

## IL PUNTO DI VISTA DI UN REGOLATORE EPIGENETICO DELL'ESPRESSIONE GENICA, COMUNE A UOMO E PIANTE

Nota del s.c. ROBERTO MANTOVANI (\*)

(Adunanza dell'8 ottobre 2020)

SUNTO. – La regolazione dell'espressione genica è al centro di tutti i processi fondamentali degli esseri viventi. Essa è controllata da Fattori Trascrizionali che, interagendo con specifiche sequenze di DNA dei genomi, richiamano macchinari proteici che modificano l'organizzazione della cromatina. I nostri studi sono concentrati sul Fattore Trascrizionale trimerico NF-Y, un trimero centrale per attivare l'espressione genica in tutti gli eucarioti. Due subunità hanno struttura simile agli istoni, gli elementi base della cromatina, la terza -NF-YA- riconosce il CCAAT box. Illusterò esempi di come la regolazione fine dei complessi NF-Y sia importante per due sistemi di espressione genica completamente diversi: la trasformazione neoplastica di cellule umane, e lo sviluppo dei fiori nelle piante.

\*\*\*

ABSTRACT. – The regulation of gene expression is at the heart of all the fundamental processes of living organisms. It is controlled by Transcriptional Factors which, by binding to specific DNA sequences of genomes, recruit protein machineries that modify the organization of chromatin (epigenetics). Our studies are focused on the NF-Y trimeric transcription factor, a crucial activator of gene expression in all eukaryotes. Two subunits have a structure similar to histones, the basic elements of chromatin, the third -NF-YA- confers specificity to the CCAAT box. I will illustrate how regulation of NF-Y complexes is important for two completely different gene expression systems: the neoplastic transformation of human cells, and the development of flowers in plants.

---

(\*) Dipartimento di Bioscienze, Università degli Studi di Milano, Italy.  
E-mail: mantor@unimi.it

## INTRODUZIONE

Per genetica intendiamo un insieme di manifestazioni “esteriori”, “fenotipiche” derivate dalla funzione di uno o più geni: nello specifico, dalla precisa sequenza di DNA presente nella loro parte codificante, o nelle aree regolative nelle immediate vicinanze. I fenomeni epigenetici si definiscono come cambiamenti nell’espressione dei geni che avvengono in maniera indipendente dalla sequenza di DNA dei geni stessi, e quindi, dalla genetica. Mentre le cause genetiche delle manifestazioni fenotipiche di tutti i viventi sono iscritte nel DNA delle loro cellule, i fenomeni epigenetici, come indica il suffisso greco *Epi*, sopra, modificano la risposta e il funzionamento genico in funzione di situazioni di stress, cellulare o dell’intero organismo, della presenza di sostanze tossiche, dell’alimentazione, del comportamento, nostro o di altri viventi che ci circondano: in due parole, dell’ambiente esterno.

Attraverso quali meccanismi molecolari l’ambiente può modificare il funzionamento dei geni, l’insieme dei quali definiamo genoma? I due meccanismi ritenuti fondamentali sono la metilazione del DNA, che avviene in aree regolative ricche in dinucleotidi CG (Isole CpG), e le modificazioni post-traduzionali degli istoni, le proteine che compattano i genomi eucariotici. Negli ultimi 20 anni, decine di modificazioni di queste proteine strutturali della cromatina sono state scoperte e caratterizzate; assieme a loro, gli enzimi che scrivono e cancellano (“Scrittori” e “Cancellatori”) e i “Lettori”, cioè le proteine che riconoscono questo -presunto- “codice” epigenetico. Molti studiosi aggiungono ai meccanismi epigenetici anche l’azione di RNA non codificanti -ncRNAs- in quanto esistono solide evidenze che alcuni di questi, in particolare miRNA, siano importanti per la trasmissibilità epigenetica trans-generazionale (cioè da genitori a figli) dei caratteri. In generale, esistono almeno due livelli di controllo epigenetico: il primo agisce nella trasmissione dei caratteri alle generazioni successive (1); il secondo, presente in ogni cellula di organismi pluricellulari, ne garantisce la trasmissione delle caratteristiche, e gli eventuali progressivi cambiamenti. Ha a che fare, insomma, con l’“identità” cellulare. Le modifiche epigenetiche vengono trasmesse alle cellule figlie, cosicchè esse ricevano un’identità definita sulla base di geni espressi, in linea con la cellula che le ha generate.

Esiste una corrente di pensiero che segnala come metilazione del DNA e modificazioni istoniche non siano gli unici, o nemmeno i fonda-

mentali, meccanismi epigenetici (2). Ciò per diversi motivi: ad esempio, *Drosophila*, e molti insetti in generale, ha meccanismi di trasmissione epigenetica transgenerazionale e di identità cellulare, pur in assenza di metilazione del DNA; si fa anche notare che i vari macchinari che scrivono, leggono e cancellano il “codice” non sono in grado di selezionare le aree del genoma su cui agire, per “aprire” o “chiudere” la cromatina, e quindi guidare l’espressione genica. Questo compito spetta a un’altra classe di proteine: i Fattori Trascrizionali -TF- dotati di una caratteristica fondamentale: la selettività per specifiche sequenze di DNA. Solo dopo il riconoscimento da parte dei TF dei loro siti sul genoma, i macchinari “epigenetici” vengono reclutati in loco a svolgere le loro funzioni. Se, dunque, i TF stanno a monte dei macchinari epigenetici, perché non considerare loro stessi alla base dei fenomeni epigenetici? A questo proposito, è illustrativo l’esperimento pubblicato di S. Yamanaka sulla “riprogrammazione” di cellule differenziate di topo a cellule staminali totipotenti (mESCs): i geni decisivi per questo drammatico cambio di identità cellulare -stesso genoma ma radicale cambio di epigenoma- codificano tutti per TF (3).

La mia carriera scientifica è in gran parte volta allo studio di un TF chiamato NF-Y, composto da tre diverse subunità: due sono strutturalmente istoni e la terza conferisce specificità per una specifica sequenza di DNA, il CCAAT box.

## IL CCAAT BOX

La sequenza CCAAT è stata, insieme a TATA, una delle prime ad essere identificata -oltre 40 anni fa- in prossimità dei siti di inizio della trascrizione di geni eucariotici, cioè nel promotore (ovalbumina, b-globina, istone H2A). Immediatamente dopo, ne è stata verificata, attraverso esperimenti genetici di mutagenesi *in vitro* ed *in vivo*, la fondamentale importanza per la funzione dei promotori. Nel corso degli anni 90, la caratterizzazione di >100 promotori di mammifero ha messo in evidenza la presenza del box CCAAT a una distanza compresa tra -60 e -100 dal sito di inizio della trascrizione, cosa che mi ha permesso di derivare una “matrice” di DNA, una sequenza consenso di nucleotidi presenti (4); successivamente, questa matrice è stata inserita nelle banche dati dei siti riconosciuti dai TF (JASPAR e TRANSFAC), ancora oggi usate nelle ricerche delle aree funzionali dei genomi di mammifero.

A partire dal completamento del sequenziamento del genoma umano, nel 2001, studi di caratterizzazione bioinformatica dei promotori hanno poi chiarito che CCAAT è presente in circa il 25% dei promotori di geni trascritti dalla RNA Pol II (mRNA, lncRNA, miRNA).

## NF-Y

Poco dopo la caratterizzazione genetica del CCAAT box, è stato individuato il TF che ne regola la funzione. La proteina che si lega con alta affinità e specificità è NF-Y (Nuclear Factor Y), formato da 3 subunità, NF-YA, NF-YB e NF-YC. Rispetto ad altri TF, NF-Y è peculiare per diversi motivi. Innanzitutto, tutte e tre le subunità sono necessarie per il legame al DNA. In secondo luogo, due sono istoni, della famiglia di H2A/H2B, come determinato da struttura cristallografica nel 2003 dai collaboratori C. Romier e D. Moras (5). In terzo luogo, NF-YA è un gene che non ha nessun omologo nei vertebrati, caso unico tra i TF, non essendo andato incontro, come le altre famiglie di TF, a successive duplicazioni ed espansioni geniche. Quarto, la struttura cristallografica del trimero legato al suo sito di DNA, risolta nel 2013 dai collaboratori M. Nardini e M. Bolognesi, ha messo in luce un legame nel solco minore del DNA, fatto raro per TF con specificità di sequenza (6). Quinto, anche grazie ad estesi contatti laterali non sequenza-specifici del dimero istonico, il DNA è piegato di circa 80°, cosa che conferisce al complesso la capacità di organizzare tridimensionalmente il promotore.

Esperimenti compiuti da vari laboratori, compreso il nostro, hanno portato alla conclusione che NF-Y sia un TF “pioniere”: questa sottoclasse di TF è in grado di raggiungere territori cromatinici chiusi, non interessati da modificazioni istoniche positive, e per questo motivo inaccessibili ad altri TF. In parallelo, NF-Y è stato inserito nel progetto ENCODE, che da 10 anni studia gli elementi funzionali del genoma umano, compresi i marcatori “epigenetici” (metilazione del DNA, modificazioni istoniche) e il legame di centinaia di TF (attualmente >350) in un certo numero di linee cellulari tumorali (7). Grazie a questa enorme massa di dati, abbiamo progressivamente identificato circa un centinaio di TF che si legano frequentemente vicino ai siti genomici riconosciuti da NF-Y. Lo scopo è ora quello di ricostruire la logica informazionale dell’associazione di questi gruppi di TF ai promotori e alle zone regolative del genoma delle diverse cellule; queste interazioni

vengono poi sezionate biochimicamente, determinandone, laddove possibile, la struttura tridimensionale.

## NF-Y NEI TUMORI

Un'area in cui il mio laboratorio si è ultimamente impegnato è quella dello studio del ruolo di NF-Y nei tumori. Analisi genetiche dei tumori umani hanno fin qui escluso che i geni delle tre subunità di NF-Y siano mutati in modo rilevante nelle cellule tumorali. Una possibile spiegazione è che NF-Y sia necessario per l'espressione di numerosi geni "cancro", cioè geni spesso sovraespressi nei tumori; questi geni codificano solitamente per proteine che inducono proliferazione e cambiamenti nel metabolismo cellulare, eventi alla base della trasformazione neoplastica. D'altra parte, analizzando le regioni regolative di tali geni è emerso che CCAAT è spesso in cima alla lista dei siti dei TF presenti. Abbiamo dunque deciso di interrogare, tra le altre, la banca dati del consorzio TCGA, che contiene dati clinici, molecolari, genetici ed epigenetici di >10.000 tumori umani: delle tre subunità, NF-YA risulta significativamente sovraespresso in quasi tutti i tipi di tumori epiteliali. Abbiamo quindi deciso di caratterizzarne alcuni più precisamente. Mi limiterò qui ad esporre i principali risultati che abbiamo ottenuto nei tumori della mammella. Il risultato principale dell'analisi bioinformatica è che la maggiore espressione di NF-YA è accompagnata da un cambiamento nelle isoforme di "splicing" presenti. Lo splicing alternativo è un fondamentale passaggio regolativo grazie al quale da un gene vengono prodotti più mRNA, e quindi proteine. NF-YA è presente con due principali isoforme, NF-YA1 e NF-YAs, frutto dell'esclusione dell'esone 3. I tumori della mammella sono comunemente classificati in 4 tipi: in tre di questi (Luminal A, Luminal B e HER2), l'isoforma preponderante è NF-YAs. Nel quarto tipo, denominato Basal-like, è espresso anche NF-YA1: esso anzi diventa prevalente in un sottotipo di questa classe, chiamato Claudin-low. Sono questi tumori caratterizzati da una trasformazione epitelio-mesenchimale, capacità di generare cellule in grado di metastatizzare e di sviluppare farmaco-resistenza. Infatti, abbiamo stabilito che alti livelli di NF-YA1 sono associati non solo a un fenotipo mesenchimale, ma anche a una prognosi peggiore. Attraverso esperimenti di editing genomico, abbiamo eliminato le sequenze dell'esone 3 dal genoma di cellule di tumore mammario

Claudin-low, ottenendo conferma della minore capacità invasiva dei cloni geneticamente modificati, che esprimono solo NF-YAs e non più NF-YAL. Lo stesso schema di splicing alternativo sembra essere comune anche ai carcinomi polmonari e del cavo orale. Questi studi ci spronano a proseguire gli studi genomici sulle isoforme di NF-YA in altri tipi di tumori e aprire un filone di studi biochimici per l'individuazione dei meccanismi molecolari alla base del comportamento differenziale delle due isoforme, dovuto ai 28 aminoacidi dell'esone 3, inseriti nella isoforma lunga NF-YAL.

### NF-Y NELLE PIANTE

Circa 20 anni fa, in collaborazione con C. Tonelli, abbiamo provveduto a identificare e catalogare tutti i geni di NF-Y di *Arabidopsis thaliana* (8), una pianta modello il cui genoma era stato appena sequenziato. Da allora, la nomenclatura che abbiamo proposto è stata seguita dalla comunità dei genetisti e biologi vegetali, che hanno studiato i genomi di >20 specie di piante nell'ultima decade. In tutte, si conferma una espansione impressionante del numero -mediamente 8/15- dei geni delle tre subunità. Ciò pone una serie di problemi biochimici ovvii: possono ciascuna delle subunità istoniche formare degli eterodimeri? Possono tutti gli eterodimeri associarsi a tutti gli NF-YA presenti? Esiste una sequenza-selettività dei diversi trimeri formati per diversi tipi di CCAAT box? Una serie di studi sulle subunità di *Arabidopsis*, compresa la risoluzione di strutture tridimensionali di dimeri e di un trimero in complesso con il DNA, ci hanno permesso di rispondere affermativamente alle prime due domande e negativamente alla terza. Un nostro contributo rilevante all'avanzamento delle conoscenze in questo campo è stata la scoperta che una seconda famiglia di TF di pianta, con somiglianza strutturale a NF-YA, si comporta in modo simile, cioè formando un trimero con le due subunità istoniche NF-YB/NF-YC (9): si tratta di una vasta famiglia -39 geni in *Arabidopsis*- denominata CCT (Constans, CO-like, TOC1). Questi trimeri legano però una sequenza leggermente, ma significativamente, differente (TGTGG). I due membri più studiati, CONSTANS e TOC1, sono geni chiave per la regolazione dei tempi dello sviluppo dei fiori e dei ritmi circadiani, rispettivamente. Questi nostri studi razionalizzano molti dati precedenti, alcuni apparentemente contraddittori, e propongono un modello per cui esi-

stono centinaia di complessi NF-Y (NF-YA/NF-YB/NF-YC) e, in parallelo, migliaia di complessi NF-CCT (CCT/NF-YB/NF-YC). Recentemente, alcuni dei complessi di Arabidopsis e riso sono stati caratterizzati mediante cristallografia, e il nostro modello confermato.

## CONCLUSIONI

A distanza di oltre 40 anni dalla identificazione del CCAAT box, e di oltre 30 anni da quella di NF-Y, questo sistema continua ad essere un modello che fornisce preziose indicazioni sui meccanismi di regolazione trascrizionale, in sistemi tanto diversi quali le cellule tumorali e lo sviluppo dei fiori. Al cuore di questo complesso ci sono due entità con ruoli potenzialmente differenti: da un lato le due subunità istoniche NF-YB/NF-YC che potrebbero “dialogare” con gli altri istoni, inserendosi in regioni genomiche chiuse, giocando quindi un ruolo di regolatore “epigenetico” classico; dall’altro, la subunità regolativa NF-YA, e CCT in piante, che indirizza il complesso verso specifiche sequenze di DNA nelle regioni regolative, giocando il ruolo tipico dei TF. Queste due caratteristiche sono dunque riunite in un’unica entità: come questo ruolo di “pioniere” venga svolto è ancora oscuro: per questo, a mio avviso, ciò garantisce che NF-Y rimanga al centro della nostra attenzione in futuro.

## BIBLIOGRAFIA

1. Boškovi A, Rando OJ. Transgenerational Epigenetic Inheritance. *Annu Rev Genet.* 52, 21-41 (2018).
2. Ptashne M. Epigenetics: core misconception. *Proc Natl Acad Sci USA.* 110, 7101-3 (2013).
3. Takahashi K, Yamanaka S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* 25, 126:663-76 (2006).
4. Mantovani, R. A survey of 178 NF-Y binding CCAAT boxes. *Nucleic Acids Res.* 26, 1135-1143 (1998).
5. Romier C., Cocchiarella F, Mantovani R., and Moras D. The crystal structure of the NF-YB/NF-YC heterodimer gives insight into transcription regulation and DNA binding and bending by transcription factor NF-Y. *J. Biol. Chem.* 278, 1336-1345 (2003).
6. Nardini M., et al. NF-Y is a sequence-specific transcription factor displaying

- histone-like DNA binding and H2B-like ubiquitination. *Cell* 152, 132-143 (2013).
7. Ronzio M, Bernardini A, Pavesi G, Mantovani R., Dolfini D. On the NF-Y regulome as in ENCODE (2019). *PLoS Computational Biology* 16, e1008488 (2020).
  8. Gusmaroli G., Tonelli C. and Mantovani R. Regulation of novel members of the *A. thaliana* CCAAT-binding NF-Y subunits. *Gene* 283, 41-48 (2002).
  9. Gnesutta N., Mantovani R., Fornara F. Plant flowering: imposing DNA specificity to Histone-fold subunits. *Trends in Plant Science*, 23, 293-301 (2018).