

Because NF-Y is localized exclusively in the nucleus (Frontini *et al.,* 2004), where it exerts its DNA-binding and transcriptional functions, we prepared and assayed nuclear extracts in all further experiments. Fig. 1A shows a dose-response of NF-YAs and NF-YAl in nuclear extracts of HCT116 cells prepared 2 hours post-transfection, monitoring protein uptake into cells by Western blot with an anti-NF-YA antibody: NF-YAs and NF-YAl enter cells at low nanomolar concentrations, as it was previously shown for TAT-NF-YA. Remarkably, transfections yielded amounts of recombinant NF-YAs and NF-YAl that are many-fold those of the endogenous NF-YAs. HCT116 cells mainly express the short isoform of NF-YA; note the superimposition of the transfected short isoform with the endogenous one (Fig. 1A).

We next performed a kinetic analysis of NF-YAl transfected protein levels: Fig. 1B shows that NF-YAl is observed at near maximum levels already after 30' post transfection, and still high at 2 hours; by 6 hours, the protein is considerably decreased.

The endogenous NF-YAs of HCT116, serves as internal control, in addition to the loading control LaminB. Interestingly, the short half-life of the overexpressed protein is very similar to the one previously calculated, including in overexpression experiments generated with DNA-mediated gene transfections (Manni *et al.,* 2008). Thus, it appears that NF-YA is rapidly internalized into cells and degraded.

3.2 Identification of NF-YA CPPs.

The assays shown below were used to identify the aminoacids responsible for internalization of NF-YA. Positively charged aminoacids are present in the primary sequence of NF-YA exclusively in the evolutionarily conserved domain (AA 262-317), which contains 7 Lysines and 11 Arginines. Specifically, there are four stretches of 3 consecutive basic residues, which were mutagenized into Alanines (Fig. 2A, YAl 31-34): YAl 31 and YAl 33 affect residues in the A1 and A2 helices, YAl 32 and YAl 34 involve residues of the linker between A1 and A2 and of the RXGGRF DNA-binding motif, respectively (Nardini *et al.,* 2013). Subsequently, we prepared double mutants in which these mutations were coupled to previously described mutants YA23 and YA29, in the subunits-interaction and DNA-binding subdomains (Mantovani *et al.,* 1994): thus we generated mutants YAl 29-31, YAl 29-32, YAl 23-33, YAl 23-34.

We produced these mutants in the NF-YAl backbone: Coomassie staining of SDS-PAGE of representative preparations are shown in Fig. S2. To verify their functionality, we initially characterized these mutants for DNA-binding in EMSAs on the high affinity CCAAT box oligo of the HSP70 promoter (Nardini *et al.,* 2013): note that mutants in the A2 and GXGGRF motif are predicted to be deficient in DNA-binding assays. Dose-response of NF-YAl on a fixed amount of

recombinant NF-YB/NF-YC dimer generated a specific band, as expected; the linker-mutated YAl 32 showed robust DNA-binding, while A1 mutated YAl 31 was decreased (Fig. 2B). On the other hand, YAl 33 and YAl 34, crippled in A2 and the GXGGRF motif, were completely negative, even at high protein concentrations (Fig. 2B). Similarly negative were the double mutants YAl 29-31, YAl 29-32, YAl 23-33 and YAl 23-34 (Fig. 2C and Fig. 2D). These results were expected, but we felt important to verify whether this was due to the renaturation process. To do so, we exploited the capacity of the DNA-binding mutants to bind normally to the HFD dimer, and prevent DNA association of wt NF-YAl, therefore acting as Dominant Negatives (Mantovani *et al.,* 1994). We performed EMSAs by incubating YAl 33 and YAl 34 with a wt NF-Y complex prior to addition of DNA: Fig. S2 shows that both mutants were able to decrease the binding of NF-Y, indeed indicating that they are functional in terms of subunits interaction. Therefore, the NF-YA mutants behaved *in vitro* as expected, based on previous extensive mutagenesis (Xing *et al.,* 1993; Xing *et al.,* 1994; Mantovani *et al.,*1994), and the analysis of the 3D structure of the NF-Y trimer in complex with DNA (Nardini *et al.,* 2013).

Thereafter, we assayed these mutants in the proteins transfection assays: kinetic analysis at 10nM concentrations performed in HCT116 cells monitored NF-YA accumulations by Western blots. Fig. 3 shows that YA 23, YAl 31, YAl 32, YAl 33 and YAl 34 were present in nuclear extracts, but at decreased levels compared to the wt NF-YAl. YA 29 was somewhat lower. The double mutants YAl 29-31 and YAl 29-32 showed levels of NF-YAl similar to the single YAl 31 and 32 mutants. In none of these mutants complete elimination of uptake was observed, suggesting the presence of more than one CPP. On the other hand, the entry of YAl 23-33 and YAl 23-34 into cells was essentially abolished. Quantifications of these experiments are shown in Fig. S3. These results indicate that there are indeed two separate CPPs in NF-YA, one in the A1, centered on the KRR sequence, and a second in the DNA-binding subdomain, encompassing the basic RHRHAMARKR stretch.

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

The experiments performed in the tumorigenic colon carcinoma HCT116 cell line were extended to other cell lines: HeLa (human epithelial cervical carcinoma) U2OS (human osteosarcoma), HT1080 (human fibrosarcoma), HaCaT (human immortalized epithelial keratinocytes) and NIH3T3 (mouse immortalized fibroblasts). We employed the 10nM dose of recombinant NF-YAl over 24 hours (Fig. 4): in all cell lines, the protein is efficiently internalized, with expression maximal, or near maximal at 30' post-transfection. Thereafter, however, there were differences in the stability of the protein. In HeLa, the levels were substantially reduced already at 2 hours after transfection and no exogenous

protein detected at 6 hours: for this reason we did not proceeded with 24 hours time point. At the opposite, slower degradation rates were observed in NIH3T3, HaCaT and, to a lesser extent, in U2OS. In NIH3T3, in particular, appreciable amounts of transfected NF-YAl was still visible after 24 hours. Note that further assessment at 48 hours post-transfection yielded negative results (See below). HT1080 showed a behavior similar to U2OS. Quantifications of these results are shown in Fig. S4. In summary, we generalize the entry of recombinant NF-YA in cells, but we notice a different half-life of transfected NF-YAl depending on the specific cell line used.

3.4 Functionality of transfected NF-YA.

Previous work with TAT-NF-YA fusions established that the recombinant fusion protein is functional upon transfection in cell lines and primary cells (Moijsin *et al.,* 2015; Domashenko *et al.*, 2010). To assess this aspect of wt NF-YA, we used the NIH3T3 system, because of the longer half-life of the protein, which could exert more lasting effects on cellular physiology. We also considered the possibility that the two NF-YA isoforms might have a different behavior: we included NF-YAs in these experiments, as well as the YAl 23-33 mutant incapable to enter cells, as a negative control. We used two doses of NF-YA, 10 and 100 nM. After transfections with the different recombinant proteins, exponentially growing NIH3T3 were monitored for cell growth at 24 and 48 time points: Fig. 5A shows that both NF-YAl and NF-YAs slowed cell growth, compared to cells mock-transfected with an irrelevant protein (BSA), or treated with YAl 23-33. The difference is particularly relevant at 24 hours, more pronounced for NF-YAl, and observed with the two protein concentrations. Note that appreciable levels of transfected NF-YA proteins are visible after 24 hours, but not at 48 hours (Fig. S5).

NF-Y controls a number of growth promoting, as well as pro- and anti-apoptotic genes: the growth suppression by recombinant NF-Y could be due to alteration in the different pathways. To shed light on this, we controlled the expression levels of genes involved in cell-cycle progression and control of apoptosis by qRT-PCR. Fig. 6 shows that G1/S (*Pcna*) and G2/M (*Ccnb1/2*) CCAAT-dependent genes are down-regulated by NF-YAl and NF-YAs; note, that while CCAAT boxes are important for phase-specific regulation of these promoters, they also mediate repression under specific conditions (Manni *et al.,* 2001; Gurtner *et al.,* 2003). On the other hand, apoptosis genes are differentially regulated: the pro-apoptotic *Bax* and *Apaf1* are up-regulated, the anti-apoptotic *Bcl2* is repressed. CCAAT-mediated regulation is not generalized, since *Topoisomerase IIa*, another NF-Y target, is unaffected, as is *Mll1*, a CCAAT-less gene. We also checked *E2f1*, an important NF-Y target, whose overexpression activates pro-apoptotic genes, notably *Apaf1* and *Bax* (Van Ginkel *et al.,* 1997; Moroni *et al.,* 2001; Giovanni *et al.,* 2000; Kabe *et al.,* 2005): indeed *E2f1* is activated by NF-YAl

and NF-YAs. In conclusion, NF-YA protein transfections in NIH3T3 mimic the effects of DNA-mediated NF-YAl overexpression, in terms of induction of pro-apoptotic genes, possibly co-mediated by up-regulation of *E2f1*.

4 DISCUSSION

CPPs were identified in several proteins, drawing considerable interest over the last two decades, not least for their far-reaching translational potential. NF-YA adds to the list of TFs entering cells without the requirement of cargos, *via* the presence of two distinct CPPs in the evolutionarily conserved region.

Our work was motivated by (i) previous experiments with GST-TAT-NF-YA fusions, showing efficient entry of the proteins, and functional proficiency in transcriptional activation, including in primary HSCs, and (ii) inspection of the NF-Y/CCAAT 3D structure. We reasoned that CPPs could be present in A1 or in A2: the transfection experiments confirm this and the mutagenesis experiments indicate that either CPPs mediate substantial cellular uptake. The mutants behave as predicted in DNA-binding assays *in vitro*, a relevant point to verify since we used proteins renatured from inclusion bodies. In transfections, elimination of two separate stretches –A1 KRR- and A2 RRK-HRH-RKR- is required to abolish the full cell penetrating capacity of NF-YA. The protein concentrations are similar to the ones used for TAT-NF-YA, in the nanomolar range: thus, considering that many proteins are internalized in the low μM range, the efficiency of NF-YA CPPs is remarkable, and 30 minutes incubation are sufficient to yield nuclear levels of NF-YA that are many-fold those of the endogenous proteins.
Under normal conditions, endogenous NF-YA is exclusively localized in the nucleus: this is unlike NF-YC, which utilizes heterodimerization with the companion NF-YB as a mechanism to enter in the nucleus (Frontini *et al., 2004;* Kahle *et al*., 2005). NF-YA nuclear localization is mediated by β-importin interactions. Interestingly, two separate signals are required for efficient nuclear localization of NF-YA, in human cells *in vivo,* and *Xenopus* cells *in vitro* (Kahle *et al.,* 2005): one is the KRR stretch in the A1, the other involves residues in the A2 (RRK and RKR); note that NF-YA mutations in either A1 or A2 retain substantial nuclear localization, whereas mutations of both essentially abolished it. These are the stretches showing CPP activity in our study. Thus, both A1 and A2 peptides are required for cellular penetration and nuclear localization. It remains to be determined whether the newly identified CPPs behave so when independently transferred to unrelated proteins.

The efficient entry is matched by a short half-life of the protein, but the half-life of the transfected NF-YA was dissimilar in the cellular contexts we used. Transformed HeLa, HCT116, HT1080, U2OS retain NF-YA for shorter periods, with respect to non-transformed HaCaT or NIH3T3 cells. Whether this is due to transformation *per se*, rather than to the specific tissue-specific contexts, is a relevant point that could be addressed by assaying larger sets of primary and transformed cells. Nevertheless, the half-life of the transfected NF-YA appears to be remarkably similar to the endogenous one and the produced by DNA-mediated transfections (Manni *et al.,* 2008), a further indication that the recombinant proteins behave similarly to internally produced NF-YA. On the other hand, the rapidity of NF-YA degradation poses problems in assessing its functionality. In general, compared to the HFD dimer, NF-YA turnover is more rapid, and subject to extensive control by degradation. The Lysines involved in poly-ubiquitination are all in the homology domain. The acetylation protects them from poly-ubiquitination, and indeed K to R mutations prolong NF-YA half-life (Manni *et al.,* 2008). Some of these Lysines are within the stretches identified here as CPPs: as basis for future experiments, it is possible that, K to R mutants, will increase the nuclear permanence and activity of transfected NF-YA. Whether they impair CPP activity remains to be seen.

As for other TFs shown to behave similarly, the fascinating question is what might be the physiological role, if any, of the cell penetrating capacity of NF-YA. Obviously, the aminoacids stretches identified here are required for key functions of the protein, nuclear localization, HFD interactions and sequence-specific DNA-binding, sufficient to justify their remarkable evolutionary conservation. However, this does not rule out that NF-YA can use the CPPs for additional purposes. For other TFs, cellular entry is matched by active export, supporting the notion that they can be used as signaling molecules in local paracrine circuits (Prochiantz and Joliot, 2003). One possible such case for NF-YA behavior might take place in the biology of macrophages. Monocytes undergo differentiation to macrophages by activation of sets of genes, in which CCAAT boxes predominate (Marziali *et al.,* 1999). Among these, the genes of the antigen presentation pathways, chiefly MHC Class II, rely on the well characterized combination of NF-Y/RFX and CIITA for activation (Zhu *et al.,*2000). Circulating monocytes are devoid of NF-YA, which is upregulated upon terminal differentiation. 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 pre-existing HFD subunits, leading to accelerated maturation of effector macrophages, and help triggering of the immune response.

The functionality of the transfected NF-YA, previously established with TAT fusions, is confirmed. The effect of transfections –inhibition of cellular growth- is at first sight surprising, since NF-Y, in general, is known to bind growth promoting genes, and, indeed, NF-YC was recently isolated as a driver oncogene (Tong *et al.,* 2015). However, it was shown that DNA-mediated overexpression of NF-YA in mouse primary fibroblasts and NIH3T3 leads to induction of apoptosis, which is dependent on p53 and E2F1. Elimination of NF-YA by shRNA, differently for the inactivation of the HFD subunits, also leads to apoptosis (Benatti *et al.,* 2011). Thus, there are specificities in NF-YA, whose levels are indeed highly constrained. In particular, NF-YA appears to play an ambiguous role in apoptosis, regulating both apoptotic –Bax, Apaf1, FasL, Bim- as well anti-apoptotic -Bcl2- genes (Hughes *et al.,* 2011; Benatti *et al.,* 2008). Furthermore, the p53-activated lncRNA PANDA, produced as an anti-sense transcript from the p21 locus, mediates anti-apoptotic signals by interacting directly with NF-YA (Hung *et al.,* 2011; Subramanian *et al.,* 2013). The effect on *E2f1* transcription observed here is in line with previous reports (Gurtner *et al.,* 2010), and with the frequent co-occurrence of the two TFs in genomic locations and regulation of common sets of genes (Elkon *et* al., 2003; Linhart *et al*., 2005; Caretti *et al*., 2003; Dolfini *et al.,* 2012). These phenomena can now be studied in detail with the reagents developed here.

The presence of CPPs in NF-YA is a potentially relevant point for practical purposes, since TAT-NF-YAs was employed to increase the stem cells pools in HSCs preparations used for bone marrow transplantation –BMT (Zhu *et al.,* 2005). In theory, a TAT-less NF-YAs could be directly used for these experiments. BMT is the treatment of choice in many malignancies in which the bone marrow has been infiltrated, making patients prone to relapse when treated with conventional therapy. One of the caveats in using NF-YA in such experiments is that it could boost proliferation of the residual cancer stem cells, in addition to the normal HSCs. Currently, there are no data as to the function of NF-YA in cancer stem cells, but the differences in NF-YA stability observed here might hint at a lesser impact, or indeed growth suppression, on cancer cells. This is an encouraging observation to pursue experiments on primary cells *in vitro*, for potential clinical development.

5 REFERENCES

Araki, D., Takayama, K., Inoue, M., Watanabe, T., Kumon, H., Futaki, S., Matsui, H., and Tomizawa, K. (2010). Cell-penetrating D-isomer peptides of p53 C-terminus: long-term inhibitory effect on the growth of bladder cancer. Urology *75,* 813-819.

Benatti, P., Dolfini, D.,Vigano`, A., Ravo, M., Weisz, A., Imbriano, C. (2011).Specific inhibition of NF-Y subunits triggers different cell proliferation defects. Nucleic Acids Research *39*, 5356–5368.

Benatti, P., Basile, V., Merico, D., Fantoni, L.I., Tagliafico, E., Imbriano, C. (2008). A balance between NF-Y and p53 governs the pro- and anti-apoptotic transcriptional response. Nucleic Acids Res. *36*, 1415-28.

Bhattacharya, A., Deng, J.M., Zhang, Z., Behringer, R., de Crombrugghe, B., Maity, S.N. (2003). The B subunit of the CCAAT box binding transcription factor complex (CBF/NF-Y) is essential for early mouse development and cell proliferation. Cancer Res *63*, 8167–8172.

Bungartz, G., Land, H., Scadden, D.T., Emerson, S.G. (2012). NF-Y is necessary for hematopoietic stem cell proliferation and survival. Blood, *119,* 1380-1389

Caretti, G., Salsi, V., Vecchi, C., Imbriano, C., Mantovani, R. (2003). Dynamic recruitment of NF-Y and histone acetyltransferases on cell-cycle promoters. J Biol Chem *278*, 30435–30440.

Crombez, L., Aldrian-Herrada, G., Konate, K., Nguyen, Q.N., McMaster, G.K., Brasseur, R., Heitz, F., Divita, G. (2008). A New Potent Secondary Amphipathic Cell–penetrating Peptide for siRNA Delivery Into Mammalian Cells Molecular Therapy *17*, 95–103.

Derossi, D., Joliot, A.H., Chassaing, G., Prochiantz, A., Vives, E., Brodin, P., Lebleu, B. (1994). The third helix of the *Antennapedia* homeodomain translocates through biological membranes. J. Biol. Chem. *269*, 10444–10450.

Derossi, D., Calvet, S., Trembleau, A., Brunissen, A., Chassaing, G., Prochiantz, A. (1996). Cell internalization of the third helix of the *Antennapedia* homeodomain is receptor-independent. J. Biol. Chem. *271*, 18188–18193.

Dietz, G.P., Bahr, M. (2004). Delivery of bioactive molecules into the cell: the Trojan horse approach. Mol. Cell. Neurosci. *27*, 85–13.

Dolfini, D., Gatta, R., Mantovani, R. (2012). NF-Y and the transcriptional activation of CCAAT promoters. Critical Reviews in Biochemistry and Molecular Biology *47*, 29-49.

Dolfini, D., Minuzzo, M., Pavesi, G., Mantovani, R. (2012). The Short Isoform of NF-YA Belongs to the Embryonic Stem Cell Transcription Factor Circuitry. STEM CELLS *30*, 2450–2459.

Domashenko, A.D., Danet-Desnoyers, G., Aron, A., Carroll, M.P., Emerson, S.G. (2010). TAT-mediated transduction of NF-Ya peptide induces the ex vivo proliferation and engraftment potential of human hematopoietic progenitor cells. Blood *116*, 2676-2683.

Duchardt, F., Fotin-Mleczek, M., Schwarz, H., Fischer, R., Brock, R. (2007). A comprehensive model for the cellular uptake of cationic cell-penetrating peptides. Traffic, *8*, 848–866.

Eiríksdóttir, E., Konate, K., Langel, Ü., Divita, G., Deshayes, S. (2010). Secondary structure of cell-penetrating peptides controls membrane interaction and insertion. Biochimica et Biophysica Acta *1798,* 1119–1128.

Elkon, R., Linhart, C., Sharan, R., Shamir, R., Shiloh, Y. (2003). Genome-wide in silico identification of transcriptional regulators controlling the cell cycle in human cells. Genome Res *13*, 773–780.

Farina, A., Manni, I., Fontemaggi, G., Tiainen, M., Cenciarelli, C., Bellorini, M., Mantovani, R., Sacchi, A., Piaggio, G. (1999). 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*, 2818-27.

Frankel, A.D., Pabo, C.O. (1988). Cellular uptake of the tat protein from human immunodeficiency virus. Cell *55*, 1189–1193.

Frontini, M., Imbriano, C., Manni, I. Mantovani, R. (2004).Cell-cycle regulation of NF-YC nuclear localization. Cell Cycle *3*, 217-222.

Futaki, S., Suzuki, T., Ohashi, W., Yagami, T., Tanaka, S., Ueda, K., Sugiura, Y. (2001). Arginine-rich peptides. An abundant source of membrane-permeable peptides having potential as carriers for intracellular protein delivery. J. Biol. Chem., *276*, 5836–5840.

Giovanni, A., Keramaris, E., Morrisi, E.J., Hou, S.T., O'Hare, M., Dysoni, N., Robertson, G.S., Slack, R.S., Park, D.S. (2000). E2F1 Mediates Death of B-amyloid-treated Cortical Neurons in a Manner Independent of p53 and Dependent on Bax and Caspase 3. The Journal Of Biological Chemistry *275*, 11553–11560.

Green, M., Loewenstein, P.M. (1988). Autonomous functional domains of chemically synthesized human immunodeficiency virus tat trans-activator protein. Cell *55*, 1179–1188.

Gurtner, A., Fuschi, P., Martelli, F., Manni, I., Artuso, S., Simonte, G., Ambrosino, V., Antonini, A., Folgiero, V., Falcioni, R., Sacchi, A., Piaggio, G. (2010). Transcription factor NF-Y induces apoptosis in cells expressing wild-type p53 through E2F1 upregulation and p53 activation. Cancer Res. *70*, 9711-9720.

Gurtner, A., Manni, I., Fuschi, P., Mantovani, R., Guadagni, F., Sacchi, A., Piaggio, G. (2003). Requirement for down-regulation of the CCAAT-binding activity of the NF-Y transcription factor during skeletal muscle differentiation. Mol Biol Cell. *14*, 2706-2715.

Hughes, R., Kristiansen, M., Lassot, I., Desagher, S., Mantovani, R. and Ham, J. (2011). NF-Y is essential for expression of the proapoptotic bim gene in sympathetic neurons. Cell Death and Differentiation *18*, 937–947.

Hung, T., Wang, Y., Lin, M.F., Koegel, A.K., Kotake, Y., Grant, G.D., Horlings, H.M., Shah, N., Umbricht, C., Wang, P., Wang, Y., Kong, B., Langerød, A., Børresen-Dale, A.L., Kim, S.K., van de Vijver, M., Sukumar, S., Whitfield, M.L., Kellis, M., Xiong, Y., Wong, D.J., Chang, H.Y. (2011). Extensive and coordinated transcription of noncoding RNAs within cell-cycle promoters. Nat Genet. *43*, 621-629.

Kabe, Y., Yamada, J., Uga, H., Yamaguchi, Y., Wada, T., Handa, H. (2005). NF-Y is essential for the recruitment of RNA polymerase II and inducible transcription of several CCAAT box-containing genes. Mol Cell Biol. *25*, 512-22.

Kahle, J., Baake, M., Doenecke, D., Albig, W. (2005). 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*, 5339-54.

Li, X.Y., Hooft van Huijsduijnen, R., Mantovani, R., Benoist ,C., Mathis, D. (1992). Intron-exon organization of the NF-Y genes. Tissue-specific splicing modifies an activation domain. J Biol Chem. *267*, 8984-90.

Linhart, C., Elkon, R., Shiloh, Y., Shamir, R. (2005). Deciphering transcriptional regulatory elements that encode specific cell cycle phasing by comparative genomics analysis. Cell Cycle *4*, 1788–1797.

Liu,S., Tobias,R., McClure,S., Styba,G., Shi,Q, Jackowski,G.(1997).Removal of endotoxin from re combinant protein preparations. Clin Biochem. *30*, 455-63.

Lu, Y.H., Dallner, O.S., Birsoy, K., Fayzikhodjaeva, G., Friedman, J.M. (2015). Nuclear Factor-Y is an adipogenic factor that regulates leptin gene expression. Molecular metabolism *4,* 392-405.

Mae, M. and Langel, U. (2006) Cell-penetrating peptides as vectors for peptide, protein and oligonucleotide delivery. Curr. Opin. Pharmacol. *6*, 509–514.

Manni, I., Caretti, G., Artuso, S., Gurtner, A., Emiliozzi, V., Sacchi, A., Mantovani, R., Piaggio, G. (2008). Posttranslational regulation of NF-YA modulates NF-Y transcriptional activity. Mol Biol Cell. *19*, 5203-5213.

Manni, I., Mazzaro, G., Gurtner, A., Mantovani, R., Haugwitz, U., Krause, K., Engeland, K., Sacchi, A., Soddu, S., Piaggio, G. (2001). NF-Y mediates the transcriptional inhibition of the cyclin B1, cyclin B2, and cdc25C promoters upon induced G2 arrest. J Biol Chem. *276*, 5570-5576.

Mantovani, R., Li, X-Y., Pessara, U., van Huijsduijnen, R., Benoist, C., Mathis, D. (1994). Dominant negative analogs of NF-YA. J. Biol. Chem. *269*, 20340-20346.

Mantovani, R., Pessara, U., Tronche, F., Li, X-Y., Knapp, A.M., Pasquali, J.L., Benoist, C., Mathis, D. (1992).Monoclonal antibodies to NF-Y define its function in MHC Class II and albumin gene transcription. EMBO J. *11*, 3315-3322.

Marziali, G., Perrotti, E., Ilari, R., Coccia, E.M., Mantovani, R., Testa, U., Battistini, A. (1999). 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*, 519-526.

Mojsin, M., Topalovic, V., Marjanovic Vicentic, J., Stevanovic, M. (2015). Transcription factor NF-Y inhibits cell growth and decreases SOX2 expression in human embryonal carcinoma cell line NT2/D1. Biochemistry (Mosc). *80*, 202-207.

Moroni, M.C., Hickman, E.S., Lazzerini Denchi, E., Caprara, G., Colli, E., Cecconi, F., Müller, H., Helin, K.. (2001) *Apaf-1* is a transcriptional target for E2F and p53. Nature Cell Biology *3*, 552 – 558.

Nardini, M., Gnesutta, N., Donati, G., Gatta, R., Forni, C., Fossati, A., Vonrhein, C., Moras, D., Romier, C., Bolognesi, M., Mantovani R. (2013). NF-Y is a sequence-specific transcription factor displaying histone-like DNA binding and H2B-like ubiquitination. Cell, *152*, 132-143.

Prochiantz, A. and Joliot, A. (2003). Can transcription factors function as cell-cell signalling molecules? Nature Reviews Molecular Cell Biology *4*, 814-818.

Roeder, G. E., Parish, J. L., Stern, P. L. and Gaston, K. (2004). Herpes simplex virus VP22–human papillomavirus E2 fusion proteins produced in mammalian or bacterial cells enter mammalian cells and induce apoptotic cell death. Biotechnology and applied biochemistry *40*, 157-165.

Subramanian, M., Jones, M.F., Lal, A. (2013). Long Non-Coding RNAs Embedded in the Rb and p53 Pathways. Cancers, *5*, 1655-1675.

Tong, Y., Merino, D., Nimmervoll, B., Gupta, K., Wang, Y.D., Finkelstein, D., Dalton, J., Ellison, D.W., Ma, X., Zhang, J., Malkin, D., Gilbertson, R.J. (2015). Cross-Species Genomics Identifies TAF12, NFYC, and RAD54L as Choroid Plexus Carcinoma Oncogenes. Cancer Cell *27*, 712–727.

Van Ginkel, P.R., Hsiao, K.M., Schjerven, H., Farnham, P.J. (1997). E2F-mediated growth regulation requires transcription factor cooperation. J Biol Chem *272*, 18367–74.

Xing, Y., Fikes, J.D., Guarente, L. (1993). Mutations in yeast HAP2/HAP3 define a hybrid CCAAT box binding domain. The EMBO Journal *12*, 4647-4655.

Xing, Y., Zhang, S., Olesent, J.T., Rich, A., Guarente, L. (1994). Subunit interaction in the CCAAT-binding heteromeric complex is mediated by a very short a-helix in HAP2. Proc. Natl. Acad. Sci. USA *91*, 3009-3013.

Yamanaka, T., Tosaki, A., Kurosawa, M., Matsumoto, G., Koike, M., Uchiyama, Y., Maity, S.N., Shimogori, T., Hattori, N., and Nukina, N. NF-Y inactivation causes atypical neurodegeneration characterized by ubiquitin and p62 accumulation and endoplasmic reticulum disorganization. Nature communications 5.

Zhu, J., Zhang, Y., Joe, G.J., Pompetti, R., Emerson, S.G. (2005). NF-Ya activates multiple hematopoietic stem cell (HSC) regulatory genes and promotes HSC self-renewal. PNAS vol. *102*, 11728–11733.

Zhu, X.S., Linhoff, M.W., Li, G., Chin, K.C., Maity, S.N., Ting, J.P. (2000). Transcriptional scaffold: CIITA interacts with NF-Y, RFX, and CREB to cause stereospecific regulation of the class II major histocompatibility complex promoter. Mol Cell Biol. *20*, 6051-61.

6 FIGURE LEGENDS

Figure 1. Transfections of recombinant NF-YA in HCT116.
(A) Western blot of NF-YA in nuclear extracts of HCT 116 cells treated with increasing concentration -2 nM, 10 nM and 50 nM- of recombinant NF-YAs and NF-YAl. Cells were collected 2 hours after treatments. (B) Kinetic analysis of NF-YAl in HCT 116 cells. LaminB was used as loading control.

Figure 2. Effect of Mutations in basic stretches of the NF-YA on CCAAT-binding.
(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 and NF-YC interactions, in CCAAT binding are reported. (B, C, D) EMSAs of NF-Y with increasing amounts of NF-YA and a fixed amount of NF-YB/NF-YC dimer. NF-YAl and YAl mutants are in 0,1:1, 0.3:1, 1:1 with the NF-YB/NF-YC dimer.

Figure 3. Entry of NF-YAl mutants in HCT 116 cells.
Western blot analysis of NF-YA in HCT116 nuclear extracts treated with 10 nM of recombinant NF-YAl and YAl mutants. Cells were collected at indicated time points after protein transfections. 10 µg of nuclear extracts were loaded on the gels; (-) control sample of untreated cells. The endogenous short form of NF-YA present in HCT116 –NF-YAs- and the transduced recombinant NF-YAl are indicated. LaminB was used as loading control.

Figure 4. Stability of NF-YAl in different cell lines.
Western blot analysis of NF-YA in nuclear extracts of different cell lines treated with 10 nM NF-YAl protein. Cells were collected at indicated time points after transfection; 10 µg of nuclear extracts were loaded. Kinetics in HeLa cells are from 0,5 to 6 hours; in U2OS, HT1080, HaCat and NIH3T3 cell lines from 0,5 to 24 hours; (-) control sample of untreated cells.

Figure 5. Growth curve of NIH3T3 cells treated with NF-YA.

NHI3T3 cells were treated with 10 nM (A) and 100 nM (B) of recombinant NF-YAl, NF-YAs and mutant YAl 23-33 in triplicate and collected 24 and 48 hours after treatments. Growth curve was determined by cell counts using the Trypan Blue exclusion method.

Figure 6. Effect of NF-YA transfection on expression of CCAAT-dependent genes.

NIH3T3 cells were treated with 10 nM of recombinant NF-YAl, NF-YAs and NF-YAl 23-33: after 24 hours, mRNA was extracted and analyzed by qRT-PCR. The transcripts of the indicated genes are shown. mRNA transcripts were normalized with housekeeping RSP15 expression levels (*p<0.1, **p<0.05).

Supplementary 1. Removal of LPS from protein preparations.

(A) SDS-page stained with Blue Coomassie of NF-YAs, NF-YAl and the indicated YAl mutants. 1 μg of recombinant proteins, produced from *E.coli* inclusion bodies, are loaded; Bovine Serum Albumin (BSA) was used as reference. (B) Western blot analysis of NF-YAs, NF-YAl and the indicated YAl mutants using antibodies against NF-YA (Upper Panel) and LPS lipid A (Lower Panel), before (-) and after (+) treatment with Triton X114.

Supplementary 2. Dominant negative effect of A2 mutants.

EMSA of competition of the NF-Y trimer with increasing amounts of NF-YAl, YAl33, YAl34 and YA23 recombinant proteins. A constant concentration of NF-YAl at a 1:1 ratio with NF-YB/YC was mixed with growing concentrations of YAl mutants (NF-YA:YAl mutants ratio of 1:1, 1:3, 1:10). The labelled CCAAT box probe was used at a 20nM concentration.

Supplementary 3. Efficiency of protein transfections with NF-YA mutants.

Quantifications of Western blot analysis of transfections of NF-YAl and different mutants (See Fig. 3). The data were obtained using Image Lab software (BioRad), normalized with LaminB and related to the amount of transfected NF-YAl –or mutant- present at the 0.5 hours time-point.

Supplementary 4. Measurement of NF-YA stability in different cell lines.

Quantifications of Western blot analysis of NF-YAl transfections in different cell lines (See Fig. 4). The data were obtained using Image Lab software (BioRad) and normalized with LaminB and related to the amount of transfected NF-YAl present at the 0.5 hours time-point.

Supplementary 5. Expression of NF-YA in NIH3T3.

Western blot analysis of NF-YA (Upper Panel) and NF-YB (Lower Panel) in nuclear extracts of NIH3T3 cells treated with 10 nM and 100 nM of NF-YAl, NF-YAs and YAl 23-33. Cells were collected 24 hours (A) an 48 hours (B). 15 μg of nuclear protein extracts were used.

Supplementary 6. Sequences of the primers used in this study are shown.

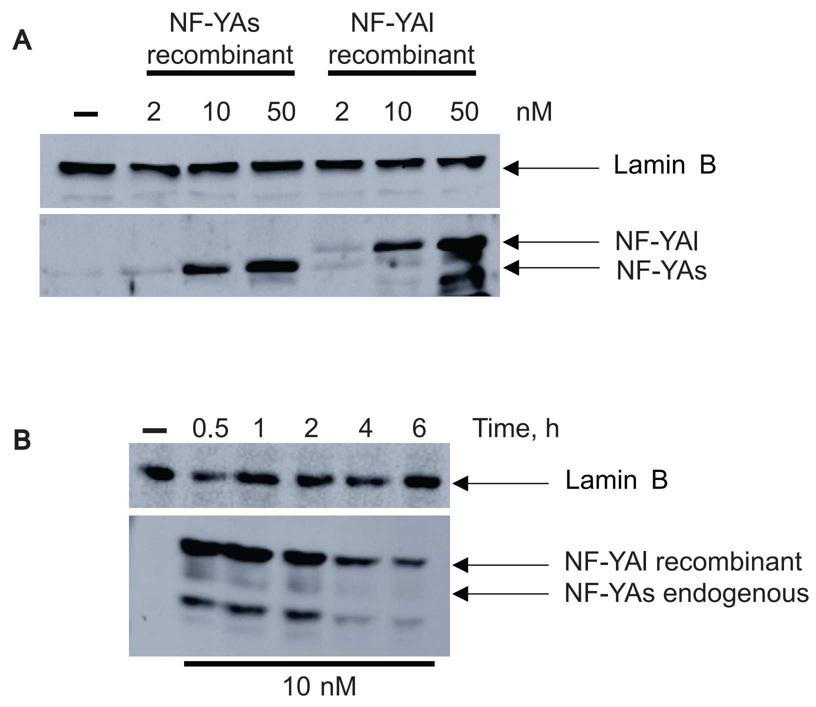


Figure 1

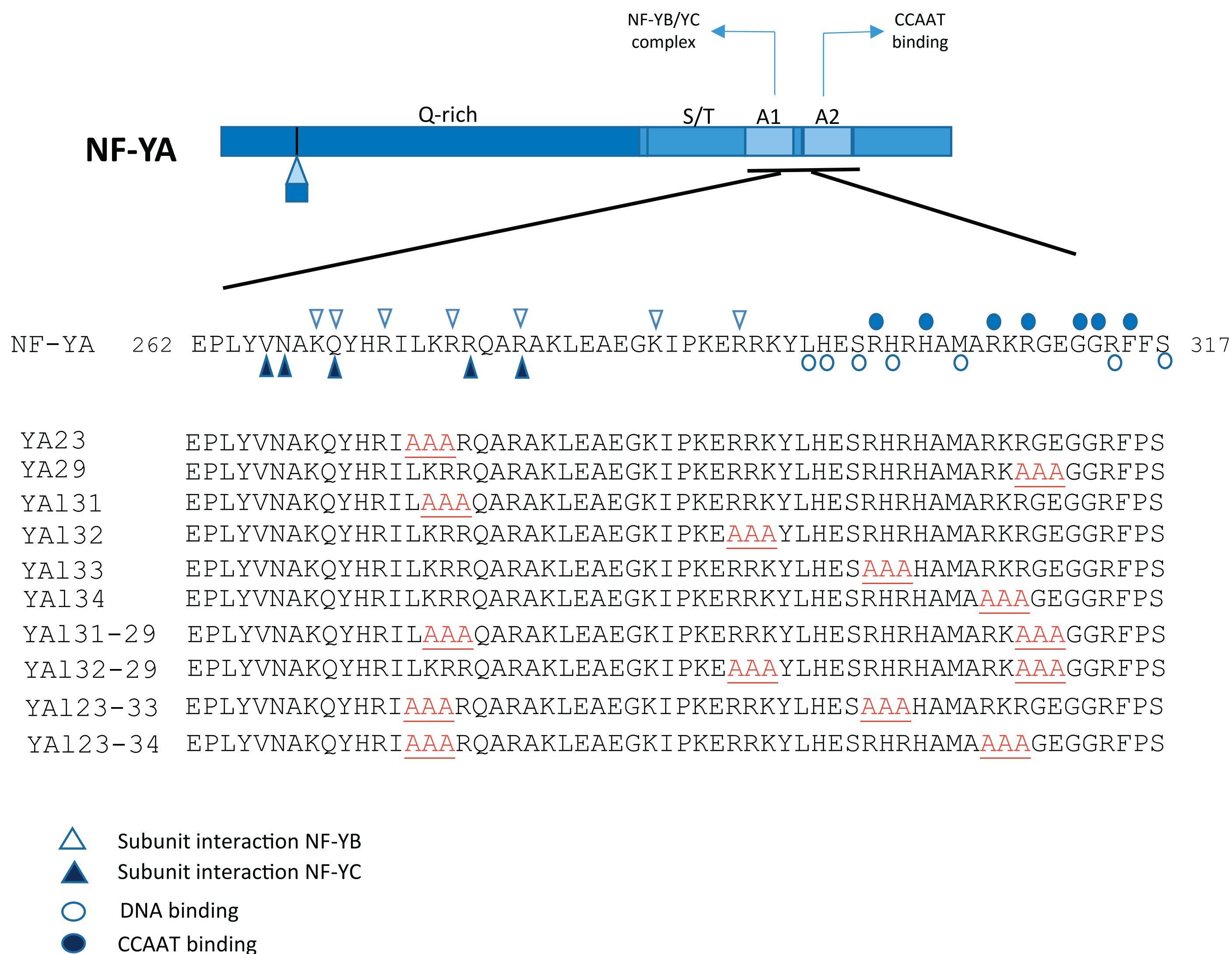
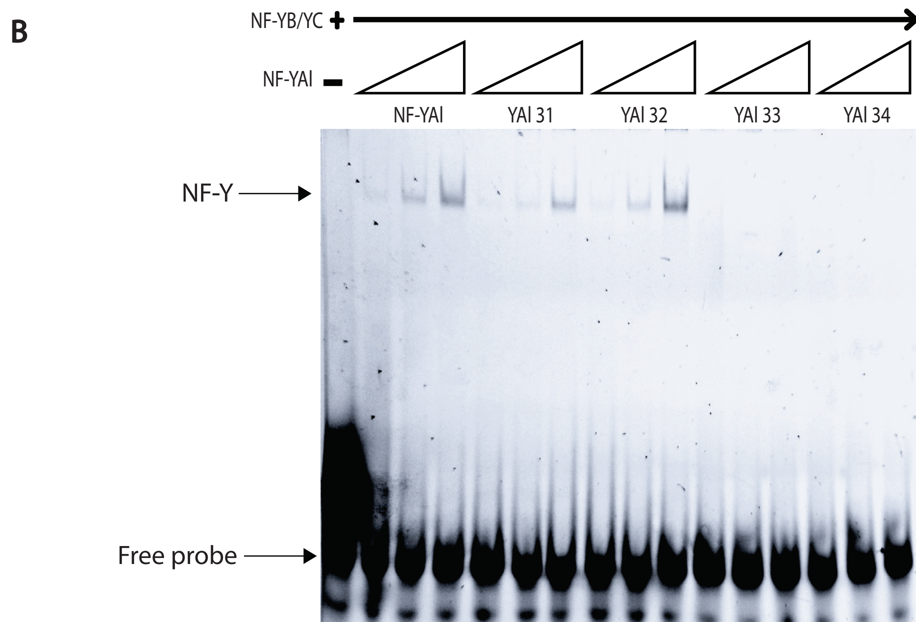
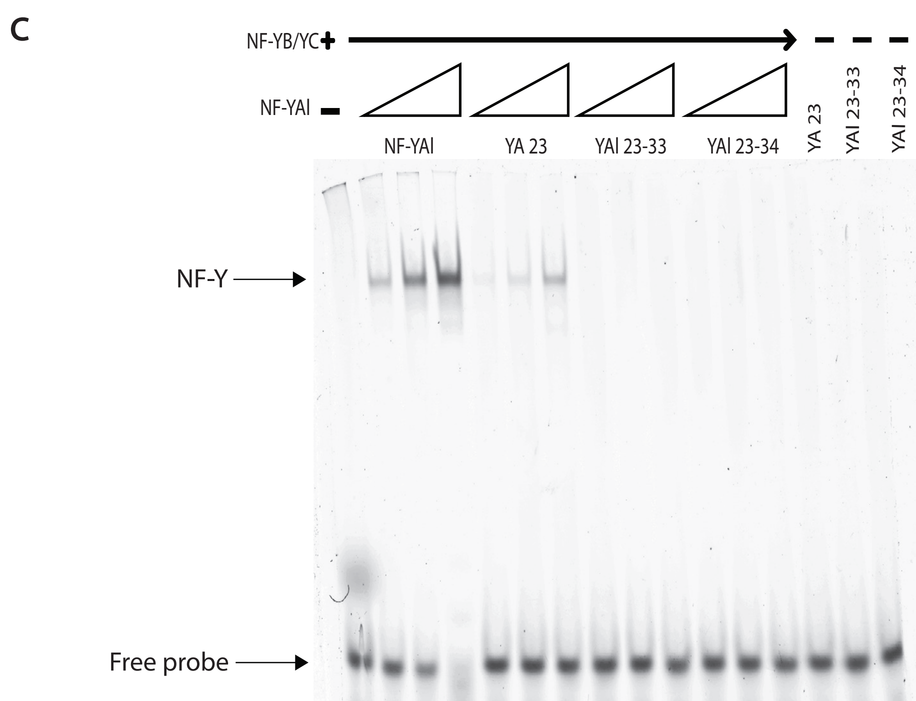


Figure 2A





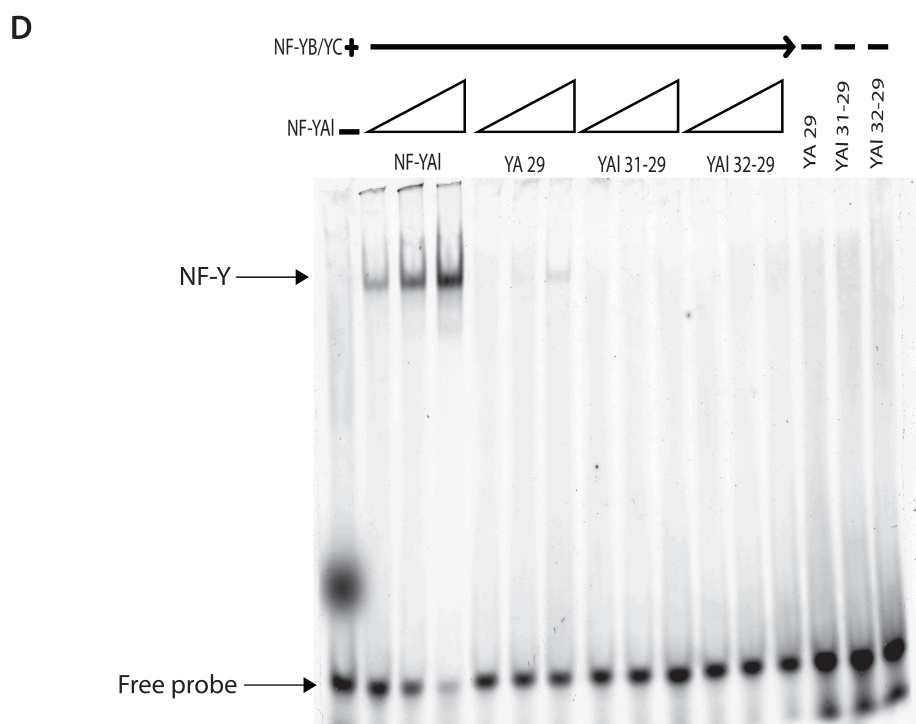


Figure 2B,C,D

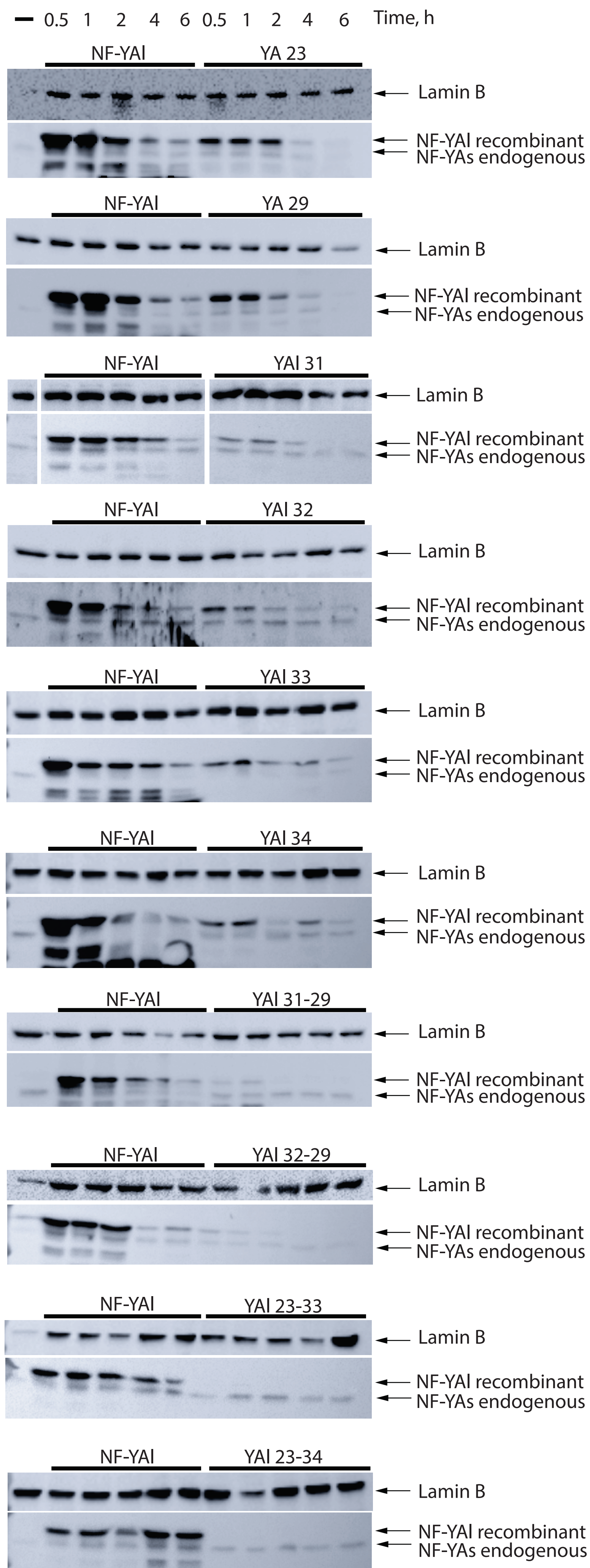


Figure 3

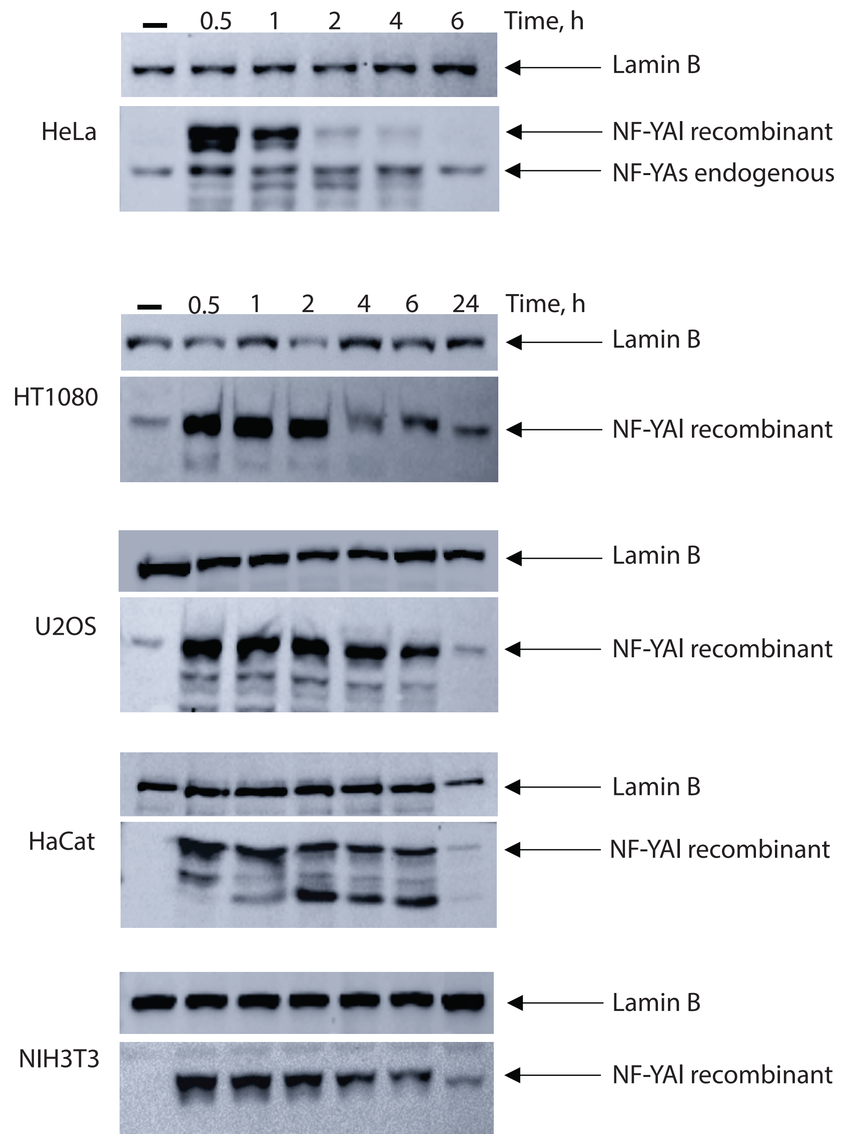
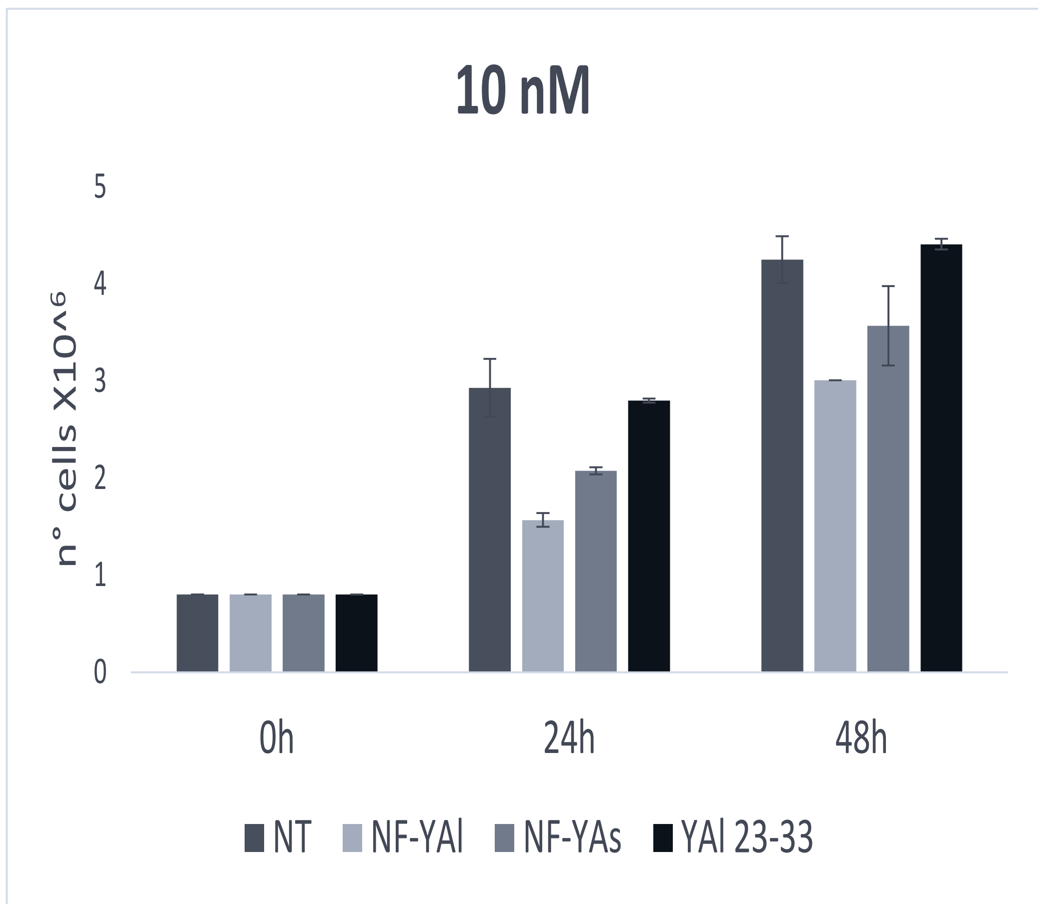


Figure 4



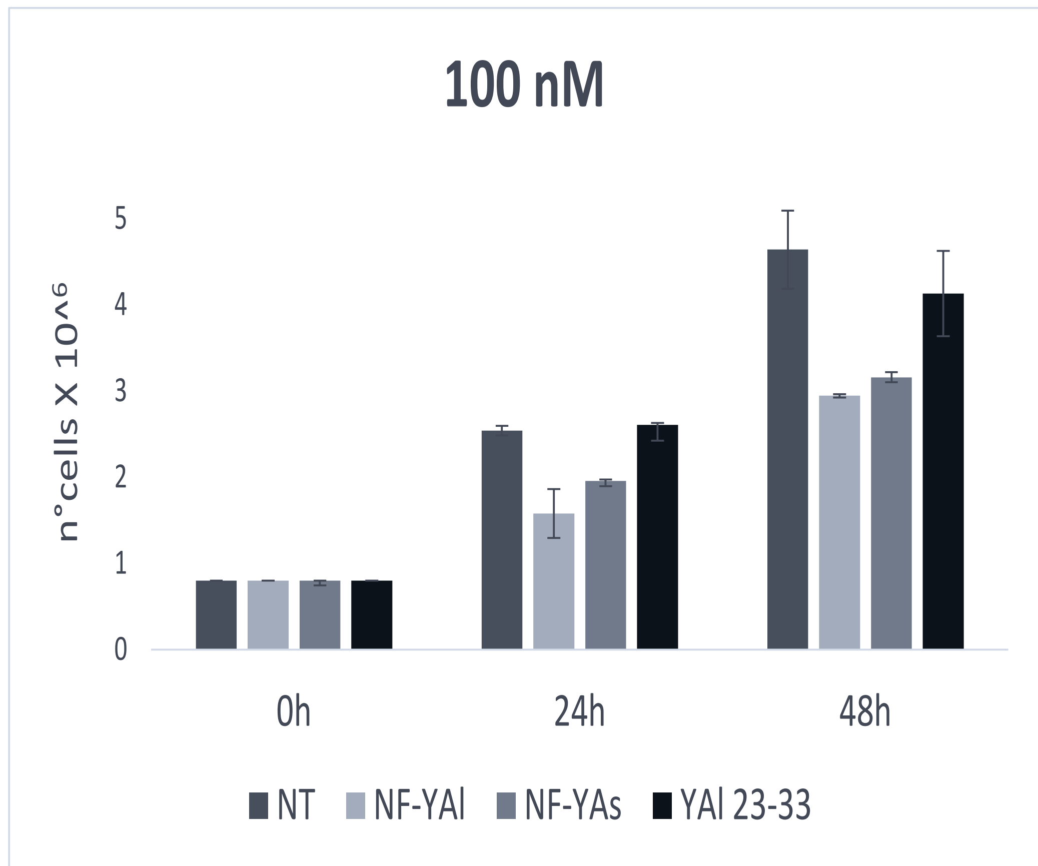


Figure 5

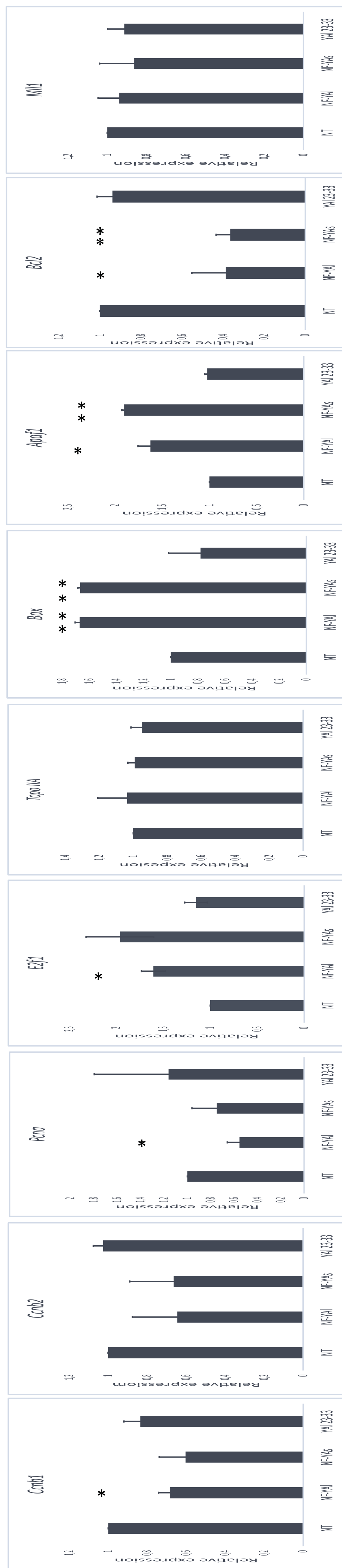
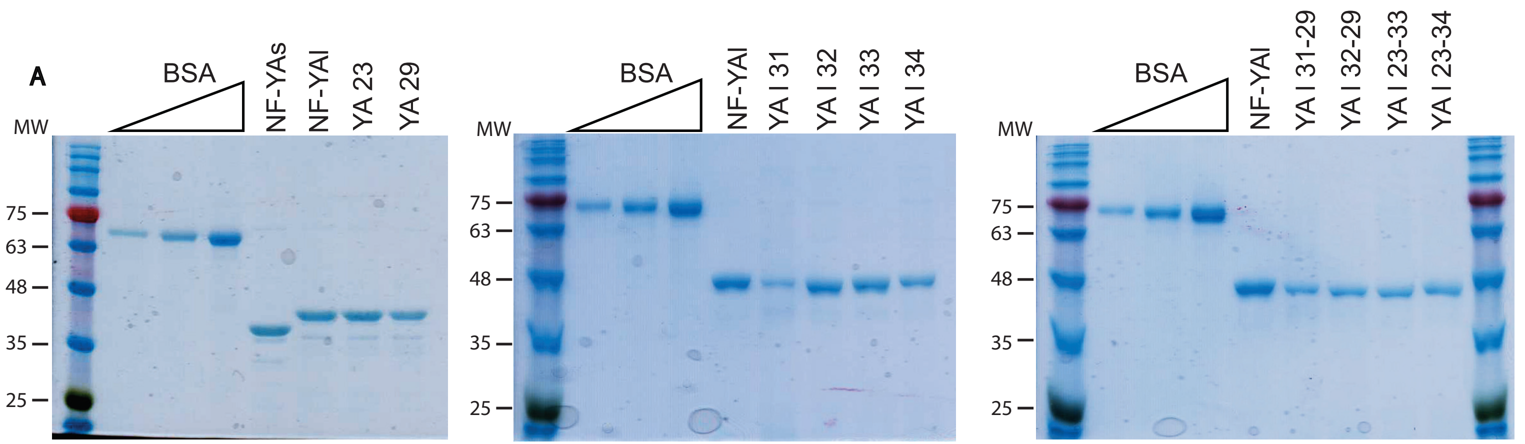
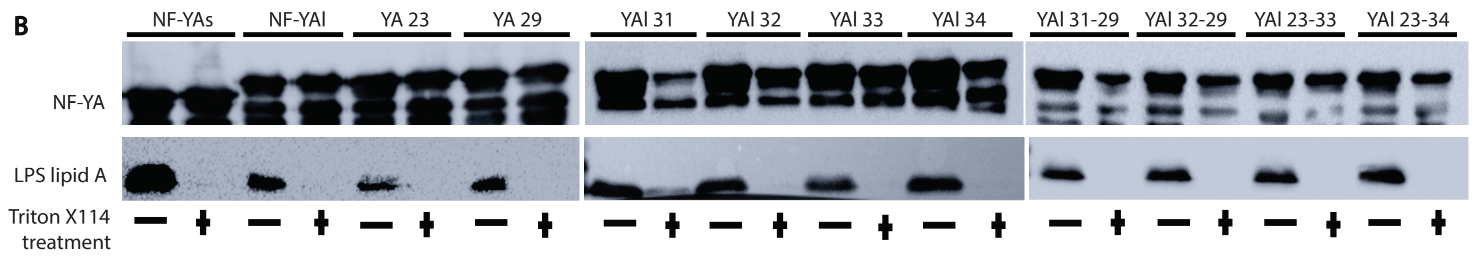
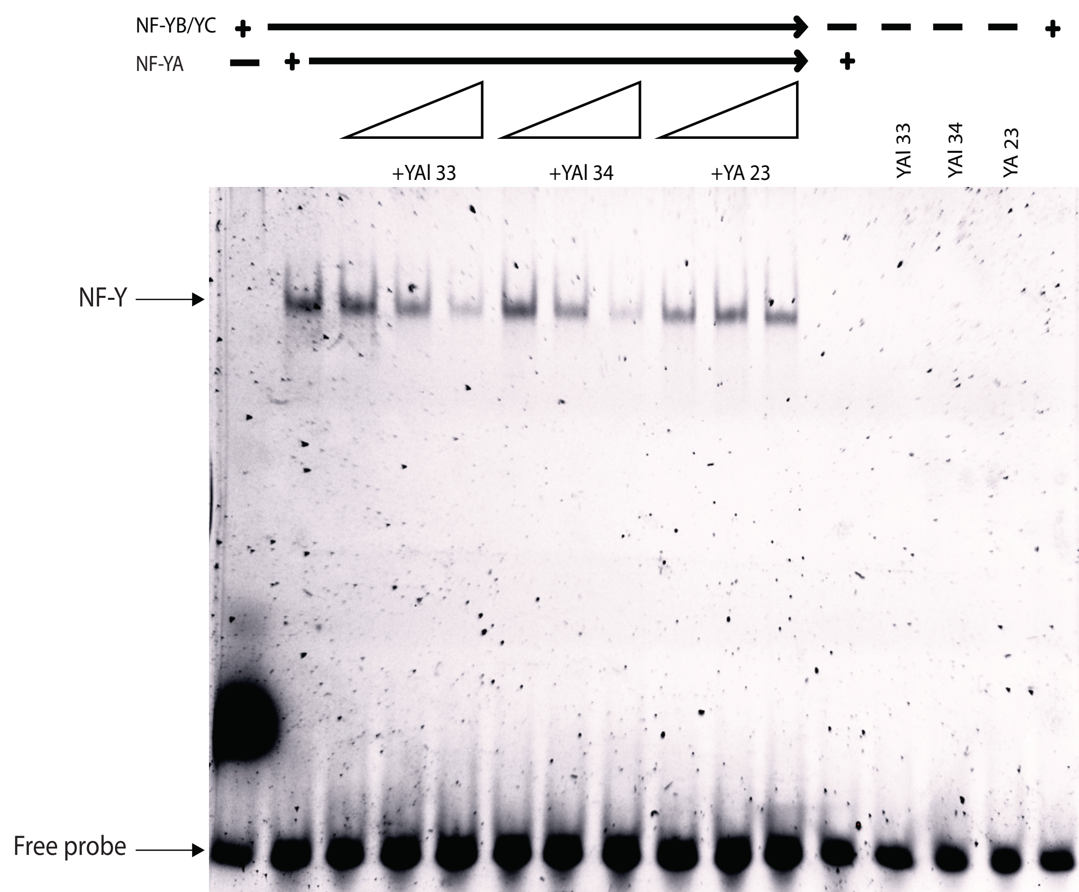


Figure 6

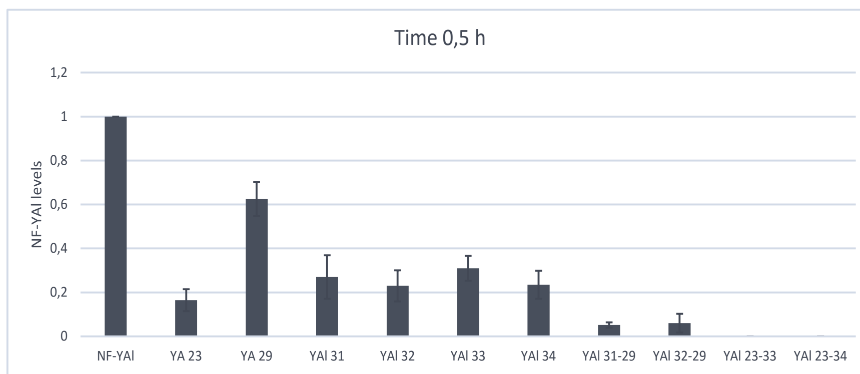


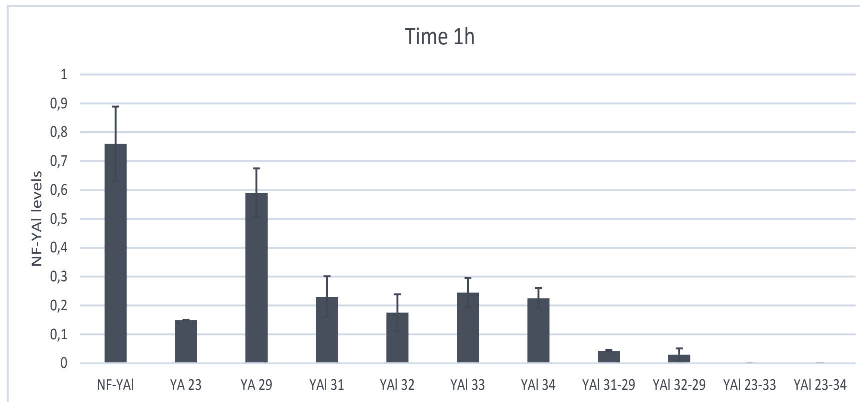


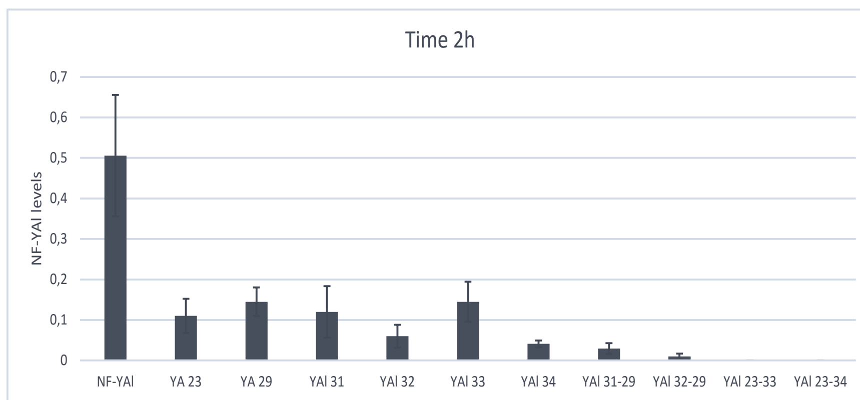
Supplementary 1

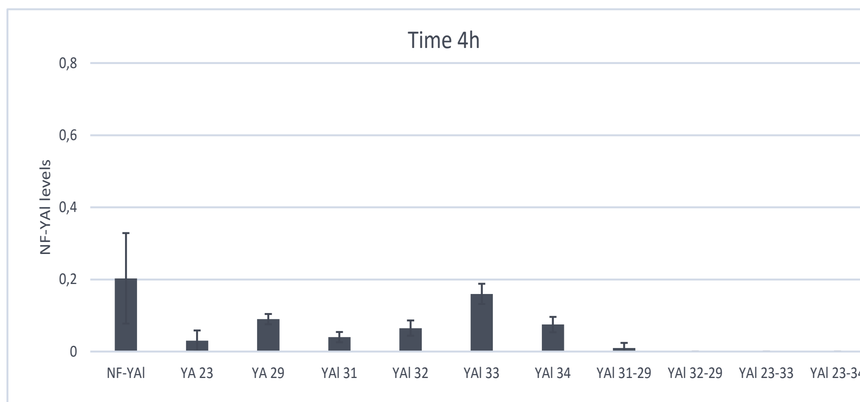


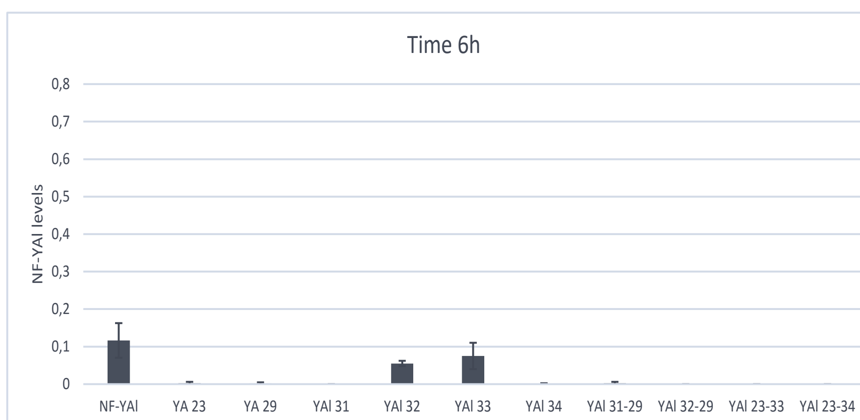
Supplementary 2



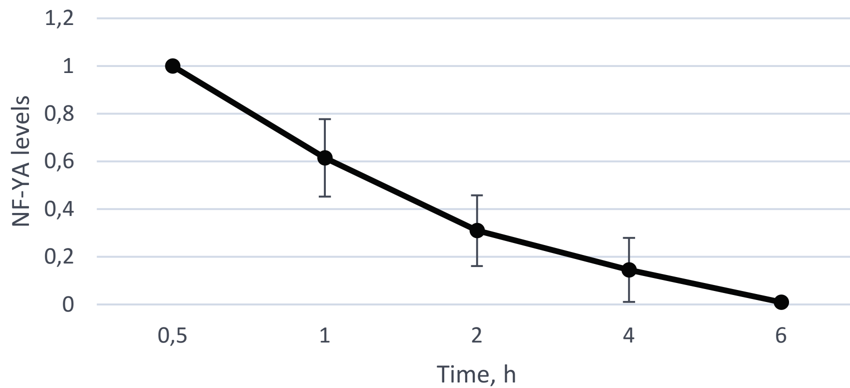


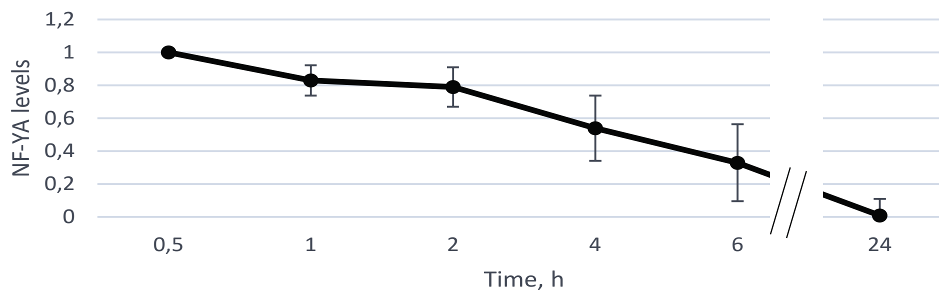


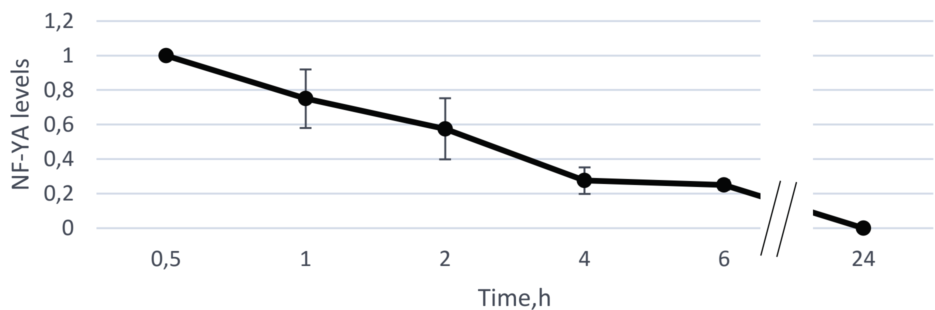


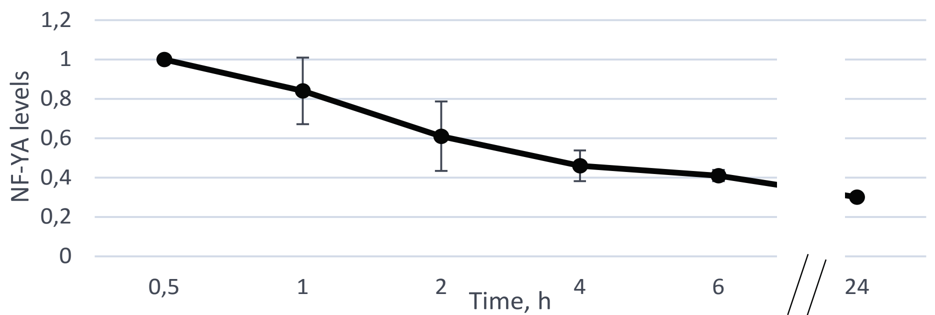


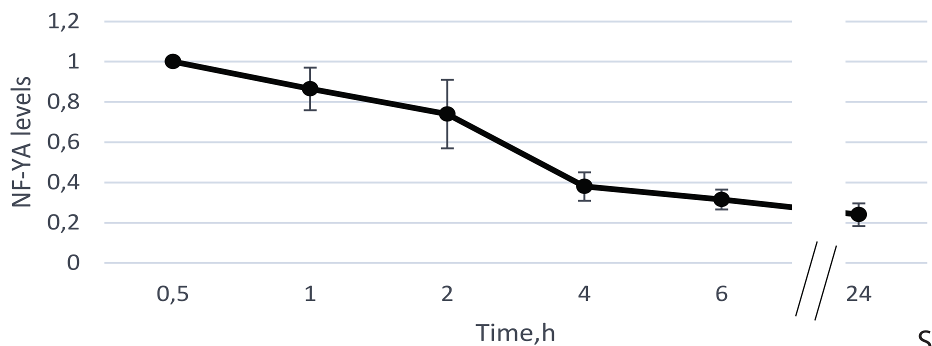
Supplementary 3



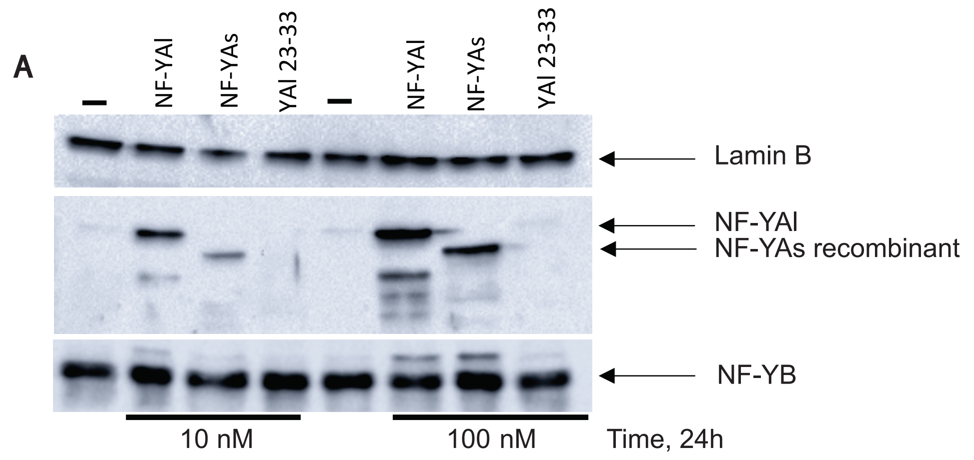


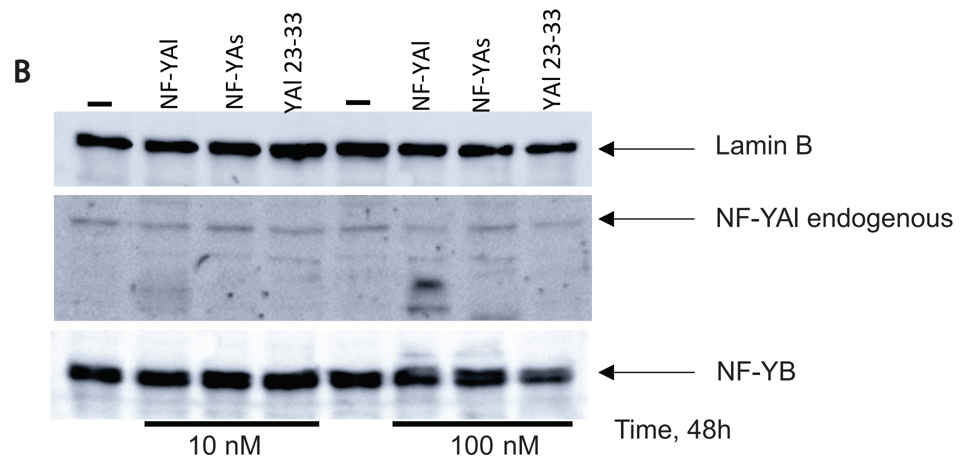






Supplementary 4





Supplementary 5

CCNB1	(F) 5’-CAGGAAGCAGGGAGTCTTCA-3’ (R) 5’-CATGCTGGACTACGACATGG-3’
CCNB2	(F) 5’-GGAAGCCAGGAGCAGAGC-3’ (R) 5’-CCTGGTGGACTGGCTGGT-3’
PCNA	(F) 5’-AGGAGACAGTGGAGTGGCTTTTG-3’ (R) 5’-GAGAGCTTGGCAATGGGAACA-3’
TOPO IIA	(F) 5’-ATCCACGTAACTGCGGAATCCT-3’ (R) 5’-GATGGATAACATGGGGAGAGCTG-3’
E2F1	(F) 5’-TCAAGCCGCTTACCAATCCC-3’ (R) 5’-CGGTGTCGTTGACCTGAACT-3’
APAF1	(F) 5’ –GATGGCCTGTGTGGATCTCT-3’ (R) 5’- GAGGATGTGGAGGTGATCGT-3’
BAX1	(F) 5’-CCCCAGTTGAAGTTGCCATC-3’ (R) 5’-CACCAAGAAGCTGAGCGAGT-3’
BCL2	(F) 5’- TCCTCCATCCCTTCATCCTCC-3’ (R) 5’- GGGGAGCAACATTCATCAGC-3’
MLL1	(F) 5’-TGACAATGTGCTTCTGCCCATC-3’ (R) 5’-CGCTCCATCCAGACAGACAAG-3’

Supplementary 6

Report

Preliminary analysis of recombinant NF-YA activity in mESCs and hints about the role of NF-Y in hiPSCs circuitry.

1. INTRODUCTION

Embryonic stem cells (ESCs) have the extraordinary feature to be indefinitely propagated *in vitro* in an undifferentiated state, retaining the ability to differentiate and generate all somatic adult tissues under the appropriate differentiation signals. Master transcription factors maintain stem cells identity by regulating the expression profiles of crucial genes that are then transmitted through cell divisions. Among these, SOX2, OCT4, and KLF4 are necessary in cell reprogramming to induced pluripotent stem cells (iPSCs) (Takahashi and Yamanaka 2006).

The TFs SOX2, OCT4 and NANOG create a regulatory feedback circuitry maintaining pluripotency both in human and mouse ESCs (Young, 2011). These three master embryonic transcription factors not only regulate each other but also are crucial for the activation of other pluripotency factors and for the repression of differentiation pathways.

Among the battery of TFs involved in stemness circuitry, the CCAAT binding complex, NF-Y, was identified and described as an essential component of the core pluripotency network (Oldfield *et al.,* 2014). NF-Y is composed by NF-YA, considered the limiting and regulatory subunit of the trimer, NF-YB and NF-YC subunits, containing a histone-fold domain (Romier *et al.,* 2003).

NF-Y plays a very important role in early mouse embryonic development. Briefly, as demonstrated by Bhattacharya, NF-YA heterozygous mice are alive and fertile while, NF-YA null mice, die at 8.5 dpc (Bhattacharya *et al*., 2003). Different studies demonstrated the role of NF-Y as important regulator of key stem cells transcription factors: the first refers to the retroviral infection of NF-YA subunit in mouse hematopoietic stem cells (HSCs). In this study, Zhu demonstrated that HSCs, overexpressing NF-YA subunits, are able to repopulate the bone marrow of immunocompromized animals enhancing the expression of transcription of several HOX genes and the activation of Notch and Wnt signaling pathways. Therefore, NF-YA is a potent cellular regulator of HSC self-renowal. Similarly, Domashenko, using TAT-fusion NF-YA protein, obtained a better engraftment and repopulation of bone marrow derived HSC in mice *ex vivo* (Domashenko *et al*., 2012). Moreover, the treatment of mESCs with TAT-NF-YA protein maintains cells in a pluripotent state in absence of leukemia inhibitory factor (LIF) as demonstrated by the activation of CCAAT stemness genes (Dolfini *et al*., 2012).

Although a TAT-fusion protein was a successful strategy to deliver a functional NF-YA protein in the nucleus, we discovered that NF-YA is able to translocate into the cells in a TAT-independent manner. The characterization of NF-YA as protein owning cell penetrating peptide (CPP) properties, was achieved by mutational analysis of basic amino acids in its evolutionary conserved domain.
NF-YA exists in two different isoforms, the long and the short, that differ for the presence of 28 amino acids codified by exon 3 of the longer isoform, in the Q-rich trans-activation domain. The conserved C-terminal domain of the protein is composed by two alpha helices, A1 and A2, rich in basic amino acids (Nardini *et al.,* 2013). We demonstrated that specific basic aminoacidic stretches, in the two helices, are important to confer to the protein the property of CPP (Chiaramonte *et al.,* manuscript in preparation in this thesis).
The identified NF-YA skills to penetrate cells, its proved stability and functionality in non-transformed cells, prompted us to investigate the possibility to use recombinant NF-YA as tool to study its activity also in mESCs.
In this report I will describe data obtained from mESCs treatments with recombinant NF-YA and I will show preliminary results about a new project concerning the study of NF-Y in human stem cells.

2. MATHERIALS AND METHODS

2.1 Generation of human induced pluripotent stem cells and embryo bodies formation.

1 x 10^5 human dermal fibroblasts were electroporated with a combination of episomal plasmid vectors including SOX2, KLF4, c-MYC, OCT4, LIN28 and NANOG using Nucleofector® Kit (Amaxa) plated in 10 cm dish in culture medium (EMEM, 15% FBS, 1% NEAA, 2 mM L-glutammine).
We trypsinized transfected cells on day 7 and reseeded them on feeder layer in embryonic stem cells medium (DMEM F12, 20% knockout medium, 2mM L-glutammine, 100 units/ml penicillin and 100 μg/ml streptomycin, 1% NEAA, 1% sodium pyruvate, 0.1 mM β-Mercaptoethanol, 20 ng/ml b-FGF) at 37°C in a humidified, 5% O_2, 5% CO_2 atmosphere.
Cells were maintained in embryonic stem cells medium until that small cell colonies became visible. Cell colonies were mechanically picked and seed in a feeder free layer and amplified on matrigel coated plates in human stem cells medium (STEMGENT).
Embryo bodies were formed by suspension culturing of hIPSCs in human Embryo Bodies Medium (DMEM F12, 20% Knock-out medium, 1% NEAA, 1mM L-glutamine, 100 units/ml penicillin and 100 μg/ml streptomycin, 1 mM sodium pyruvate, 0.1 mM β) at 37°C in a humidified 5% CO_2 atmosphere for 7 days.

2.2 Lentivirus production and cell infection.

pLKO.1 non-target short hairpin RNA (shRNA) (SHC002),
shRNA-NF-YA, targeting exon 6 (CCATCGTCTATCAACCAGTTA),
shRNA-NF-YB targeting exon 5 (GCTATGTCTACTTTAGGCTTT), were designed by Sigma-Aldrich. For lentivirus production, the shRNA vector plasmid (20 µg) and second-generation packaging plasmids (5 µg of pMD2-VSVG and 15 µg of pCMV_R8.91) were co-transfected into HEK293T cells. Lentiviral particles were collected from supernatant 36 hours after transfection, centrifuged at 1500 rpm to remove cell debris for 5', 0.45 µm filtered and stored at -80°C.
60% confluent hiPSCs were infected with two ml of collected virus in 6 multi well plates and harvested 72 hours for protein total extract and RNA extraction.
The experiments were repeated twice.

2.3 mESCs culture.

Mouse ESCs were maintained in glasgow minimum essential medium (Sigma-Aldrich) supplemented with 10% ESC qualified fetal bovine serum (Gibco), 1 mM L-glutamine, 0.1 mM non-essential amino acids, 100 units/ml penicillin, 100 µg/ml streptomycin, 1 mM sodium pyruvate, 0.1 mM 2-mercaptoethanol and 1,000 units/ml of recombinant leukemia inhibitory factor (LIF) (Gene Spin).

2.4 Recombinant protein production and transduction.

NF-YAl, NF-YAs and mutant YAl 23-33 were purified as previously described (Chiaramonte *et al.,* manuscript in preparation in this thesis).
Protein concentration was determined by the Bradford method (Bio-Rad Protein Assay).
Cells were treated with the indicated amount of recombinant proteins. Cells were washed with PBS and recombinant proteins were added in DMEM without serum for 30'. The treatment was repeated twice. After the second treatment cells were grown in ES medium and kept at 37°C in 5% CO_2 for 24 hours. Recombinant proteins were added at 24 hours intervals for four days.
Alkaline phosphatase assays were performed using a colorimetric system (SIGMA).
The experiments were repeated twice.

2.5 Western blot analysis.

Western blot analyses of nuclear and total extract were performed according to the standard procedures using the primary antibodies and peroxidase-conjugated antibody.

Primary antibodies used: anti-NF-YA (Santa Cruz), anti-NF-YB (GeneSpin), anti-KLF4 (GeneSpin), anti-SOX2 (Santa Cruz), anti-OCT4 (Santa Cruz), anti-GAPDH (Santa Cruz), ananti-Lamin B (Santa Cruz), anti-Vinculin (Santa Cruz). Visualization of images was obtained with the ChemiDoc^{TM} MP imaging system (BioRad).

2.6 RNA extraction and RT-PCR analysis.

Total RNA was extracted using Trizol reagent (Sigma) according to the manufacturer's instructions. 1 μg of RNA was retrotranscribed with SuperScript^{TM} reverse transcriptase (BioRad) and random primers. cDNA was diluted 1:20 with water. cDNA was analyzed in triplicate by q-PCR analysis and a comparative $2^{-\Delta\Delta Ct}$ was used to detect relative gene expression.

3. MAIN RESULTS AND DISCUSSION

3.1 Recombinant NF-YA protein transduction in mESCs.

Before testing recombinant NF-YA transfection in mouse embryonic stem cells, we evaluated the ability of the protein to translocate and enter *nuclei* of different cell lines. For all the treated mouse and human cell types, we observed a rapid uptake and accumulation of NF-YA in the *nuclei* at very low concentrations. One of the restriction use of recombinant NF-YA is due to the half-life of the protein inside to the cells. For many of the transformed cell lines tested, the protein disappears few hours (4-6h) after transduction making hard to evaluate possible effects of its overexpression (Chiaramonte *et al*., manuscript in preparation in this thesis). Curiously, we observed that the trend was different for non-transformed HaCaT and NIH3T3 cells that retain a relatively abundant amount of protein in the *nucleus* until 24 hours from treatment.

Therefore, we decided to investigate the stability of recombinant NF-YA also in mESCs.

Figure 1 shows the kinetic performed in order both to evaluate the capability of recombinant NF-YA to translocate mESCs and to determine the half-life of the protein after translocation.

As we alredy knew, by previous experiments, that recombinant NF-YAl and NF-YAs have the same behavior in terms of transduction efficiency and time of cell residence (half-life), this experiment was performed using the long isoform of NF-YA. Protein transduction experiments were performed providing recombinant NF-YAl to mESCs in DMEM medium without serum, one time, every 30 minutes, for one hour. After treatments, DMEM was removed, cells were cultured in complete mESCs medium and collected at different time points. Western blot analysis shows that recombinant NF-YA translocates efficiently and rapidly at very low concentrations (20 nM) in the *nuclei* where it is can be detected until 24 hours from treatments. The kinetic carried out up to 48 hours, revealed that

the protein completely disappeared two days after treatments. Quantification of these results are shown in Fig 1B.

3.2 Effects of recombinant NF-YA on mESCs.

Once established that recombinant NF-YA enters mESCs *nuclei* efficiently and rapidly and that it is present until 24 hours from treatments, we went to evaluate its molecular activity. In particular, we verified whether it was able, similarly to TAT-NF-YAs, to maintain mESCs in an undifferentiated state in absence of LIF (Dolfini *et al*., 2012). To note, experiments of TAT-NF-YA protein transduction were performed only with the short isoform of NF-YA.

In this work, we wanted to evaluate whether the transduction of short and the long isoform of NF-YA showed different effects in terms of stemness maintaining. For this reason, mESCs were treated with recombinant NF-YAs, NF-YAl and with the mutant of NF-YA unable to penetrate cells -YAl 23-33-. Every 24 hours, recombinant proteins were provided to mESCs as described above. After treatments, mESCs were grown in ESCs medium supplemented or not with LIF (Fig.2A, schematic representation of cell treatments). These treatments were conducted for 4 days and the first observations about the effects of NF-YA transduction were marked by cell morphology.

LIF is necessary for ESCs growth and pluripotency state stability and it is known that mESCs, in healthy conditions, grow as nice small colonies. We observed that control and YAl 23-33 treated cells, cultured for 4 days in withdrawal of LIF, lost their typically growing aspect. Daily treatments with NF-YA, in particular with recombinant NF-YAs, in absence of LIF, instead, maintained cells with their classic morphology (Fig. 2B).

Moreover, we performed alkaline phosphatase assay as a further indication: in presence of LIF, cells treated with NF-YAs, NF-YAl, YAl 23-33 and control cells were positive to alkaline phosphatase assay; in absence of LIF positivity was weaker for control and YAl 23-33 treated cells while it was maintained for cells treated with NF-YAl and much more for that treated with NF-YAs (Fig. 2C).

These findings pushed us to check the expression levels of stem cell genes by qRT-PCR (Fig. 2D).

What emerged from mRNA analysis is that all considered stemness genes maintain the same expression levels when cells are cultured in presence of LIF. The *scenario* changes when LIF is removed from culture medium. The expression of *Nanog, Klf4* and *Klf5* are maintained to high levels (near to normal expression) for cells treated with recombinant NF-YAs and NF-YAl while they are dramatically reduced for control and YAl 23-33 treated cells.

The expression levels of *Oct4* and *Sall4* are similar for all the treatments performed in absence of LIF while we did not observe a reduced expression of *Sox2*. This could be explained with the major

stability of mRNA of these genes therefore, treatments for over several days, will be performed to evaluate this point.

The *Ccnb2* expression level, instead, is drastically affected in control and YAl 23-33 treated cells while the transfection of NF-YAl and in particular NF-YAs maintains high its expression. This is in agreement with what was observed: mESCs treated with NF-YAs grow faster compared to the others.

These preliminary data show that:

i) recombinant NF-YA penetrates mESCs efficiently and rapidly at low concentrations (20 nM) and it is detectable until to 24 hours from treatment;

ii) when cells are cultured in absence of LIF, the short isoform of NF-YA seems to have a more pronounced effect, compared to the long one, on maintaining mESCs morphology and in promoting cell growth. However, we did not observe a relevant difference, between the two treatments, on the expression levels of analyzed stemness genes. In both cases, they are maintained to higher levels compared to the control cells. It is clear that other genes involved in pluripotency, as well as genes involved in differentiation, have to be analyzed.

The activity of the protein, highlighted by the modulation of important stemness and CCAAT containing genes, makes recombinant NF-YA a very important tool to study its effects in different cell systems avoiding problems related to classic methods of gene delivery. Moreover, this system allows to have controlled protein expression levels and, in general, these results lay the groundwork to use recombinant NF-YA also in other stem cell systems, such as human stem cells.

For this reason, in parallel to this study, we started to investigate NF-Y in human induced pluripotent stem cells.

3.3 Generation and characterization of human induced pluripotent stem cells (hiPSCs).

The clear involvement of NF-Y in mESCs circuitry and, in particular, the importance of NF-YA subunit to maintain their pluripotent state, pushed us to turn the attention on the role of NF-Y also in human stem cells.

Considering the clear ethical problems for the use of human embryons and the difficulties to get human embryonic stem cells, human induced pluripotent stem cells were used as model to start this study.

Human dermal fibroblasts were electroporated with a mix of episomal plasmid vectors containing the Yamanaka factors: SOX2, KLF4, OCT4, L-MYC, LIN28 and the shp53 that was demonstrated to increase the efficiency of hiPSCs generation (Keisuke Okita *et al.*, 2011). 1 x 10^5 electroporated cells

were plated in 10 cm dish, trypsinized after 7 days and reseeded onto a feeder layer (Fig. 3A, schematic representation of hiPSCs generation).

Cells were grown in ESC medium and colonies became visible two weeks after transfection; the single colonies were mechanically picked and seeded into a feeder free (matrigel coated) system.

The majority of obtained colonies were expandable, had a classic morphology of human ESCs characterized by large nuclei and scanty cytoplasm, and were positives to alkaline phosphatase assay (Fig. 3B).

qPCR analysis revealed that colonies obtained by reprogramming express the main stemness transcription factors: SOX2, OCT4, KLF4 and NANOG. The expression levels of stemness genes were normalized on GAPDH housekeeping gene (Fig. 3C).

We further characterized the pluripotency of obtained colonies by testing their differentiation potential. Along the hiPSCs differentiation process, embryo bodies (EBs) formation is a routine founder step to direct cell differentiation toward a specific lineage. EBs are three-dimensional aggregates of hiPSCs that mimic the structure of human developing embryos able to differentiate to all the three primary germ lines.

hiPSCs dissociated to single cells, were seeded into a ultra-low adherence petri dish in EBs medium; dissociated hiPSCs spontaneously aggregate to form spheroid bodies (Fig. 4A) further demonstrating that colonies obtained by reprogramming process are pluripotent stem cells able to start the differentiation process under appropriate *stimuli.*

3.4 From hiPSCs to EBs: a switch from short to long isoform of NF-YA

As mentioned before, NF-YA subunit exists in two different isoforms generated by an exon skipping event and different cell lines preferentially express one of them (Ceribelli *et al.,* 2009).

Western blot analysis in hiPSCs reveals that, similarly to mESCs, the short isoform of NF-YA is that expressed in hiPSCs (Fig. 4B); very interesting, we observed that after differentiation to EBs, NF-YAl becomes the mainly expressed isoform (Fig. 4C). SOX2, OCT4 and KLF4 protein expression levels have been used to monitor the differentiation process.

These data reflect what observed in mESCs: a shift from short to long isoform of NF-YA occurs when cells are committed to EBs (Dolfini *et al*., 2012) suggesting that the short isoform of NF-YA governs stemness circuitry while the long one drives the differentiation process.

3.5 NF-Y subunits silencing affect the expression of important stemness genes

In order to determine the role of NF-Y in regulation of hiPSCs stemness genes, NF-Y subunits (NF-YA and NF-YB) were silenced using lentiviral delivery of shRNAs. A non-target shRNA (sh-CTR)

was used as control vector. HiPSCs were collected 72 hours after lentiviral infection and the expression levels of NF-YA and NF-YB subunits were analyzed by western blot and qRT-PCR analysis (Fig. 5). We observed that at 72 hours from infection NF-YA and NF-YB protein subunits were depleted (Fig. 5A, 5C). qRT-PCR analysis highlight that after NF-YB silencing the expression levels of SOX2 are dramatically affected; a modest effect was observed for the expression of SOX3 and KLF4 while, the expression levels of KLF5, NANOG and OCT4 remain unchanged; GAPDH and PCNA were used as positive controls as they are housekeeping genes target of NF-Y (Fig.5 D). Interestingly, we observed that the silencing of NF-YA strongly affected the expression of all stemness genes analyzed (Fig. 5B).

Curiously, the expression level of NF-YA and in particular the long isoform of the subunit, increased after NF-YB silencing, *et vice versa,* as if there was a sort of compensation between the subunits, whose mechanism of regulation is still not well understood.

We speculated that the emerged predominant effect of NF-YA subunit on the expression levels of stemness genes compared to NF-YB depletion could be due to the different gene silencing levels obtained for the two proteins. A second hypothesis is that the two proteins may also cover different roles other that NF-Y trimer activity. The latter is supported by experiments carried out in other cell lines: NF-YA and NF-YB depletion leads different effects on HCT116 cells. While NF-YA silencing affects the progression of S phase and induces DNA damages, NF-YB deletion induces a delay in G2/M phase progression (Benatti *et al*., 2011).

4. CONCLUSIONS AND FUTURE PERSPECTIVES

The results reported in the first part of the work, regarding recombinant NF-YA transduction in mESCs, provide clear evidences that the protein is functional in this cell system because, similarly to TAT-NFYA, it is able to maintain the pluripotency properties of mESCs when they are grown in absence of LIF.

This effect may be considered as a reflection of a role of NF-YA in preventing cell exit from the stem pool within the culture. This result was observed to be typical of other TFs belonging to the core of embryonic stem cells circuitry: KLF4 and TCF3, another important TF for ESCs, show effects similar to NF-YA when overexpressed in mESCs (Tam *et al.,* 2008; Zhang *et al.,* 2010).

The use of CPPs as carriers for the transport of molecular drug inside to the cells is a developing field for clinical trials; a functional recombinant NF-YA gives the advantage to use a protein as near as possible to the endogenous one and we cannot exclude to undertake studies for future therapeutic and clinical applications.

Data obtained from hiPSCs highlighted a good parallelism among mESCs and hiPSCs regarding the expression of NF-YA isoform before and after differentiation. Moreover, we observed that after NF-YA subunit depletion, hiPSCs lost their pluripotency properties as demonstrated by the drastic reduction of stemeness genes expression levels.

In order to have a broader and complete view about NF-Y activity in human stem cells network, we want to proceed with gene expression profiles experiments of hiPSCs upon NF-YA and NF-YB depletion and with ChIP analysis in order to better understand the relationship between NF-Y and the other master stemness TFs.

Apart from Yamanaka factors, other TFs have been identified to enhance reprogramming of differentiated cells into iPSCs, among these, SALL4 (Wong *et al.,* 2008). SALL4, like NF-YA, exists in two different isoforms which relative abundance changes according to the pluripotent or differentiated state of the cells (Rao *et al.,* 2010). Moreover, it shows effects similar to NF-YA when transfected in HSCs (Yang *et al.*, 2011), it is involved in leukemogenesis and in other types of cancers (Gao *et al.,* 2011). The similes between NF-YA and SALL4 suggest us that NF-YA may be investigated for its potential role also in reprogramming process.

Finally, the use of recombinant NF-YA to overexpress NF-YA subunit and the induction of hiPSCs to differentiate toward a specific lineage will provide further indications about the role of NF-Y in stemness and differentiation processes.

6. References

Dolfini, D., Minuzzo, M., Pavesi, G., and Mantovani, R. (2012). The Short Isoform of NF-YA Belongs to the Embryonic Stem Cell Transcription Factor Circuitry. STEM CELLS *30*, 2450–2459.

Domashenko, A.D., Danet-Desnoyers, G., Aron, A. et al. (2012). TAT-mediated transduction of NF-Ya peptide induces the ex vivo proliferation and engraftment potential of human hematopoietic progenitor cells. Blood *116*, 2676–2683.

Gao, C., Kong, N.R., and Chai, L. (2011).The role of stem cell factor SALL4 in leukemogenesis. Crit Rev Oncog *16*, 117–127.

Oldfield, A.J., Yang, P., Conway, A.E., Cinghu, S., Freudenberg, J.M., Yellaboina, S., and Jothi, R.. Histone-Fold Domain Protein NF-Y Promotes Chromatin Accessibility for Cell Type-Specific Master Transcription Factors Molecular Cell Molecular Cell *55*, 708–722.

Rao, S., Zhen, S., Roumiantsev, S. et al (2010). Differential roles of Sall4 isoforms in embryonic stem cell pluripotency. Mol Cell Biol *30*, 5364–5380.

Romier, C., Cocchiarella, F., Mantovani, R., and Moras, D. (2003). The NF-YB/NF-YC structure gives insight into DNA binding and transcription regulation by CCAAT factor NF-Y. J Biol Chem. *10*, 1336-1345.

Takahashi, K. and S Yamanaka, S. (2006). Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors. Cell *126*, 663–676.

Tam, W.L., Lim, C.Y., Han, J. et al. (2008). T-cell factor 3 regulates embryonic stem cell pluripotency and self-renewal by the transcriptional control of multiple lineage pathways. Stem Cells *26*, 2019–2031.

Wong, C., Gaspar-Maia, A., Ramalho-Santos, M. et al. (2008). High-efficiency stem cell fusion-mediated assay reveals Sall4 as an enhancer of reprogramming. PLoS One 3:e1955.

Yang, J., Aguila, J.R., Alipio, Z. et al. (2011). Enhanced self-renewal of hematopoietic stem/progenitor cells mediated by the stem cell gene Sall4. J Hematol Oncol 4:38.

Young, R.A. (2011). Control of the embryonic stem cell state. Cell *144,* 940-954.

Zhang, P., Andrianakos, R., Yang, Y. et al. (2010). Kruppel-like factor 4 (Klf4) prevents embryonic stem (ES) cell differentiation by regulating Nanog gene expression. J Biol Chem *285*, 9180–9189.

Zhu, J., Zhang, Y., Joe, G.J., et al. (2005). NF-Ya activates multiple hematopoietic stem cell (HSC) regulatory genes and promotes HSC self-renewal. Proc Natl Acad Sci USA *102*, 11728–11733.

7. FIGURE LEGENDS

Figure 1. Transfection of recombinant NF-YA in mESCs.

(A) Western blot of NF-YA in nuclear extracts of mESCs treated with 20 nM of recombinant NF-YA. Cells were collected at indicated time points after protein transfection; 10 μg of nuclear extracts were loaded. Lamin B was used as loading control.

(B) Quantification of western blot analysis of NF-YAl transfection in mESCs. The data were obtained using Image software (BioRad), normalized with Lamin B and related to the amount of transfected NF-YAl present at 0.5 hours time-point.

Figure 2. Effects of recombinant NF-YA protein in mESCs in withdrawal of LIF.

(A) Recombinant NF-YAs, NF-YAl and the mutant NF-YAl 23-33, were transfected in mESCs. The addition of recombinant proteins was performed daily, every 24 hours for four days at the final concentration of 20 nM and mESCs were grown in presence or absence of LIF.

(B) Morphology of mESCs colonies treated with recombinant NF-YA proteins after four days of indicated treatments.

(C) Alkaline phosphatase assay of mESCs after four days of indicated treatments.

(D) qRT-PCR analysis of embryonic stem cells transcription factors from mESCs transfected with recombinant NF-YA in presence or absence of LIF. mRNA of ESCs transcription factors were analyzed four days after indicated daily treatments. mRNA transcripts were normalized with housekeeping RSP20 expression levels.

Figure 3. Generation and characterization of human induced pluripotent stem cells (hiPSCs).

(A) Schematic representation of hiPSCs generation protocol.

(B) Phase contrast images of (1) human dermal fibroblasts before transfection with OKSM vectors, (2) hiPSCs clone originated by human fibroblasts reprogramming on a feeder layer, (3) iPSCs clone seeded on a feeder free layer, (4) alkaline phosphatase assay of hiPSCs at p4.

(C) qRT-PCR analysis for pluripotent cell markers were performed on total RNA extraction from colonies obtained by reprogramming of human fibroblasts. The expression levels of stemness genes are relative to GAPDH. mRNA from HaCaT cells was used as negative control for the expression of stemness genes.

Figure 4. NF-YA isoform in hiPSCs before and after differentiation.

(A) Phase contrast images of hiPSCs colony grown on a feeder free layer and embryo body (EB) originated by induction of hiPSCs differentiation.

(B) Western blot analysis of B panel compare hiPSCs with mouse embryonic stem cells (mESCs) for the expression of NF-YA subunits. NF-YB was used as internal loading control.

Western blot analysis of the indicated proteins before and after 7 days of hiPSCs differentiation to EBs.

For western blot analysis 20 μg of total protein extracts were loaded on the gel.

Figure 5. Effects of NF-Y subunits depletion on the expression of stemness genes.

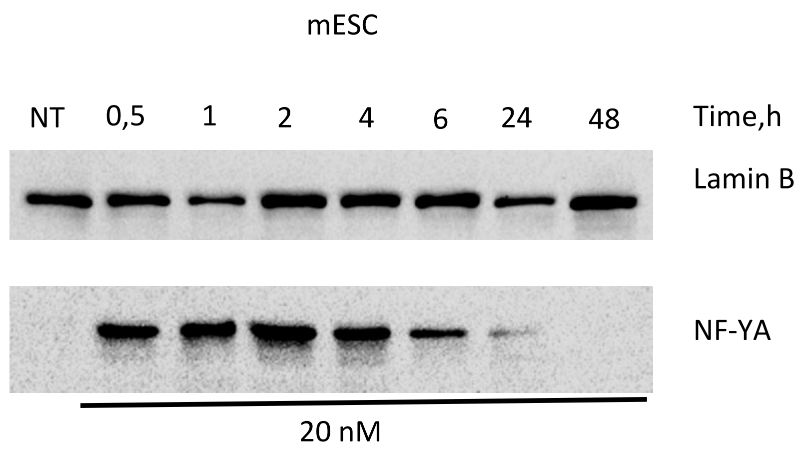
(A) and (C) Western blot analysis of NF-YA and NF-YB subunits after 72 hours of shRNA lentiviral infection of hiPSCs.

(B) and (D) mRNA levels of NF-YA and NF-YB subunits, stemness genes and housekeeping NF-Y target genes, were analyzed by qRT-PCR after inactivation of NF-YA (B panel) and NF-YB (D panel) depletion by shRNA lentiviral infection of hiPSCs. mRNA transcripts were normalized with housekeeping RSP20 expression levels. Error bars represent SD of qPCR replicates (six analyses). Experiments were performed twice both for NF-YA and NF-YB inactivation.

Supplementary 1. Primers used in this study are shown.

Figure 1: Transfection of recombinant NF-YAl in mESCs

A



B

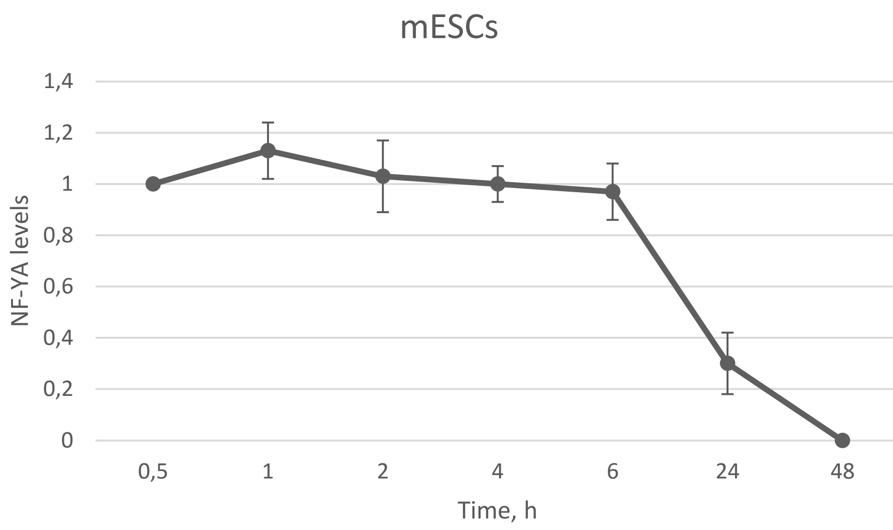


Figure 2: Effects of recombinant NF-YA in mESCs in withdrawal of LIF.

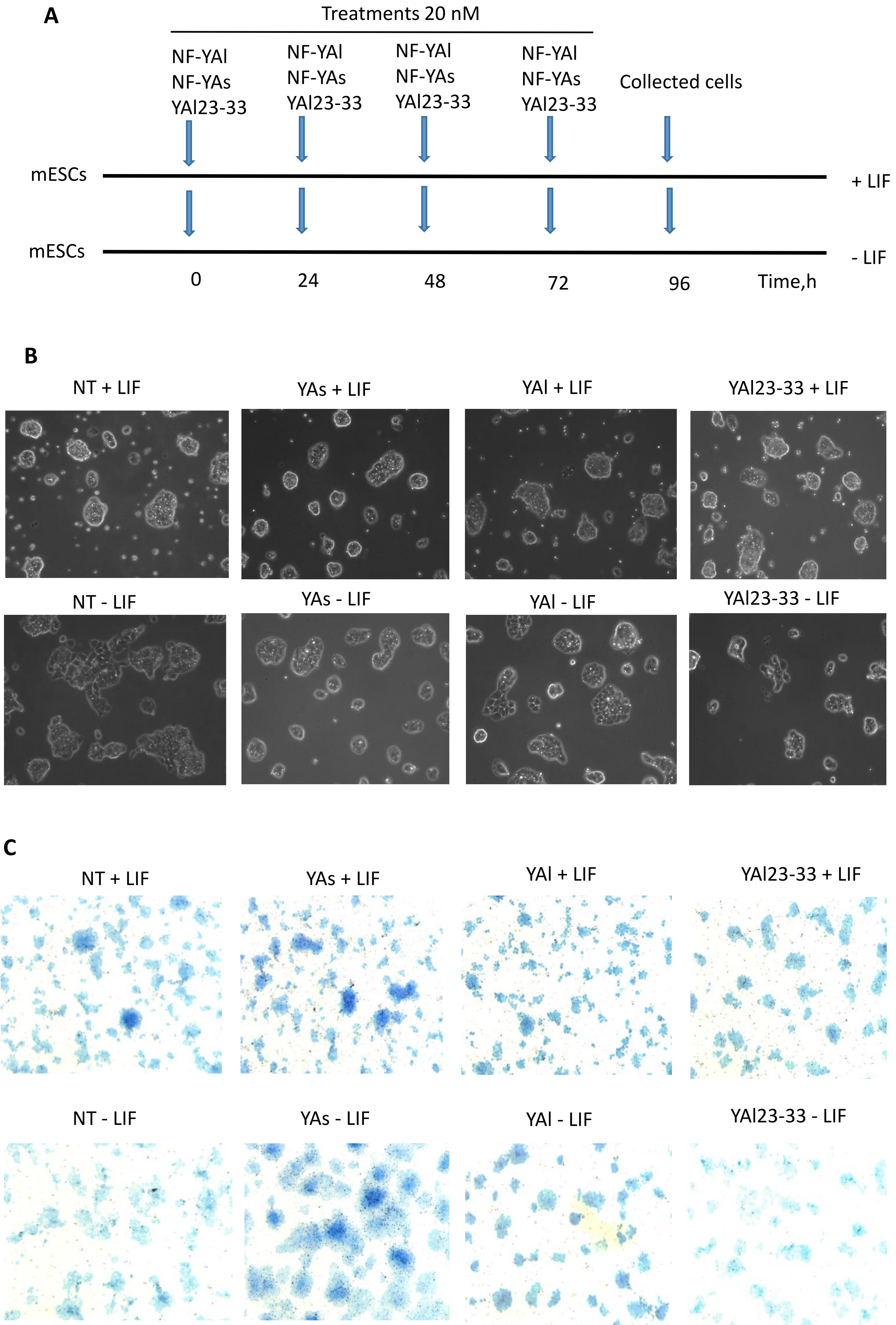


Figure 2: Effects of recombinant NF-YA in mESCs in withdrawal of LIF.

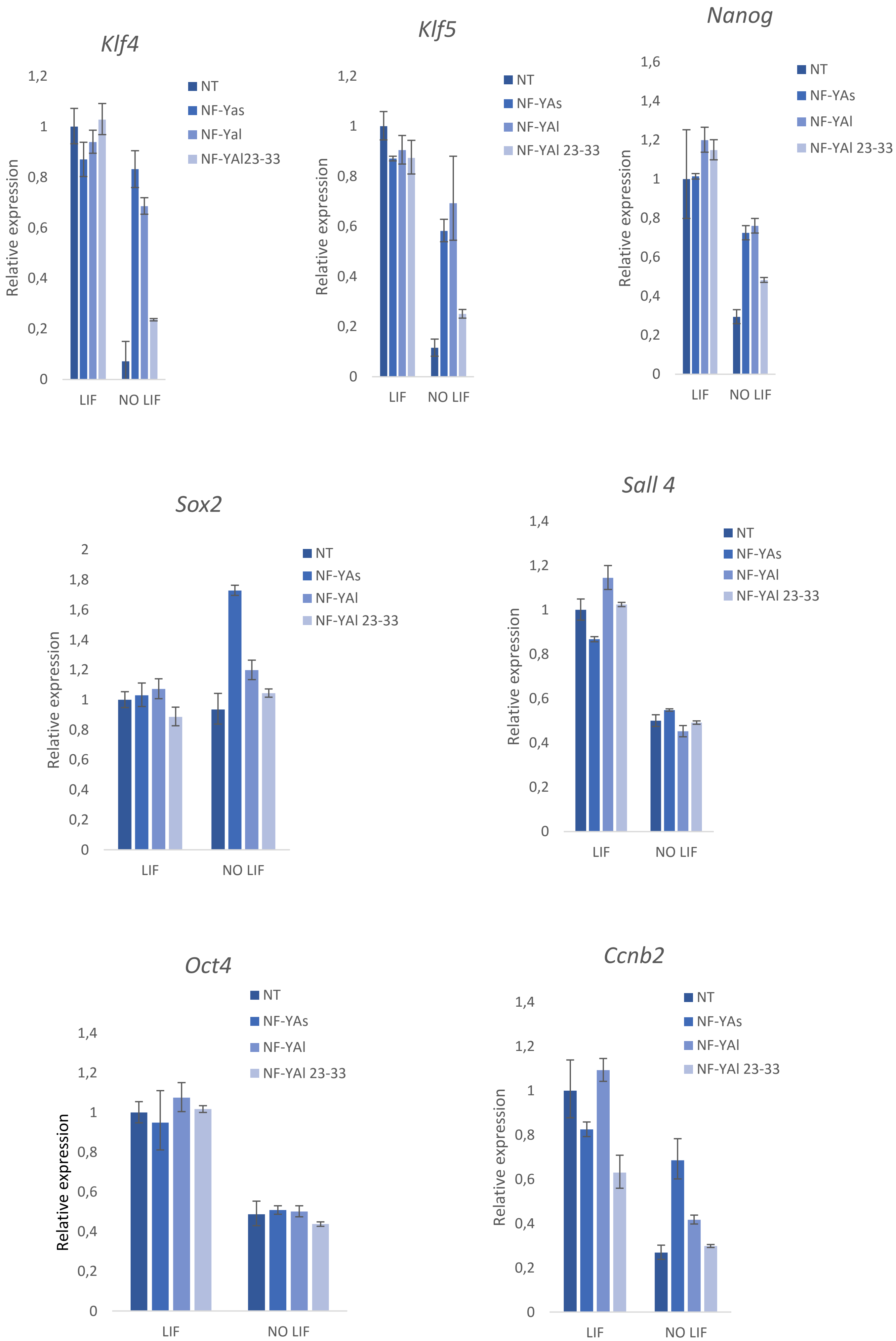
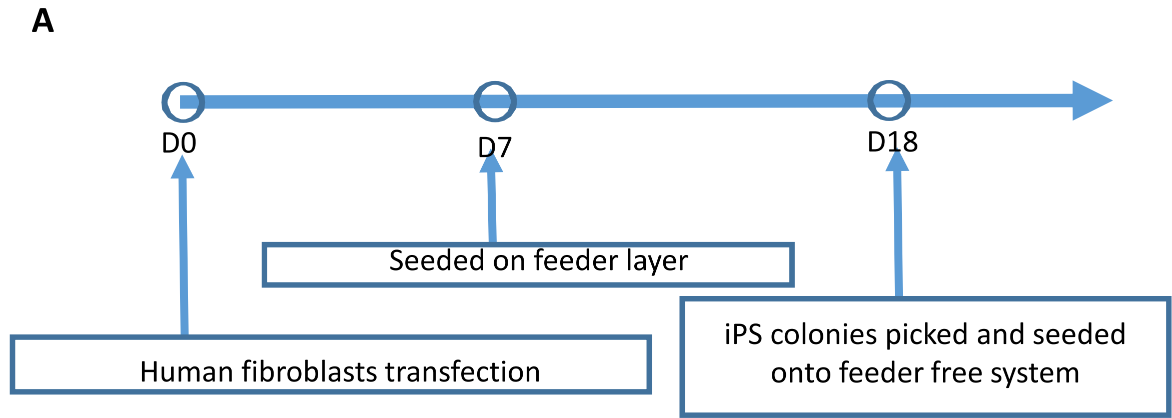
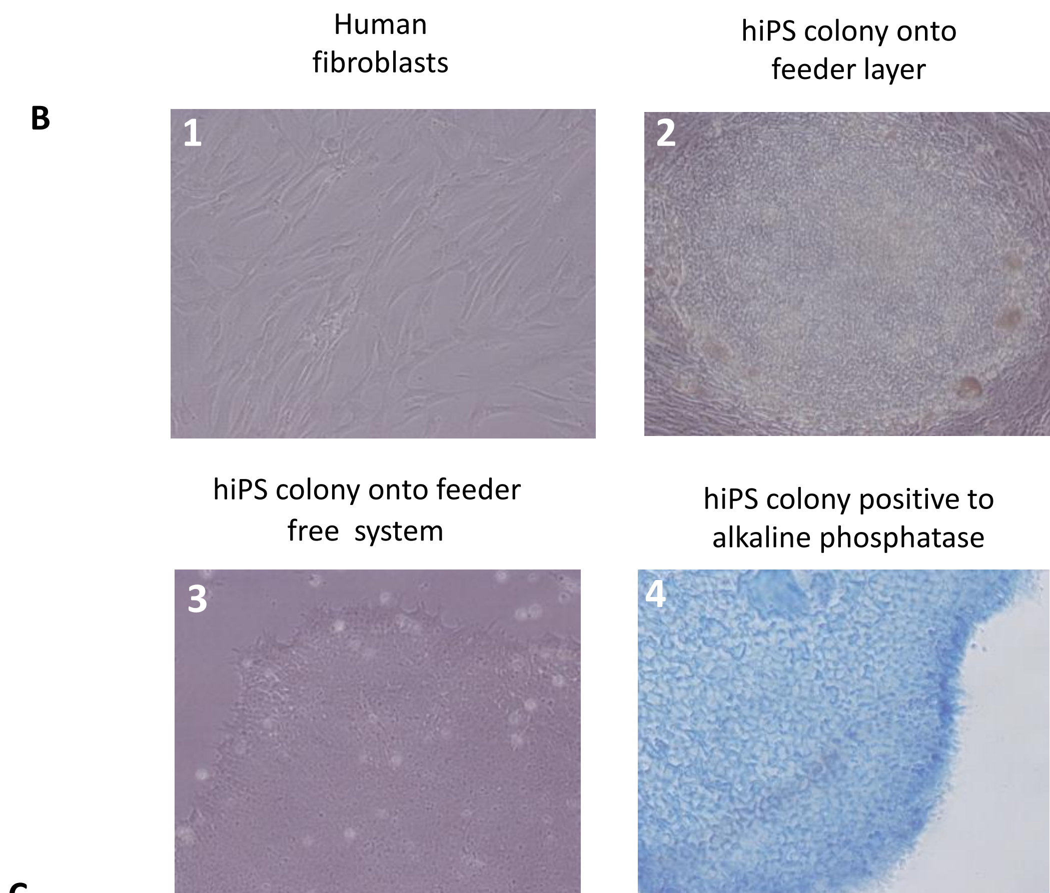


Figure 3: Generation and characterization of human induced pluripotent stem cells (hiPSCs).

A



B



C

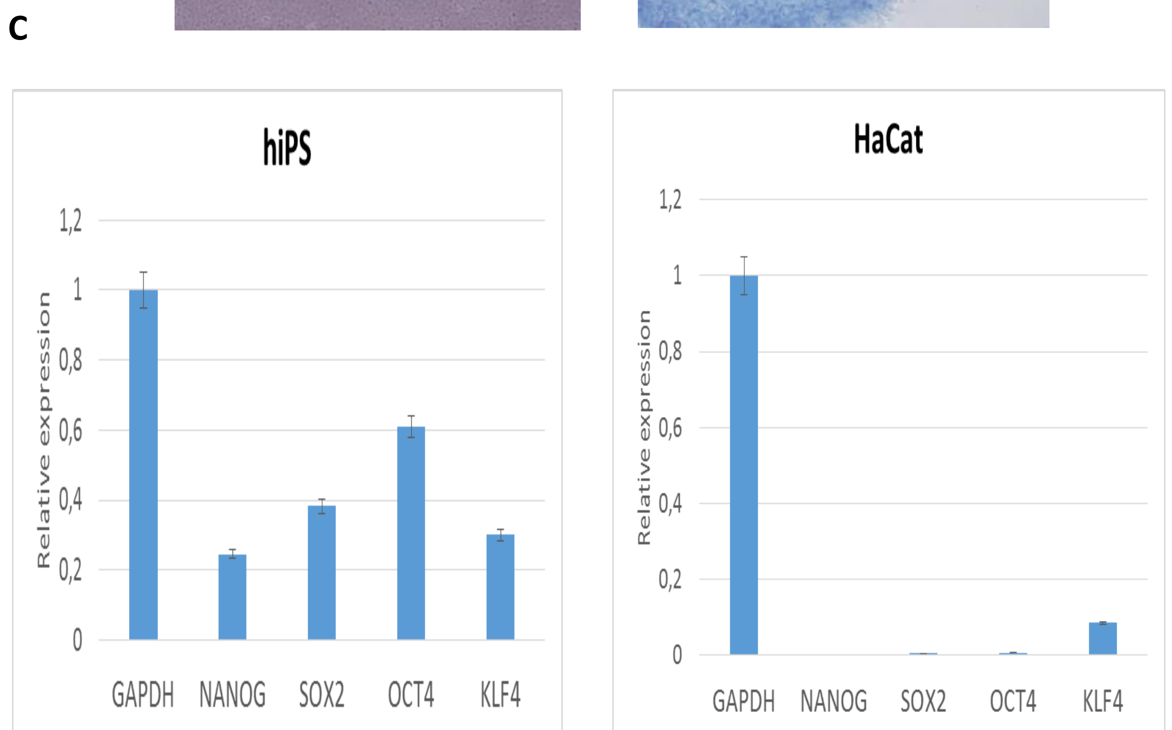
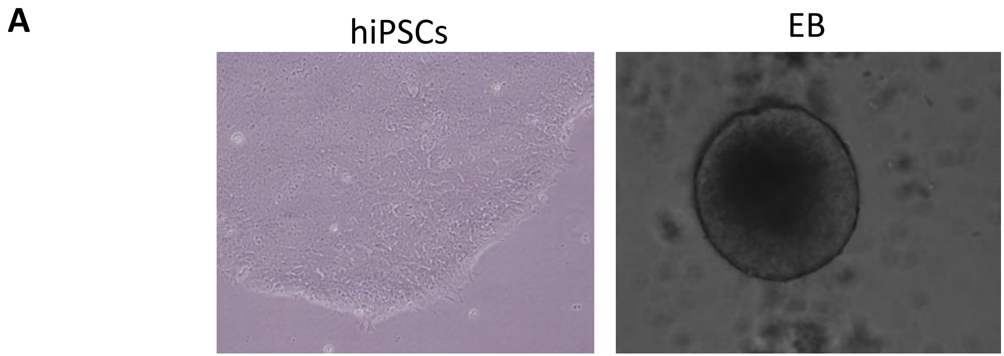
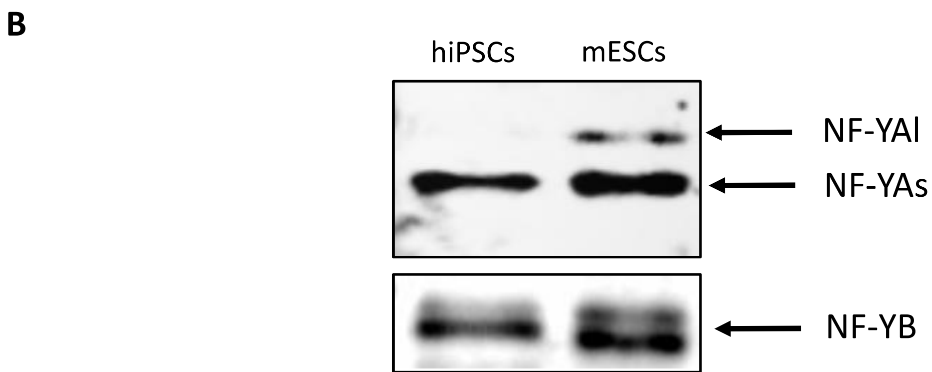


Figure 4: Generation of embryo bodies from hiPSCs and analysis of NF-YA isoforms before and after differentiation.

A



B



C

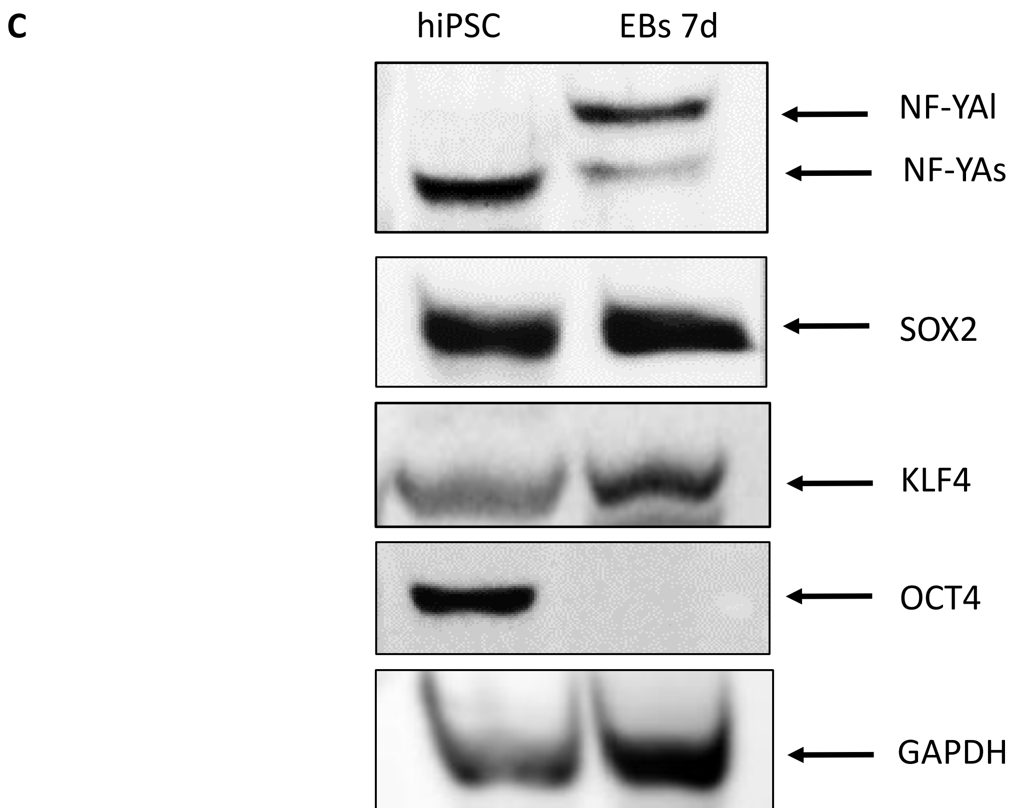
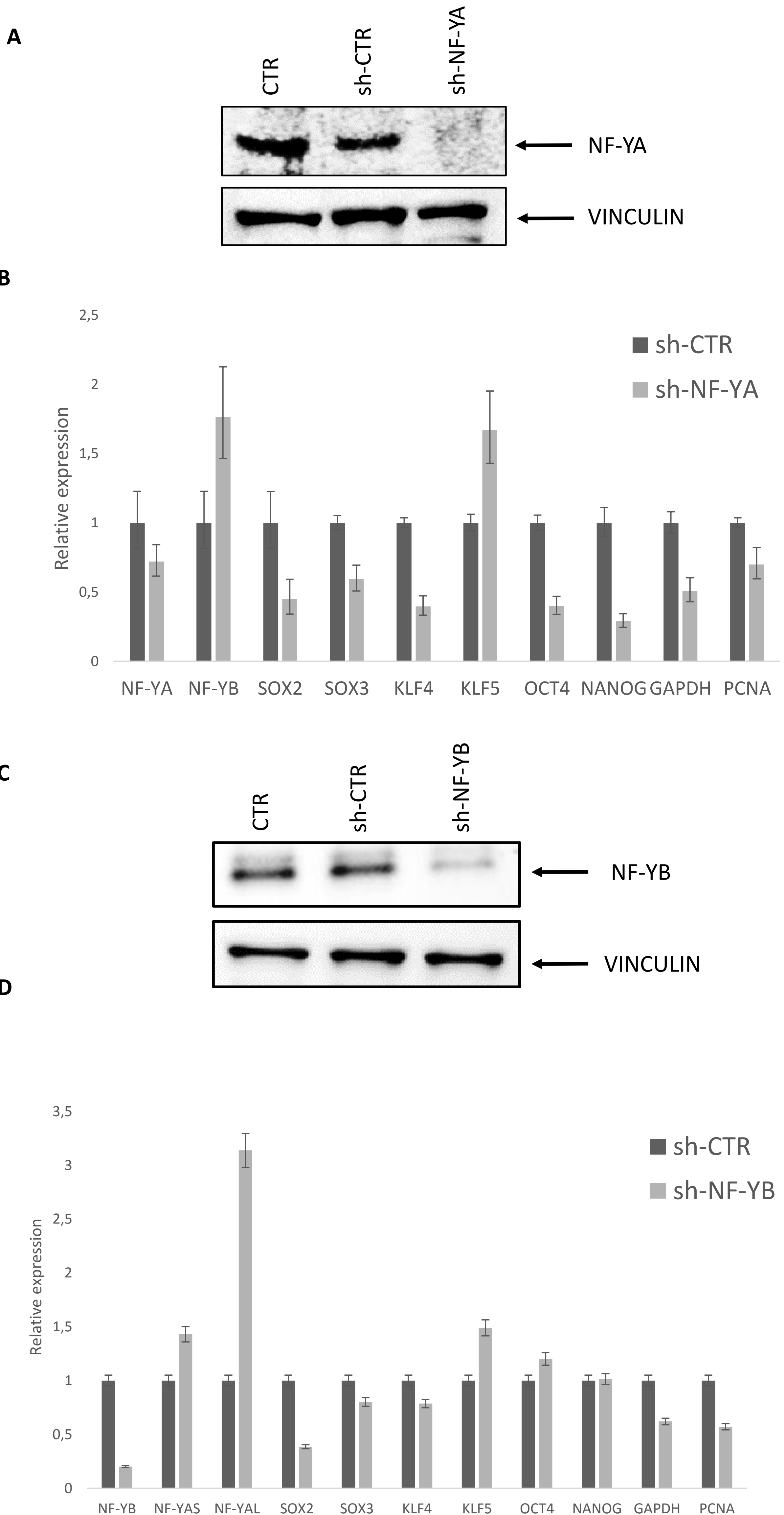


Figure 5: Effects of NF-Y subunits depletion on the expression of human stemness genes.



Supplementary 1: primers used in this study

Klf4	(F) 5'-AGAGAGTTCCTCACGCCAAC-3' (R) 5'-GCGAGTCTGACATGGCTGT-3'
Klf5	(F) 5'-CAGGTGCACTTGTAGGGCTT-3' (R) 5'-AGCGACGTATCCACTTCTGC-3'
Nanog	(F) 5'-GCGGACTGTGTTCTCTCAGG-3' (R) 5'-CCACCGCTTGCACTTCATCC-3'
Sox2	(F) 5'-GACCAGCTCGCAGACCTACA-3' (R) 5'-CTGGAGTGGGAGGAAGAGGT-3'
Sall4	(F) 5'-GGCGCTCTCTTTCCCTCAG-3' (R) 5'-GCGAACACACACGGGAGAG-3'
Oct4	(F) 5'-CAGCCAGACCACCSTCTGC-3' (R) 5'-TGGTCTCCAGACTCCACCTC-3'
Ccnb2	(F) 5'-GGAAGCCAGGAGCAGAGC-3' (R) 5'-CCTGGTGGACTGGCTGGT-3'
NF-YA	(F) 5'-TCAATTCAGGAGGGATGGTC-3' (R) 5'-GCCGAGACTCATGCAGGTAT-3'
NF-YB	(F) 5'-AGGTGCCATCAAGAGAAACG-3' (R) 5'-TGTTGTTGACCGTCTGTGGT-3'
SOX2	(F) 5'-ACCCTCCCCAGGTTTTCTCT-3' (R) 5'-CACCCACAGCAAATGACAGC-3'
SOX3	(F) 5'-GGGGAGGGCTGAAAGTTTTG-3' (R) 5'-ACACAGCGATTCCCAGCCTA-3'
KLF4	(F) 5'-ACCAGGCACTACCGTAAACACA-3' (R) 5'-GGTCCGACCTGGAAAATGCT-3'
KLF5	(F) 5'-CTTCCACAACAGGCCACTTACTT-3' (R) 5'-AGAAGCAATTGTAGCAGCATAGGA-3'
OCT4	(F) 5'-TCTCCAGGTTGCCTCTCACT-3' (R) 5'-GCTTTGAGGCTCTGCAGCTT-3'
NANOG	(F) 5'-CTGCGTCACACCATTGCT-3' (R) 5'-ATACCTCAGCCTCCAGCAGA-3'
PCNA	(F) 5'-GAGTCAAAGAGGCGGGGAGAC-3' (R) 5'-CTTGCGGGGAAGACTTTAGGG-3'
GAPDH	(F) 5'-GCCTCAAGATCATCAGCAATGC-3' (R) 5'-CCACGATACCAAAGTTGTCATGG-3'