

UNIVERSITA' DEGLI STUDI DI MILANO  
FACOLTA' DI SCIENZE MATEMATICHE, FISICHE E NATURALI

LAUREA TRIENNALE IN BIOTECNOLOGIE INDUSTRIALI E AMBIENTALI

ELABORATO FINALE

***Mappe ad alta risoluzione di metilazioni e acetilazioni istoniche  
nel genoma umano***

**Relatore: Professore Giulio Pavese**

**Marco Notaro  
765324**

**anno accademico 2011- 2012**

## **RINGRAZIAMENTI**

*Desidero ringraziare il professore Giulio Pavesi per la disponibilità, la cortesia e per tutti i consigli forniti durante la stesura.*

*Un sentito ringraziamento ai miei genitori che con il loro sostegno morale ed economico, mi hanno permesso di raggiungere questo primo e importante traguardo nel mondo della ricerca.*

*Un ultimo, ma non meno importante ringraziamento ad Elisa per avermi sempre incoraggiato e sostenuto nei momenti difficili.*

# INDICE

RINGRAZIAMENTI .....	1
RIASSUNTO .....	3
INTRODUZIONE.....	4
IMMUNOPRECIPITAZIONE DELLA CROMATINA .....	7
<b>CONFRONTO TRA CHIP-ON-CHIP, CHIP-SAGE E CHIP-SEQ .....</b>	<b>12</b>
Risoluzione.....	12
Quantificazione.....	12
Costi.....	12
METILAZIONE.....	13
<b>CORRELAZIONI DELLE MODIFICHE ISTONICHE SULLE REGIONI PROMOTRICI CON L'ESPRESSIONE GENICA .....</b>	<b>13</b>
Pol II Binding .....	14
Metilazioni di H3K4.....	15
Metilazioni di H3K27.....	15
Metilazioni di H3K36.....	15
Metilazioni di H3K79.....	15
Metilazioni di H3K9.....	16
Metilazioni di H4K20.....	16
Metilazioni di H2BK5 .....	16
Variante istonica H2A.Z .....	16
<b>CONFRONTO TRA ATTIVITÀ GENICA E METILAZIONE ISTONICA IN REGIONI TRASCritte .....</b>	<b>16</b>
Modificazioni associate con il silenziamento genico .....	18
Modificazioni associate con l'attivazione genica.....	18
Pattern di metilazione istonica di geni attivi e silenti.....	18
Metilazione istonica su Insulator.....	19
Metilazione istonica su Enhancers .....	21
<b>BREAKPOINT LOCALIZZATI IN DOMINI DI CROMATINA ATTIVA .....</b>	<b>22</b>
<b>PREDIZIONE DI NUOVE UNITÀ TRASCRIZIONALI USANDO I PATTERNS DI METILAZIONE ISTONICA .....</b>	<b>23</b>
ACETILAZIONE.....	25
CONCLUSIONI .....	27
BIBLIOGRAFIA .....	28

## RIASSUNTO

Le proteine istoniche sono soggette a numerose modificazioni covalenti localizzate principalmente a livello della coda N-terminale, che influenzano l'attivazione o la repressione della trascrizione genica. Ad oggi sono stati proposti svariati modelli per spiegare e comprendere il ruolo di queste modifiche, quali ad esempio il codice istonico, il signalling-pathway e il modello di neutralizzazione di carica. Tuttavia, fino a pochi anni fa la mancanza di dati sulla distribuzione delle modifiche nel genoma rendeva impossibile la piena comprensione del grado di complessità dei patterns combinatoriali istonici.

Negli ultimi anni si è d'altra parte assistito ad un marcato incremento della quantità di informazioni sulla funzione di queste modificazioni istoniche, grazie allo sviluppo di approcci genome-wide che combinano tecniche di immunoprecipitazione della cromatina (ChIP) con analisi di DNA microarray (ChIP-on-chip) e con tecniche di high-throughput sequencing (ChIP-Seq). E' stato così possibile generare mappe ad alta risoluzione della distribuzione genomica di singole modifiche istoniche, come la metilazione dei residui di lisina e arginina e l'acetilazione dei residui di lisina, oltre a varianti istoniche come H2A.Z, l'RNA Pol II, e alle regioni di legame (sequenze insulator) della proteina CTCF. Ad esempio, utilizzando tecniche di next-generation sequencing (negli studi presi in esame è stato utilizzato il sequenziatore Solexa 1G Genome Analyzer) è stato possibile identificare tipici patterns di metilazione e acetilazione istonica a livello dei promotori, degli enhancer e delle regioni trascritte, e le loro correlazioni con la trascrizione. Si è così osservato, ad esempio, che le monometilazioni di H3K27, H3K9, H4K20, H3K79 e H2BK5 sono collegate con l'attivazione genica; le trimetilazioni di H3K27, H3K9 e H3K79 sono correlate con la repressione genica; la variante istonica H2A.Z è associata a elementi regolatori funzionali; i siti di legame di CTCF definiscono chiaramente i confini tra domini cromatinici attivi (marcati da H3K27me1) e repressivi (marcati da H3K27me3); tutti e tre gli stati di metilazione di H3K4 sono associati con i punti di rottura cromosomici in cellule umane T cancerogene. La tendenza di queste modificazioni a coesistere suggerisce che la maggior parte di esse siano correlate le une con le altre a livello di singolo nucleosoma. Sebbene i geni associati con le stesse modificazioni tendano ad avere simili livelli di espressione, le modificazioni istoniche di per sé non determinano il livello esatto di trascrizione, ma potrebbero funzionare cooperativamente per preparare la cromatina all'attivazione trascrizionale mediata dai fattori di trascrizione.

*Articolo 1:* <http://www.ncbi.nlm.nih.gov/pubmed/17512414?dopt=Abstract>

*Articolo 2:* <http://www.ncbi.nlm.nih.gov/pubmed/18552846?dopt=Abstract>

## INTRODUZIONE

All'interno di una cellula eucariotica il DNA interagisce molto strettamente con proteine istoniche, che condensano e ordinano il DNA in unità strutturali di impacchettamento superiori, dette nucleosomi. Gli istoni sono proteine con masse molecolari comprese tra 11 e 21 kDa e aventi caratteristiche basiche, in modo tale da facilitare l'interazione con lo scheletro zucchero-fosfato acido del DNA. Come mostrato in Figura 1, ogni nucleosoma è costituito da circa 200 paia di basi, di cui 146 sono legate, mediante 1.8 giri di superavvolgimento sinistrorso a solenoide, ad un core istonico (struttura proteica ottamerica formata dagli istoni H2A, H2B, H3, H4, ciascuno presente in duplice copia), mentre le rimanenti 54 paia di basi servono da collegamento tra i diversi nuclei istonici. Un quinto istone, l'istone H1, è infine legato al DNA linker.

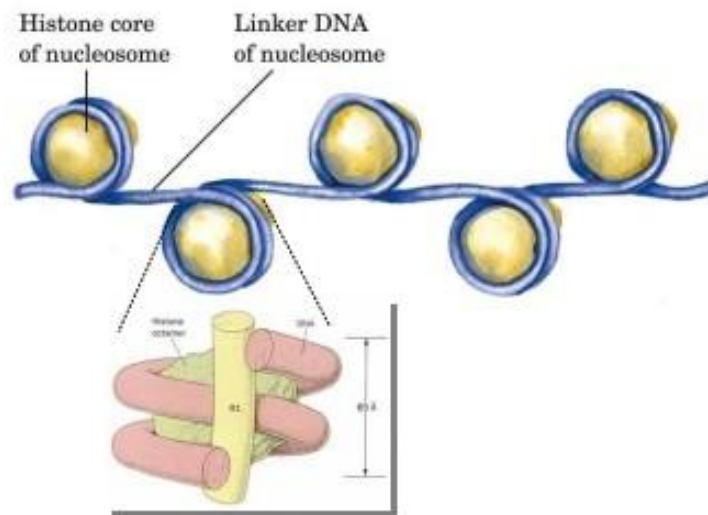


Figura 1A: illustrazione schematica di alcuni nucleosomi spazati ad intervalli regolari costituiti da complessi di istoni legati al DNA [1]. Dettaglio di un singolo nucleosoma [2].

Ciascun istone, in particolare a livello della coda N-terminale (strettamente conservata nella sua sequenza), può essere soggetto a numerose modificazioni post traduzionali quali metilazione, acetilazione, fosforilazione, glicosilazione, sumoilazione, ed ubiquitazione. Queste modificazioni cambiano la forma e la carica elettrica netta delle molecole di istone e di conseguenza la struttura e le proprietà funzionali della cromatina, giocando un ruolo rilevante nella regolazione della trascrizione genica. Infatti, le code istoniche di un nucleosoma possono prendere contatto con il core di un nucleosoma adiacente mediandone l'interazione, stabilizzando la struttura della cromatina nella fibra 30nm e contribuendo ad una maggiore compattazione del DNA (Figura 1B), che richiede la presenza di una molecola di istone H1 per nucleosoma. La struttura della fibra 30nm sembra anche variare a seconda dall'attività trascrizionale della particolare regione di DNA: zone

contenenti geni in cui si sta verificando un processo di trascrizione si presentano apparentemente in uno stato meno ordinato che contiene una quantità trascurabile dell'istone H1.

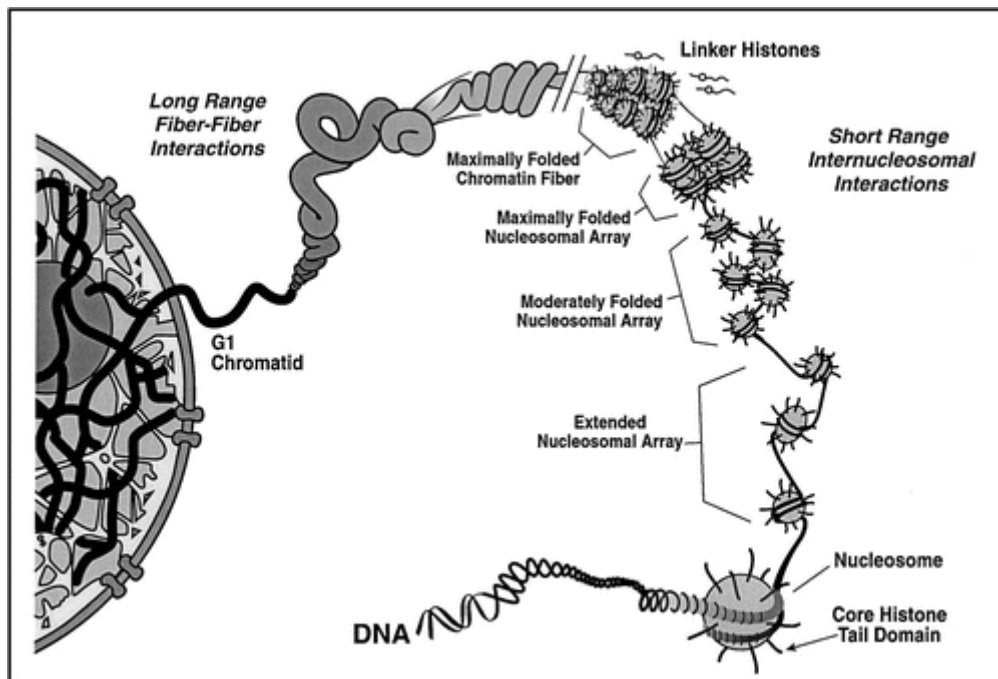


Figura 1B: rappresentazione schematica della condensazione in fibre di cromatina [3].

Le modificazioni istoniche, determinando nella sua globalità la struttura cromatinica, sono in grado di influenzare l'espressione genica e la funzione genomica, orchestrando tutti i processi biologici che coinvolgono il DNA. Tra le varie modificazioni istoniche conosciute, la metilazione e l'acetilazione sembrano essere quelle più importanti. In letteratura l'acetilazione delle code istoniche è solitamente associata con l'attivazione genica. Infatti il gruppo acetile, dotato di una carica negativa, sembra in grado di neutralizzare la carica positiva della lisina, che, a sua volta, causa una destabilizzazione del nucleosoma interferendo con l'interazione istone-DNA avvolto. La metilazione delle code istoniche, invece, è associata sia alla repressione che all'attivazione genica, in base alla posizione e allo stato in cui essa si trova. In genere la metilazione dell'arginina attiva la trascrizione, mentre la metilazione della lisina la inibisce. La teoria del codice istonico suggerisce che diverse modifiche istoniche possano agire in combinazione, determinando stati cromatinici distinti.

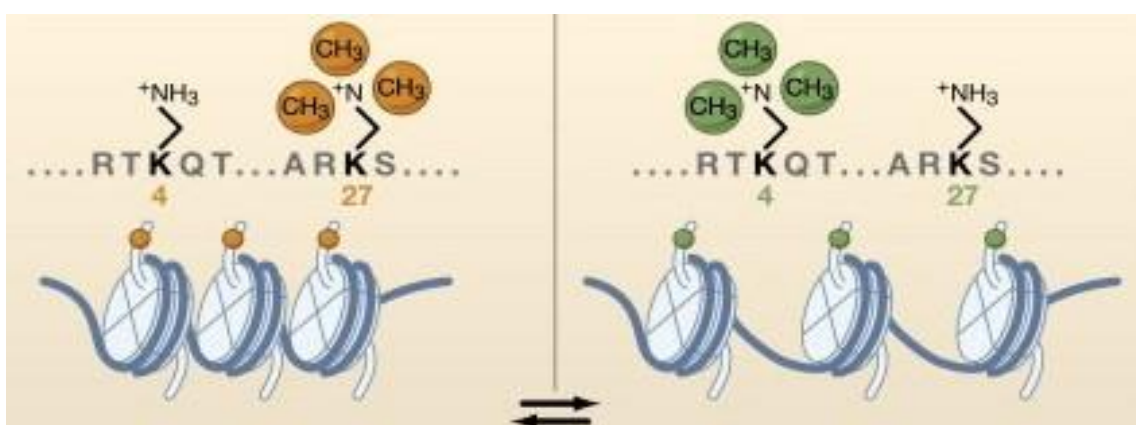
Il quadro generale emerso dagli studi su lievito compiuti in questi ultimi anni ha permesso di osservare che a livello delle regioni promotrici di geni attivi si ha una maggiore acetilazione istonica e una minore occupazione nucleosomale. Per quanto riguarda i livelli di metilazione, invece, si sono osservati alti livelli di tutti e tre gli stati di metilazione di H3K4 (lisina 4 dell'istone

H3) nelle regioni circostanti il TSS ed elevati picchi di H3K36me3 (trimetilazione della lisina 36 dell'istone H3) vicino l'estremità 3' dei geni.

Significativi progressi sono anche stati compiuti nella caratterizzazione globale dei livelli di modificazione istonica nei mammiferi. Infatti, recenti studi su larga scala hanno evidenziato complesse relazioni tra espressione genica e modifiche istoniche. In linea generale, alti picchi di metilazione su H3K27 sono stati correlati con la repressione genica, mentre elevati livelli di acetilazione istonica e di metilazione su H3K4 sono stati individuati in corrispondenza delle regioni promotrici di geni attivi (*Figura 1C*). In aggiunta alle regioni promotrici queste modificazioni sono state anche osservate nelle regioni intergeniche. Inoltre, i segnali di acetilazione dell'istone H3 e H3K4me1 sono stati correlati con enhancer funzionali. Ancora, sebbene la metilazione di H3K9 sia stata collegata alla formazione dell'eterocromatina e al silenziamento genico, recenti analisi su larga scala hanno osservato come H3K9me3 sembri essere presente anche in promotori attivi.

In passato erano state compiute approfondite analisi di localizzazione per solamente poche metilazioni istoniche su una limitata copertura e risoluzione genomica, mentre sono necessarie analisi di colocalizzazione globali e di alta risoluzione di tutte le metilazioni istoniche in un intero genoma (ad esempio quello umano) per comprendere appieno le diverse correlazioni funzionali tra le varie metilazioni istoniche coinvolte in processi quali la trascrizione e la riparazione del DNA.

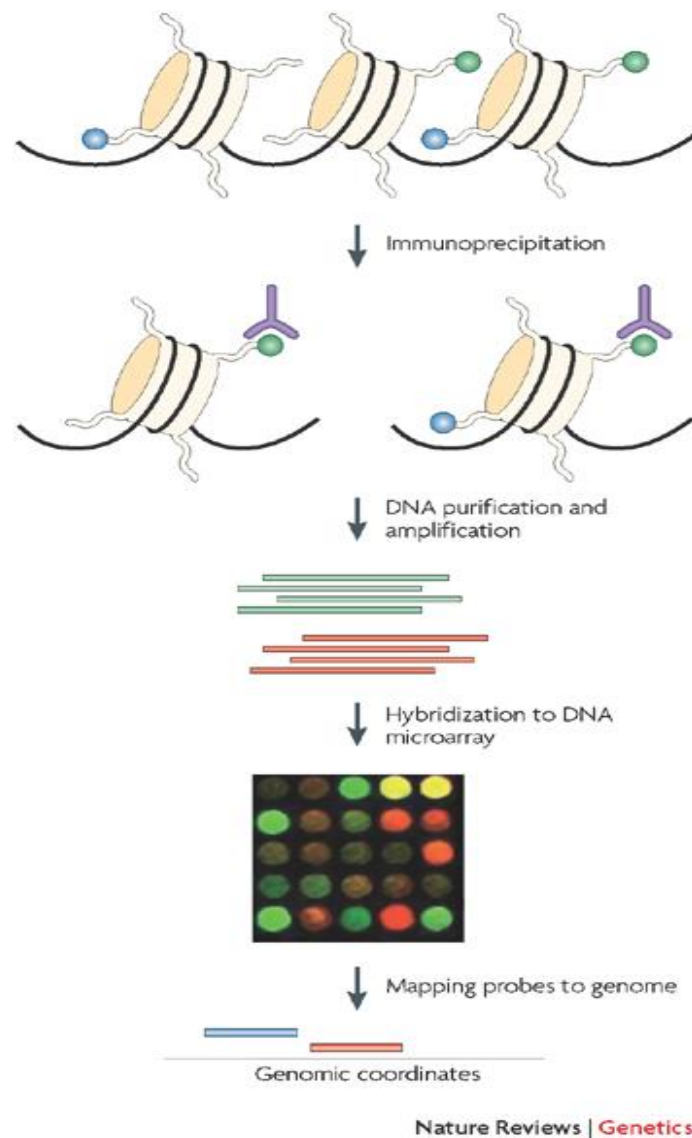
Infine, oltre a variare l'ambiente cromatinico, le modifiche istoniche sono in grado di modificare l'accesso di altre proteine regolatorie addizionali, le quali, a loro volta, contribuiscono o ad un ulteriore compattamento o alla maggiore accessibilità di un tratto del DNA assemblando complessi coattivatori o corepressori.



*Figura 1C: rappresentazione della struttura e degli effetti di due diverse modificazioni sull'istone H3. H3K27me3 è collegata con la repressione genica (metilazione colorata di giallo, sulla sinistra), mentre H3K4me3 è correlata con l'attivazione genica (metilazione colorata di verde, sulla destra) [4].*

## IMMUNOPRECIPITAZIONE DELLA CROMATINA

Importanti progressi nella caratterizzazione delle modifiche istoniche su scala genomica sono stati resi possibili, in questi ultimi anni, grazie a nuove metodiche di indagine, prima tra tutte la tecnica ChIP-on-chip, che combina l'immunoprecipitazione della cromatina (ChIP) con analisi microarray (chip). La figura di sotto riportata (*Figura 2A*) illustra brevemente la tecnica ChIP-on-chip.



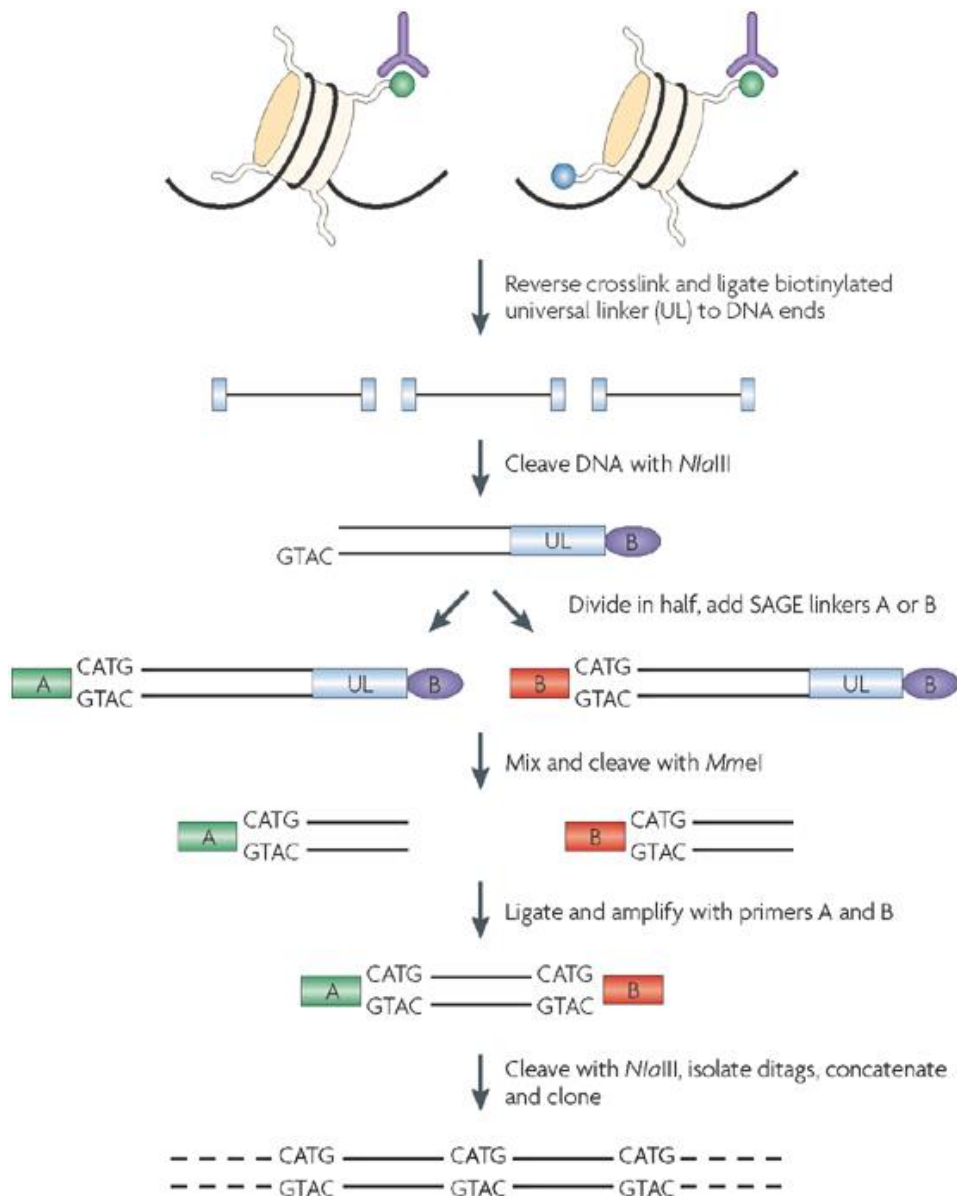
*Figura 2A: esperimento ChIP-on-chip inteso a studiare le modificazioni istoniche. La cromatina è purificata mediante immunoprecipitazione utilizzando un anticorpo specifico per una particolare modificazione istonica (mostrata in verde). Successivamente il DNA viene purificato ed amplificato. Infine il ChIP DNA viene marcato e ibridato in un microarray che contiene come sonde tutte o parti delle sequenze genomiche di un particolare tipo di cellula. Le sonde ibridate su microarray possono poi essere mappate sul genoma per ottenere le coordinate genomiche [5].*



Con la tecnica ChIP-on-chip è stato ampiamente studiato il pattern di modifiche istoniche a livello del genoma di lievito. Sulle cellule umane, invece, è stata effettuata l'analisi di localizzazione di solo poche metilazioni istoniche e solamente su una piccola parte del genoma. Dunque, nonostante l'associazione riscontrata tra modificazioni istoniche e attività cromatinica in genomi meno estesi, la tecnica non permetteva lo studio completo dei patterns combinatoriali presenti nell'intero genoma umano.

Per comprendere le complesse correlazioni tra modifiche istoniche e funzioni cellulari è necessaria un'analisi di colocalizzazione ad alta risoluzione. La tecnica ChIP-on-chip, nonostante il grande contributo apportato, presenta una serie di limitazioni, come la necessità di molti array per coprire l'intero genoma umano o come il possibile errore introdotto con l'amplificazione mediante PCR, che rendevano poco realistico il suo uso per analisi su interi genomi.

Di conseguenza è stata utilizzata una nuova tecnica (*Figura 2B*), denominata ChIP-SAGE o GMAT (genome-wide mapping technique), che combinava la ChIP con analisi seriali di espressione genica (SAGE, serial analysis of gene expression). Tuttavia anche questo metodo presentava una serie di limitazioni, come elevati costi di sequenziamento e la creazione di mappe di modificazioni istoniche di relativamente bassa risoluzione (500-1000 bp), inadeguate per il genoma umano. Per superare questi problemi è stata introdotta una nuova tecnica, detta ChIP-Seq, che utilizza tecnologie di next-generation sequencing (NGS), come il sequenziatore Solexa 1G Genome Analyzer, per sequenziare direttamente i filamenti ChIP DNA generati dalla digestione della cromatina nativa da parte della nucleasi micrococcica (MNasi). La tecnica del ChIP-Seq (*Figura 2C*) è un metodo completo, quantitativo e conveniente per analizzare le modificazioni istoniche e i siti target delle proteine in un ampio genoma, così come lo è quello umano.



Nature Reviews | Genetics

Figura 2B: combinazione dell'immunoprecipitazione della cromatina con analisi seriali d'espressione genica (ChIP-SAGE o GMAT). La ChIP-SAGE incomincia con un ChIP step, ovvero con la purificazione delle regioni cromatiniche associate con una specifica modificazione istonica (mostrata in verde) e procede come segue. Innanzitutto i crosslinks sono invertiti e le estremità del DNA sono legate a linker universali biotinilati (UL). Quindi l'enzima di restrizione *Nla*III, che riconosce la sequenza CATG, viene usato per digerire il DNA ed un linker, contenente la sequenza di riconoscimento di *Mme*I, viene legato alle estremità scisse del DNA. La digestione mediante *Mme*I produce sequenze tag di 21-22 bp dai frammenti immunoprecipitati; le sequenze tag sono concatenate, clonate in un vettore e sequenziate. Le sequenze tag possono poi essere mappate sul genoma per identificare le regioni modificate [6].

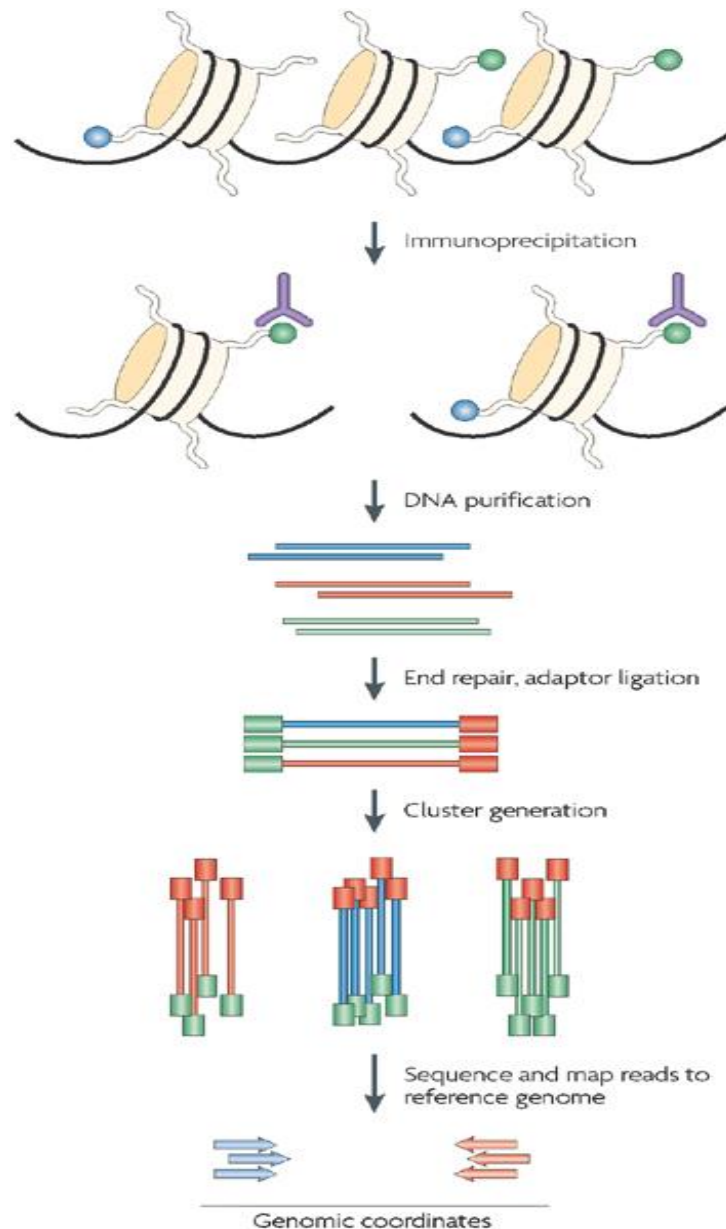


Figura 2C: immunoprecipitazione della cromatina combinata con tecniche high-throughput sequencing (ChIP-Seq). Il primo passaggio consiste nella purificazione della cromatina modificata mediante immunoprecipitazione utilizzando un anticorpo specifico per una particolare modificazione istonica (mostrata in verde). Le estremità dei filamenti ChIP DNA sono riparate e legate ad un paio di adattatori e sottoposte a limitati cicli di amplificazione mediante PCR (circa diciassette). Le molecole di DNA vengono ibridate a oligonucleotidi, covalentemente legati su una superficie solida e complementari alle sequenze adattatrici. Segue il cluster di molecole di DNA e il sequencing-by-synthesis. Le sequenze risultanti vengono mappate in corrispondenza di un genoma di riferimento per ottenere le coordinate genomiche che corrispondono ai frammenti immunoprecipitati [7].

Per determinare l'affidabilità e la riproducibilità della tecnica ChIP-Seq sono stati eseguiti due esperimenti indipendenti di analisi della distribuzione di H3K4me3 su un locus del cromosoma 19. Per un esperimento si è utilizzata la tecnica ChIP-SAGE e per l'altro la ChIP-Seq. È stato riscontrato che i dati ottenuti con la ChIP-Seq sono ampiamente coerenti con quelli ottenuti con la ChIP-SAGE (Figura 2D).

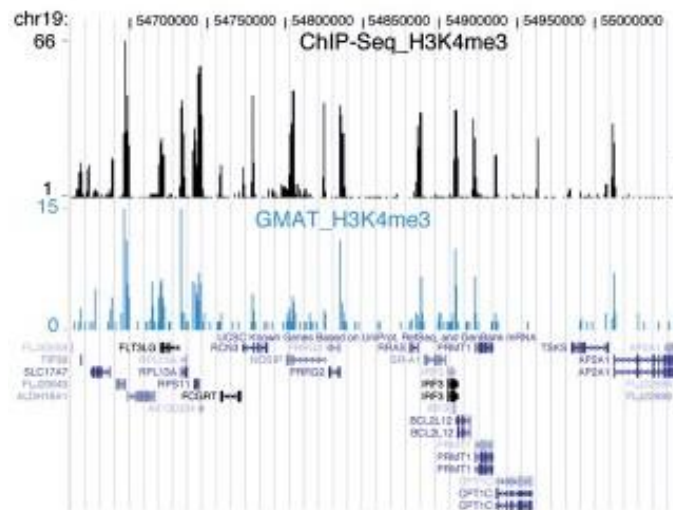


Figura 2D: confronto dei patterns della modificazione H3K4me3 in una regione di 350 Kb sul cromosoma 19 usando la ChIP-Seq (pannello superiore) e la ChIP-SAGE (pannello inferiore). I dati sono mostrati come custom tracks sul browser genomico UCSC [8].

Inoltre, è emerso come la ChIP-Seq sia più sensibile, abbia una risoluzione nettamente migliore e generi meno regioni false negative rispetto alla tecnica ChIP-SAGE. Per esempio, su un locus del gene IL-2 è stato individuato a circa 1.5 Kbp a monte del TSS un alto picco del segnale di H3K4me3 utilizzando la tecnica ChIP-Seq, mentre è stato individuato solamente un singolo tag utilizzando il metodo GMAT (Figura 2E). Il miglioramento della sensibilità e della specificità può essere attribuito all'utilizzo di mononucleosomi nativi per ChIP.

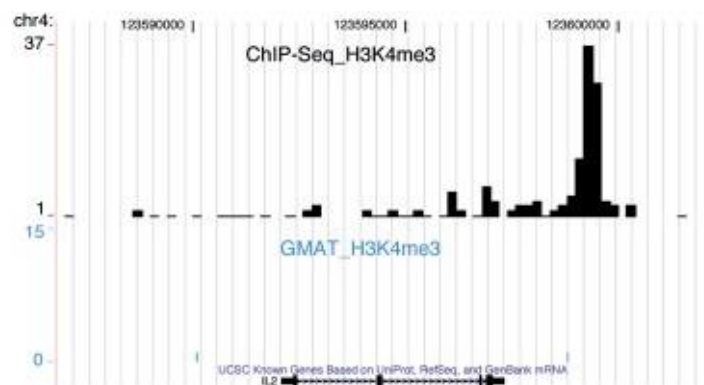


Figura 2E: confronto tra ChIP-Seq e GMAT su un locus del gene IL-2 [9].

## **CONFRONTO TRA CHIP-ON-CHIP, CHIP-SAGE E CHIP-SEQ**

### *Risoluzione*

La risoluzione di ChIP-Seq dipende dalla dimensione dei frammenti di cromatina che sono usati per la ChIP (immunoprecipitazione della cromatina) oltre all'efficienza del sequenziamento. Usando i mononucleosomi generati dalla digestione delle nucleasi micrococciche, i segnali delle modificazioni istoniche individuati dalla ChIP-Seq possono essere assegnati a singoli nucleosomi nel genoma. La risoluzione del ChIP-on-chip dipende sia dalla dimensione dei frammenti di cromatina usati per la ChIP che dalle sonde usate per i microarray. La ChIP-Seq non ha bisogno di una pre-selezione di regioni genomiche, mentre la ChIP-on-chip può analizzare solo le porzioni del genoma per cui sono state disegnate probe presenti sul microarray. La risoluzione della ChIP-SAGE dipende da quanto frequentemente i siti degli enzimi di restrizione si trovano nel DNA sottoposto a ChIP. Anche se la ChIP-SAGE può essere utilizzata per studiare interi genomi, è limitata alle regioni che hanno siti di riconoscimento per l'enzima di restrizione usato per scindere il ChIP DNA.

### *Quantificazione*

Per la ChIP-on-chip la quantificazione dipende dall'efficienza di ibridazione delle molecole di ChIP DNA alle sonde del microarray, le quali possono variare notevolmente a seconda della sequenza. Nessuna ibridazione è richiesta per la ChIP-Seq e il ChIP DNA è minimamente amplificato per generare cluster di molecole che possono essere direttamente contate attraverso la procedura di sequenziamento. La ChIP-SAGE, similmente alla ChIP-Seq, non richiede nessuna ibridazione, ma richiede molti più cicli di amplificazione mediante PCR. Come conseguenza la ChIP-Seq e la ChIP-SAGE sono, probabilmente, tecniche molto più quantitative rispetto alla ChIP-on-chip.

### *Costi*

La ChIP-Seq è sicuramente la tecnica più economica per realizzare mappe di risoluzione nucleosomica in genomi di mammifero, considerato l'odierno costo dei whole-genome tiling array (WGA). La ChIP-SAGE è più costosa della ChIP-Seq in quanto utilizza costosi metodi di sequenziamento. La ChIP-on-chip potrebbe essere la tecnica più conveniente per tracciare i profili di regioni subgenomiche.

## METILAZIONE

### CORRELAZIONI DELLE MODIFICHE ISTONICHE SULLE REGIONI PROMOTRICI CON L'ESPRESSIONE GENICA

La regione vicina ai siti d'inizio della trascrizione (*TSSs, transcription start sites*) contiene elementi critici regolatori necessari per la trascrizione. Precedenti studi su scala genomica hanno rivelato che alcune modificazioni istoniche, come H3K4me e H3K36ac, si trovano in corrispondenza di queste regioni, mentre altre, come la modificazione istonica H3K36me3, sono localizzate in regioni trascritte. Per individuare ulteriori patterns di metilazione istonica a livello dei promotori, degli enhancers, degli insulator e delle regioni trascritte e per valutare il contributo di altre modificazioni e varianti istoniche in cellule umane, sono state generate mappe genomiche ad alta risoluzione per l'analisi della distribuzione di venti istoni metilati ai residui amminoacidici di lisina (K) e arginina (R), oltre alla variante istonica H2A.Z, alla RNA polimerasi II e alla regione di legame per la proteina CTCF (insulator). Per correlare queste modificazioni con la trascrizione genica, sono stati analizzati 12726 geni di cellule umane T CD4<sup>+</sup> a riposo, i cui livelli di espressione erano noti. Questi geni sono stati raggruppati, in base al loro livello d'espressione, in 12 cluster da 1000 geni ciascuno e allineati in base al TSS. Infine, per ciascun cluster è stato rappresentato l'arricchimento ottenuto dalla ChIP-Seq in corrispondenza del TSS e della regione trascritta. I grafici dei dati così ricavati riportano per ogni modificazione studiata il numero medio di read per paio di basi individuato nella regione nell'esperimento corrispondente. I valori rappresentati sono direttamente proporzionali al livello di modificazione presente nella regione, relativamente al TSS dei geni. Di seguito vengono illustrati i grafici delle principali metilazioni istoniche presenti a livello della regione circostante il TSS di solamente quattro dei dodici gruppi dei geni considerati (*Figura 3A-L*): quello riguardante i geni altamente espressi, quello riguardante i geni silenti e quelli riguardanti gruppi di geni con un profilo di espressione intermedio.

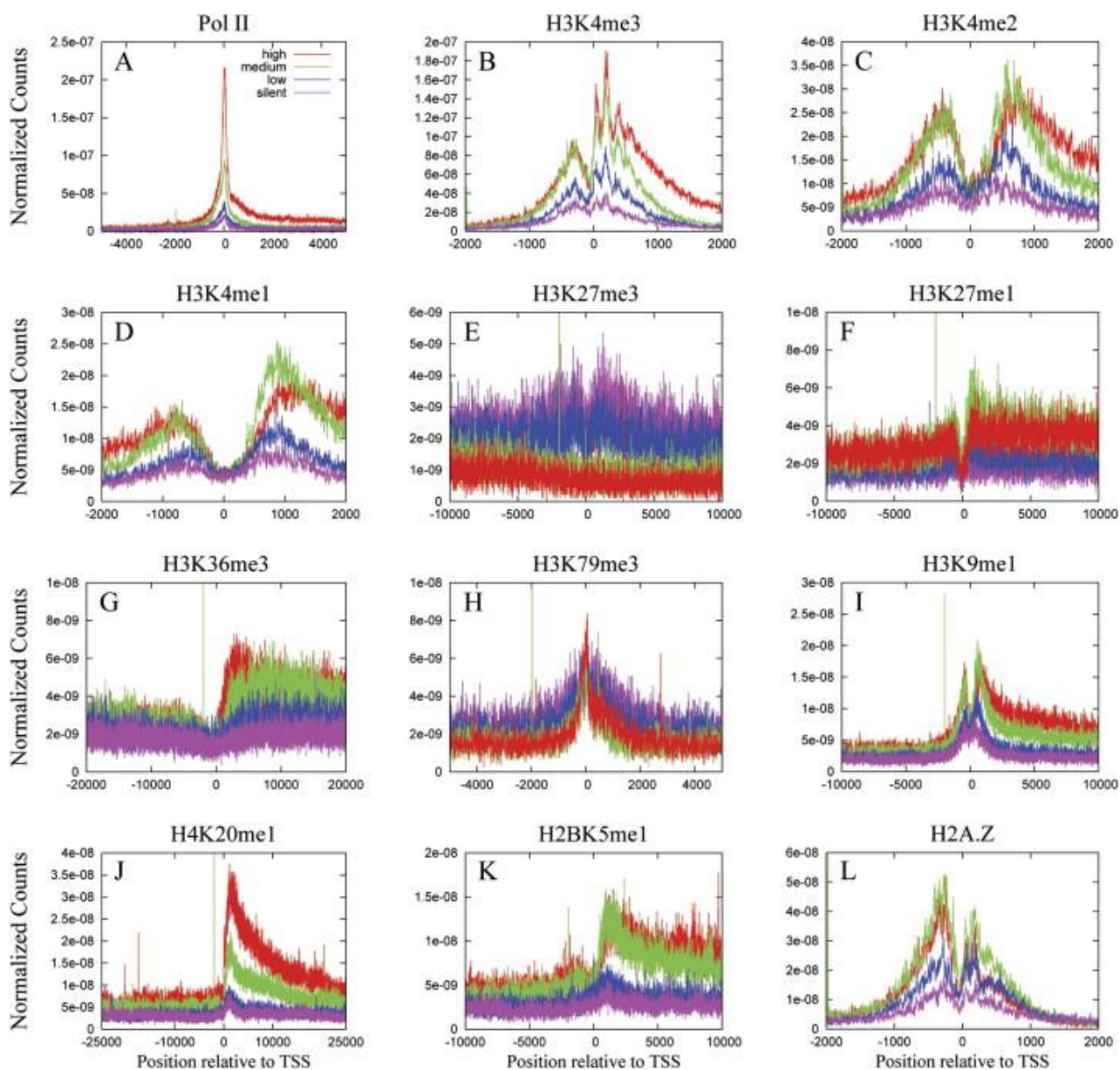


Figura 3A-L: Metilazione istonica vicino al punto d'inizio della trascrizione. Il colore rosso indica geni con un alto profilo d'espressione, il verde geni con un profilo d'espressione intermedio, il blu geni con un basso profilo d'espressione e il viola geni silenti. Per ulteriori dettagli si veda il testo [10].

### Pol II Binding

Sono stati identificati 35961 picchi (ovvero, regioni arricchite nel sequenziamento) per Pol II. Come da atteso, i segnali di Pol II sono positivamente correlati con un alto livello d'espressione genica (Figura 3A) e localizzati sull'inizio della trascrizione. È interessante notare che il 37% dei promotori dei geni silenti contengono comunque segnale per Pol II, suggerendo che essi potrebbero essere pronti per l'attivazione genica o per il mantenimento della memoria di una passata trascrizione. Inoltre, circa il 91% dei segnali di Pol II sono correlati con picchi di H3K4me3.

### *Metilazioni di H3K4*

Tutti e tre gli stati di metilazione di H3K4 sono elevati nelle zone circostanti i TSSs dei geni analizzati (*Figure 3B, 3C, 3D*). H3K4me3 è positivamente correlata con l'espressione genica. Per quanto riguarda questa modificazione si è osservato un profondo avvallamento del segnale tra -200 bp e +50 bp, il quale è correlato con la perdita del nucleosoma a livello del TSS di geni attivi. Sono stati individuati anche una serie di picchi del segnale di H3K4me3 a +50, +210 e +360, collegati col posizionamento del nucleosoma metilato vicino al TSS.

Mentre i livelli di H3K4me1 e H3K4me2 sono positivamente correlati con i livelli di trascrizione, i segnali a valle del TSS sono più elevati nei gruppi dei geni con livello d'espressione medio rispetto ai gruppi dei geni con alti livelli d'espressione. Questo potrebbe essere causato dall'alto livello di trimetilazione di H3K4 nei gruppi dei geni altamente espressi, dal momento che i tre stati di metilazione competono tra loro per ogni singola lisina.

Per ogni modificazione si sono osservati due picchi più elevati: per H3K4me1 a -900 e a +100, per H3K4me2 a -500 e a +700 e per H3K4me3 a -300 e a +100. I segnali sono progressivamente più localizzati al TSS man mano che si passa da una monometilazione, ad una dimetilazione e infine ad una trimetilazione.

### *Metilazioni di H3K27*

Dalla letteratura è noto che la metilazione di H3K27 è correlata con la repressione genica. Infatti si è riscontrato che i livelli di H3K27me3 sono più alti in promotori silenti che in promotori attivi. Tuttavia, inaspettatamente, i segnali di H3K27me1 sono più alti in promotori attivi che in promotori silenti, in particolare a valle del TSS (*Figura 3F*).

### *Metilazioni di H3K36*

In accordo con precedenti osservazioni, le quali associano H3K36me3 con regioni attivamente trascritte, i picchi di H3K36me3 sono nettamente elevati a valle dei TSSs di geni attivi, nella regione trascritta. (*Figura 3G*).

### *Metilazioni di H3K79*

I picchi di H3K79me3 sono più localizzati in promotori attivi che in promotori silenti, anche se sembrano essere più elevati in promotori silenti, salvo che in una stretta regione circostante il TSS (*Figura 3H*).



### *Metilazioni di H3K9*

Studi precedenti hanno individuato che le metilazioni di H3K9 sono implicate nella formazione dell'eterocromatina e nel silenziamento genico. Infatti sia i segnali di H3K9me2 che di H3K9me3 sono più alti in geni silenti che in geni attivi, in una regione di 10 Kb circostante il TSS. Al contrario più alti livelli di H3K9me1 sono stati individuati in promotori attivi circostanti il TSS (*Figura 3I*), suggerendo che questa modificazione potrebbe essere associata con un'attivazione trascrizionale.

### *Metilazioni di H4K20*

Dalla letteratura è noto che la metilazione di H4K20 è associata a domini di cromatina repressiva. Inoltre, recenti studi suggeriscono che la monometilazione e la trimetilazione di H4K20 hanno distribuzioni differenti lungo tutto il genoma. H4K20me3 è associata con l'eterocromatina, mentre H4K20me1 si trova in regioni promotrici o codificanti di geni attivi. Infatti H4K20me1 è stata trovata colocalizzata con H3K9me1 risultando essere strettamente correlata con l'attivazione genica e la trascrizione in regioni a valle di TSS (*Figura 3J*).

### *Metilazioni di H2BK5*

Le metilazioni di H2BK5 non sono mai state riportate in studi precedenti. Come si nota in *Figura 2K* i segnali di H2BK5me1 sono associati con le regioni dei promotori attivi localizzate immediatamente a valle del TSS. Questi dati suggeriscono che H2BK5me1 potrebbe essere un nuovo marcatore di attivazione.

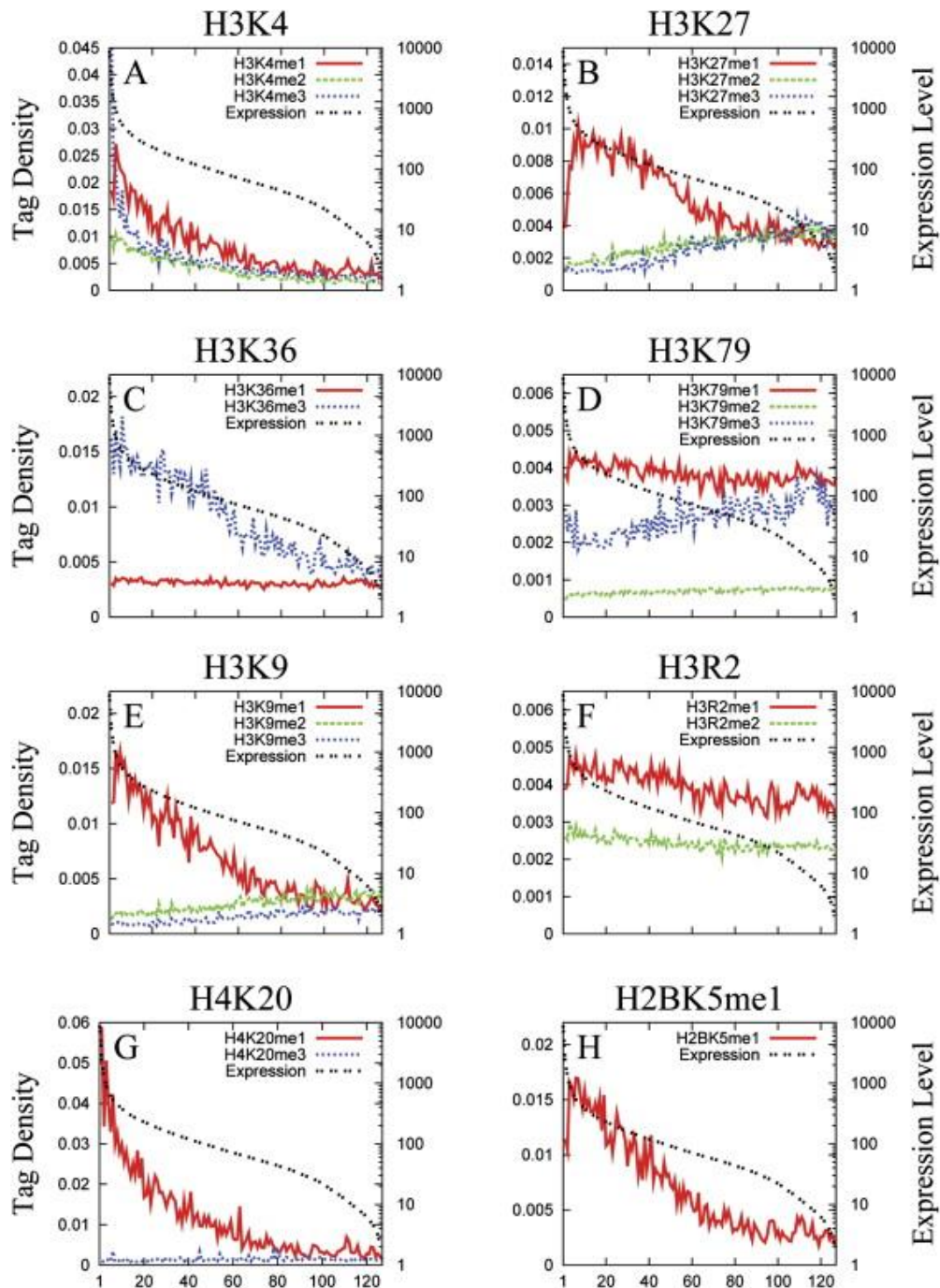
### *Variante istonica H2A.Z*

L'istone H2A.Z, variante istonica dell'istone H2A, è rimasto altamente conservato durante l'evoluzione. L'istone H2A.Z è generalmente associato a regioni cromatiniche in attiva trascrizione. L'incorporazione dell'istone H2A.Z stabilizza l'ottamero nucleosomico, ma impedisce alcune interazioni cooperative tra i nucleosomi necessari a compattare il cromosoma. Questo porta ad una struttura cromatinica più aperta, che facilita la trascrizione dei geni. I picchi di H2A.Z sono infatti elevati in regioni promotrici sia monte che a valle del TSS di geni trascritti (*Figura 3L*).

## **CONFRONTO TRA ATTIVITÀ GENICA E METILAZIONE ISTONICA IN REGIONI TRASCritte**

Le analisi effettuate indicano che i livelli di alcune modifiche ai promotori non sono correlate con l'attività genica. Precedenti studi in lievito hanno suggerito che alcune modificazioni, come H3K36me3, sono elevate solamente nelle regioni attivamente trascritte. Per studiare queste

corrispondenze in cellule umane T, 12726 geni sono stati suddivisi ancora in differenti gruppi in base alla loro espressione e la media dei loro livelli di metilazione è stata plottata in grafico contro le loro regioni trascritte (*Figura 4A-H*).



*Figura 4 (A-H): i geni sono stati raggruppati in gruppi da 100 in base al loro livello d'espressione genica, dal livello più alto a quello più basso (da sinistra a destra sull'asse delle ascisse). L'asse destro delle ordinate indica il livello d'espressione di ogni gruppo. L'asse sinistro delle ordinate mostra il livello di metilazione istonica che è stato individuato riassumendo tutti i read trovati all'interno delle regioni trascritte e normalizzati sul numero totale di paia di basi dei 100 geni [11].*

### *Modificazioni associate con il silenziamento genico*

Dalle analisi effettuate è stato osservato che H3K27me2 e H3K27me3 sono correlate con geni silenti (*Figura 4B*). È stata notata anche una moderata correlazione tra i livelli di H3K9me2 e H3K9me3 e i geni silenti (*Figura 4E*). Tuttavia, alti picchi di H3K9me3 sono stati osservati in alcuni geni attivi, come STAT1 e STAT4 (*Figura 5A*).

Inaspettatamente, i livelli di H3K79me3, che in lievito erano stati associati a geni attivamente trascritti, sono moderatamente correlati con geni silenti nelle cellule umane T, tranne che per un piccolo gruppo di geni attivi che sono associati con alti livelli di H3K79me3 (*Figura 4D*).

### *Modificazioni associate con l'attivazione genica*

Come atteso, H3K4me1, H3K4me2, H3K4me3 e H3K36me3 sono correlate con l'attivazione genica (*Figura 4A e 4C*). Coerentemente con le analisi di allineamento al TSS (fare riferimento a [Figure 3F, 3I, 3J, 3K](#)), le modificazioni H3K27me1, H3K9me1, H3K20me1 e H2BK5me1 sono positivamente correlate con i livelli d'espressione genica (*Figure 4A, 4E, 4G, e 4H*). Sia H3R2me1 che H3R2me2 sono moderatamente correlate con l'attivazione genica (*Figura 4F*).

### *Pattern di metilazione istonica di geni attivi e silenti*

Un tipico esempio di metilazione istonica in corrispondenza di loci genomici di geni attivi (STAT1 e STAT4) e di geni silenti (MYO1B) è mostrato in *Figura 5A*. I geni STAT1 e STAT4 sono marcati da alti livelli di H3K4me1, H3K4me2, H3K4me3 e Pol II nelle regioni promotrici e da alti livelli di H3K4me1, H3K4me2, H3K27me1, H3K9me1, H3K36me3, H2BK5me1 e H4K20me1 all'interno delle regioni trascritte. Al contrario, il gene MYO1B è associato con alti livelli di H3K27me2, H3K27me3, H3K9me2 e H3K79me3. È interessante notare che la variante istonica H2A.Z mostra picchi elevati in corrispondenza delle regioni promotrici dei geni STAT1 e STAT4 e nelle regioni intergeniche tra i geni STAT4 e MYO1B, mentre mostra picchi a livelli più bassi, ma uniformemente distribuiti, lungo tutto il gene MYO1B.

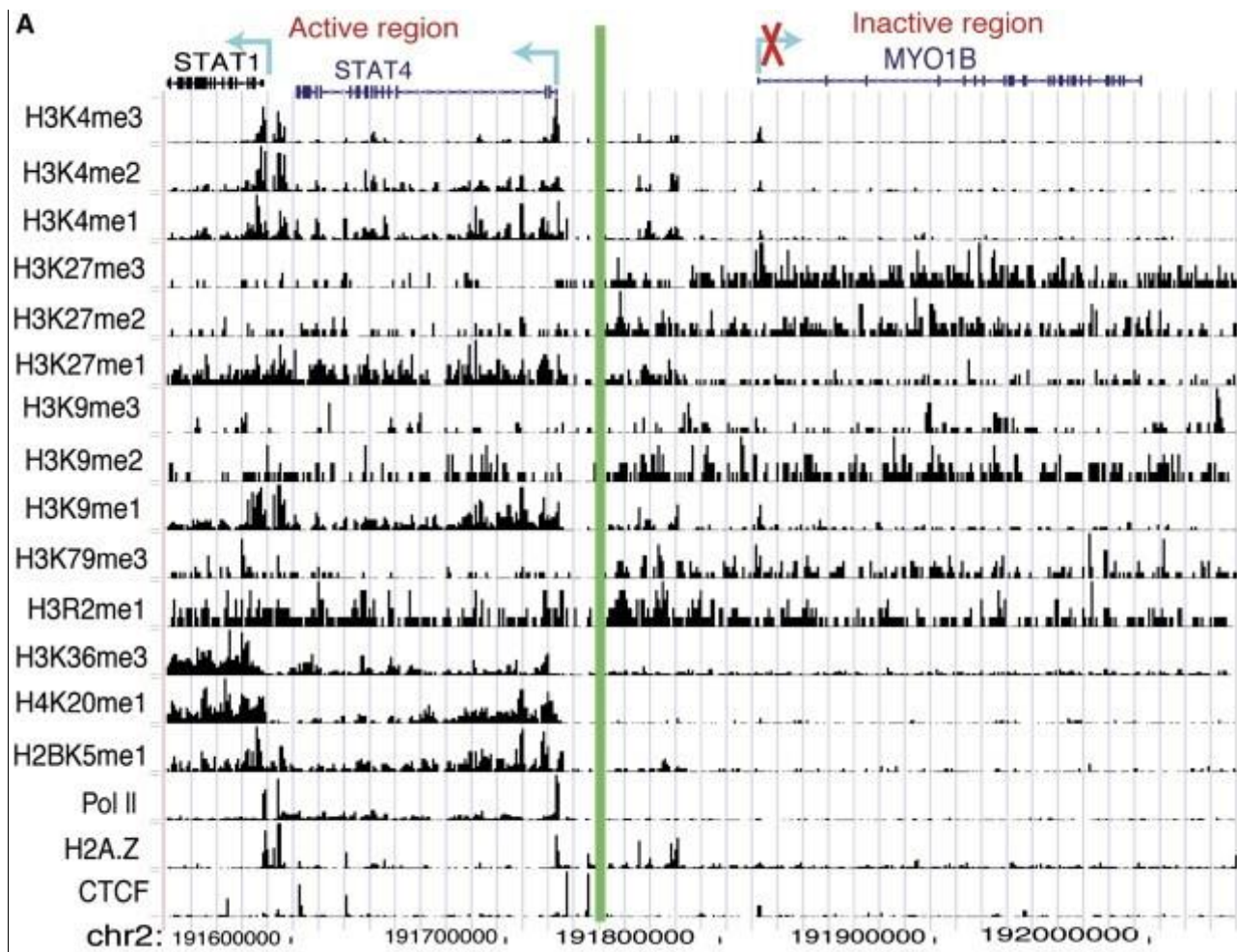


Figura 5A: esempio di pattern di metilazione istonica sia su geni attivi che inattivi mostrato come custom tracks su browser genomico UCSC [12].

### Metilazione istonica su Insulator

Gli insulator sono elementi di DNA che separano differenti domini cromatinici. CTCF è una proteina a dominio zinc-finger che riconosce e lega le regioni insulator nei vertebrati. Per determinare come gli insulator, legati da CTCF, demarcano domini cromatinici definiti da metilazione istonica, è stata determinata anche la distribuzione genomica dei siti di legame CTCF nelle cellule umane T CD4<sup>+</sup>, sfruttando sempre la tecnica ChIP-Seq. Sono stati così individuati 20262 siti di CTCF, dei quali 8308 sono localizzati nelle regioni intergeniche, 6305 in regioni trascritte e 5649 all'interno di una regione circostante 2 Kb il TSSs. Inoltre, si è notato che alcuni dei siti di legame di CTCF si trovano tra domini di cromatina attiva e silente. Per esempio il gene attivamente trascritto PPP5C, marcato da alti livelli di H3K27me1, è compreso tra domini di cromatina repressiva, la quale è marcata da alti livelli di H3K27me3. I siti di legame di CTCF tracciano chiaramente i confini tra questi due differenti domini cromatinici, come è possibile notare in *Figura 5.1*.

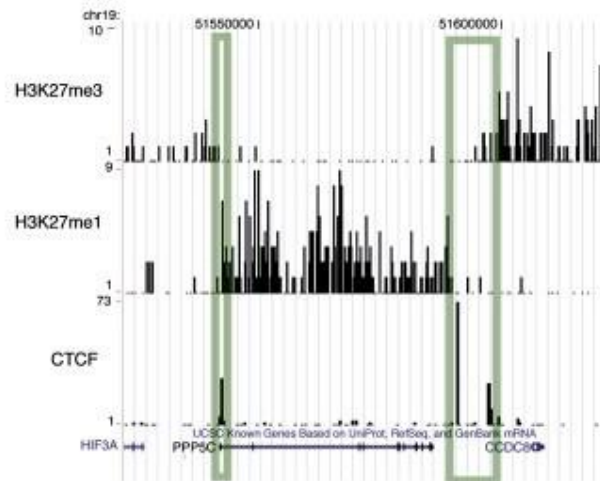


Figura 5.1: Il legame di CTCF segna chiaramente i confini tra domini cromatinici attivi e inattivi. I segnali di H3K27me3 (marca domini cromatinici repressivi), H3K27me1 (marca domini cromatinici attivi) e CTCF sono mostrati sul gene PPP5C [13].

Per individuare il pattern generale di metilazione istonica su sequenze insulator sono state allineate tutte le isole CTCF individuate (fatta eccezione per quelle vicino ai segnali di Pol II) e comparate con tutte le metilazioni istoniche. Come mostrato in Figura 5B, tutti e tre gli stati di metilazione di H3K4 sono elevati. Inoltre è interessante notare come la variante istonica H2A.Z e H3K9me1 siano altamente arricchite, ma non H3K9me2 e H3K9me3.

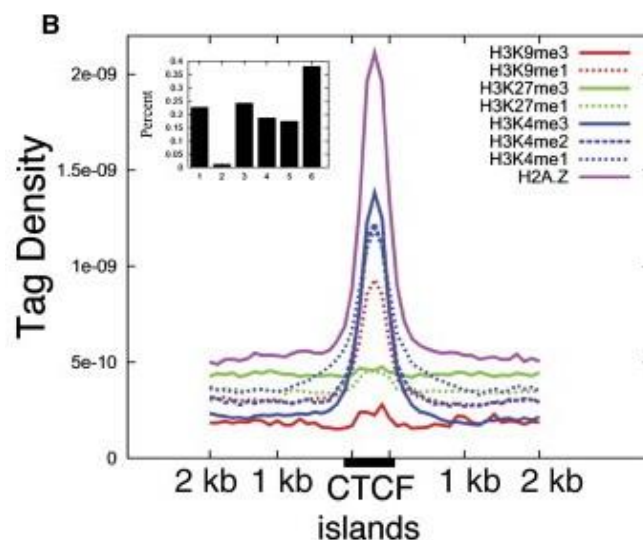


Figura 5B: patterns di metilazione istonica su insulator legati da CTCF. Tutte le isole CTCF, escluse quelle sovrapposte ai segnali di Pol II, sono state allineate. I read delle metilazioni istoniche sono stati misurati lungo tutte le isole. L'inserto in alto a sinistra mostra la percentuale delle isole CTCF associate con le varie modificazioni; i numeri riportati sull'asse delle ascisse corrispondono a come indicato di seguito: 1=H2A.Z, 2=H3K9me3, 3=H3K9me1, 4=H3K4me3, 5=H3K4me2, 6=H3K4me1 [14].



### Metilazione istonica su Enhancers

Per descrivere il pattern di modificazione istonica su elementi enhancers, sono stati esaminati alcuni noti enhancers funzionali per stabilire la loro associazione con le metilazioni istoniche, quali H3K4me3 e H3K4me1, e la variante istonica H2A.Z. Tre enhancers distali, nominati CNS1, CNS2 e CNS22, che regolano l'espressione dell'interferone  $\gamma$  nelle cellule T, sono stati associati con H3K4me1, H3K4me3 e H2A.Z (Figura 5C).

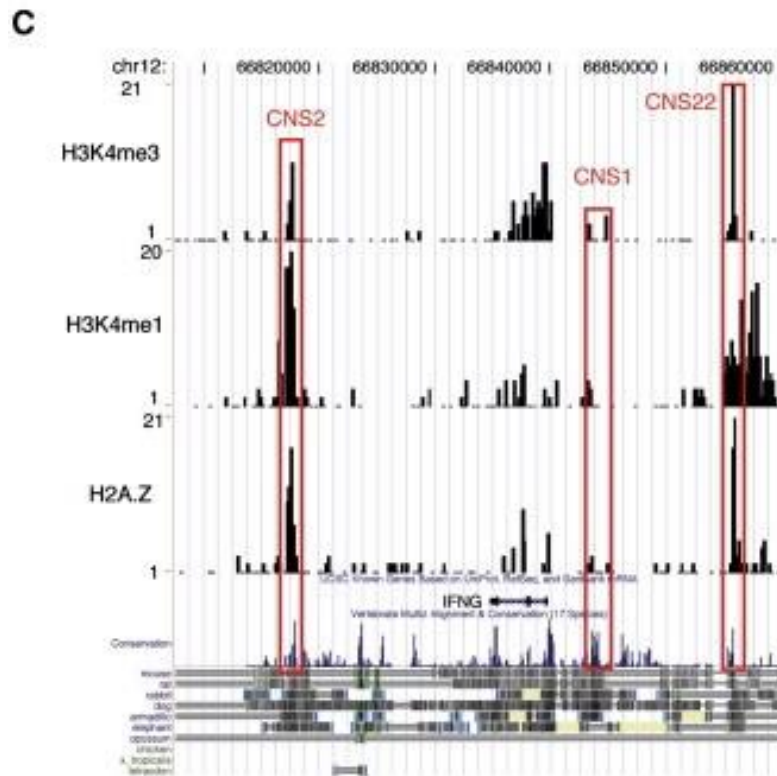


Figura 5C: associazione di H3K4me3, H3K4me1 e H2A.Z con enhancers noti del gene IFN- $\gamma$ . I tre enhancers conosciuti del gene IFN- $\gamma$  sono evidenziati nel rettangolo rosso e nominati rispettivamente CNS1, CNS2, CNS22. I dati sono visualizzati come custom track su browser genomico UCSC [15].

Basandosi sulle analisi finora effettuate, è possibile riassumere le caratteristiche epigenetiche su insulator, enhancers, promotori e regioni trascritte di geni sia attivi che inattivi (Figure 5E e 5F). Geni attivi sono caratterizzati da alti livelli di H3K4me1, H3K4me2, H3K4me3, H3K9me1 e H2A.Z in regioni circostanti i TSSs ed elevati livelli di H2BK5me1, H3K36me3, H3K27me1 e H4K20me1 in regioni a valle del TSS e lungo tutte le regioni trascritte. Al contrario, geni inattivi sono caratterizzati da bassi o trascurabili livelli di metilazione di H3K4 nelle regioni promotrici; da alti livelli di H3K27me3 e H3K79me3 in regioni promotrici e nel gene-body; da bassi o trascurabili livelli di H3K36me3, H3K27me1, H3K9me1 e H4K20me1; da bassi livelli, ma uniformemente distribuiti, della variante istonica H2A.Z.

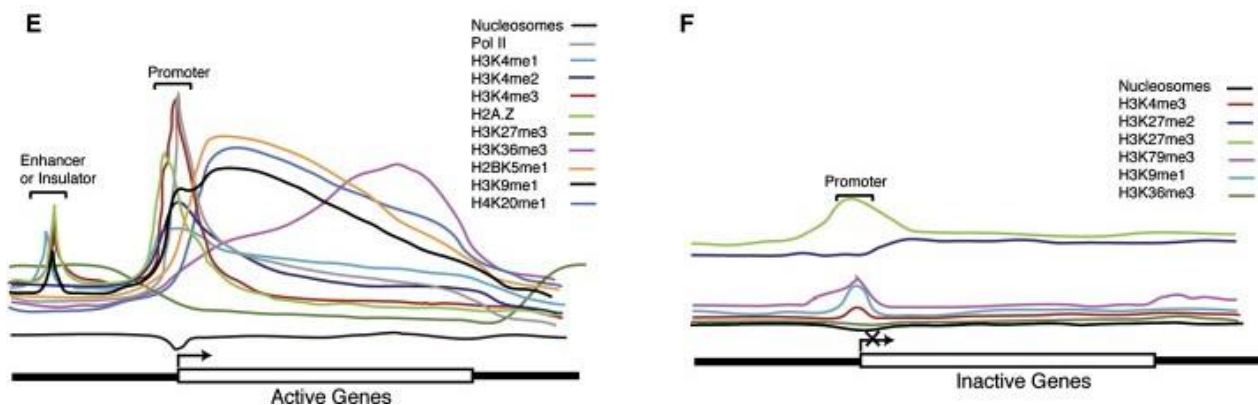


Figure 5E e 5F: patterns caratteristici di metilazione istonica di geni sia attivi che inattivi [16].

### BREAKPOINT LOCALIZZATI IN DOMINI DI CROMATINA ATTIVA

Le traslocazioni cromosomiche sono la causa maggiore di diversi sottotipi di linfomi e leucemie. Specifici riarrangiamenti cromosomici si verificano frequentemente in corrispondenza di alcuni punti caldi (hot spot) del genoma. È stato ipotizzato che questi specifici punti di rottura possano essere associati con domini di cromatina attiva. Così sono state comparate le posizioni dei punti di rottura cromosomici elencati nel Gross Rearrangement Breakpoint Database con alcune mappe di metilazioni istoniche. Dalla letteratura è noto che la traslocazione cromosomica all'interno della catena pesante IgA del gene IGHA1 è collegata a molte leucemie e linfomi. Dal confronto è emerso che il punto di rottura, sul locus del gene IGHA1, associato al linfoma delle cellule T, è localizzato al livello di modificazioni istoniche, come H3K4me1, H3K4me2, H3K4me3, e H3K9me1, mentre i punti di rottura in corrispondenza di questo locus in cellule T non cancerogene, non sono associati con queste modificazioni istoniche attivatrici. Inoltre il punto di rottura sul locus del gene IGHA1 è stato associato anche con la variante istonica H2A.Z e i siti di legame di CTCF (Figura 6C). Tra gli 84 punti di rottura individuati nelle cellule T cancerogene, il 62% (52 breakpoint) sono stati colocalizzati con i segnali di H3K4me3; mentre, tra i 180 punti di rottura individuati in cellule T non cancerogene, solo il 26% (47 breakpoint) è stato associato con alti livelli di H3K4me3. Questi risultati suggeriscono che la struttura aperta della cromatina, mantenuta dalla famiglia delle proteine MLL, potrebbe contribuire alla rottura cromosomica e potrebbe spiegare perché certe traslocazioni cromosomiche sono individuate in determinati tipi di cellule cancerogene.

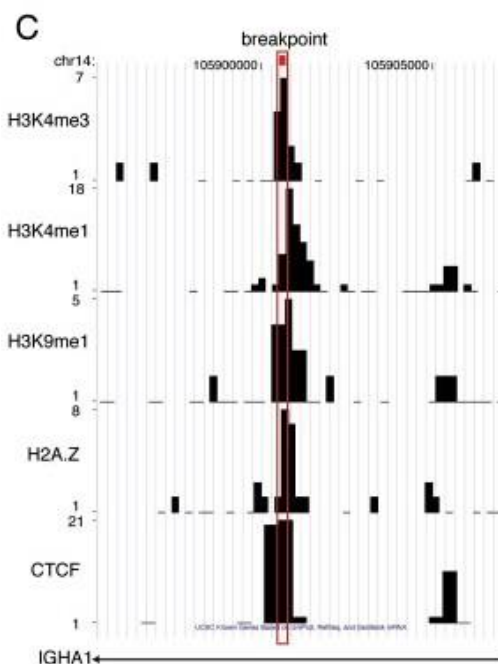


Figura 6C: punti di rottura cromosomica individuati in regioni in cui le modificazioni istoniche sono associate con strutture di cromatina aperta. I punti di rottura identificati all'interno del gene *IGHA1* e associati con le cellule T cancerogene sono evidenziati nel rettangolo rosso [17].

#### PREDIZIONE DI NUOVE UNITÀ TRASCRIZIONALI USANDO I PATTERNS DI METILAZIONE ISTONICA

L'accuratezza della predizione genica è necessaria per comprendere il genoma umano. Recenti studi hanno suggerito l'esistenza di un largo numero di trascritti intergenici non caratterizzati. A tutt'oggi non si conosce se si tratti di prodotti di scarto della trascrizione o se siano originati da TSSs di unità trascrizionali non note. Inoltre, anche la maggior parte dei geni conosciuti presenta TSSs alternativi, e non è ancora chiaro quali di questi siano utilizzati in particolari tessuti. Si è così cercato di studiare la possibilità di confermare TSS noti o di predirne di nuovi utilizzando i patterns di metilazione istonica circostanti. Sono stati quindi esaminati alcuni loci genomici con pattern di metilazione istonica caratteristici di TSSs attivi, e l'esistenza di possibili trascritti è stata valutata mediante RT-PCR. Ad esempio, il gene *DHX37* è caratterizzato da alti livelli di H3K36me3 e H2BK5me1 (Figura 7A), che suggeriscono che il gene sia attivamente trascritto. Tuttavia, al TSS del gene sono stati individuati solo bassi livelli di Pol II binding e H3K4me3 (regioni evidenziate in rosso in Figura 7A). Inoltre i livelli di H3K36me3 sono più alti vicino al TSS di *DHX37* che all'estremità 3' dello stesso gene, il che è l'opposto del pattern atteso (si veda Figura 3G). In effetti, sono identificabili segnali molto forti a circa 10 Kbp a valle del gene *DHX37*, i quali non sono stati colocalizzati con alcuna unità genica predetta o trascritta. Analisi di RT-PCR (Figura 7B) confermano l'esistenza di un trascritto nella regione colorata in rosso in Figura 7A.



Complessivamente, i dati suggeriscono l'esistenza di un nuovo TSS in direzione opposta al gene DHX37.

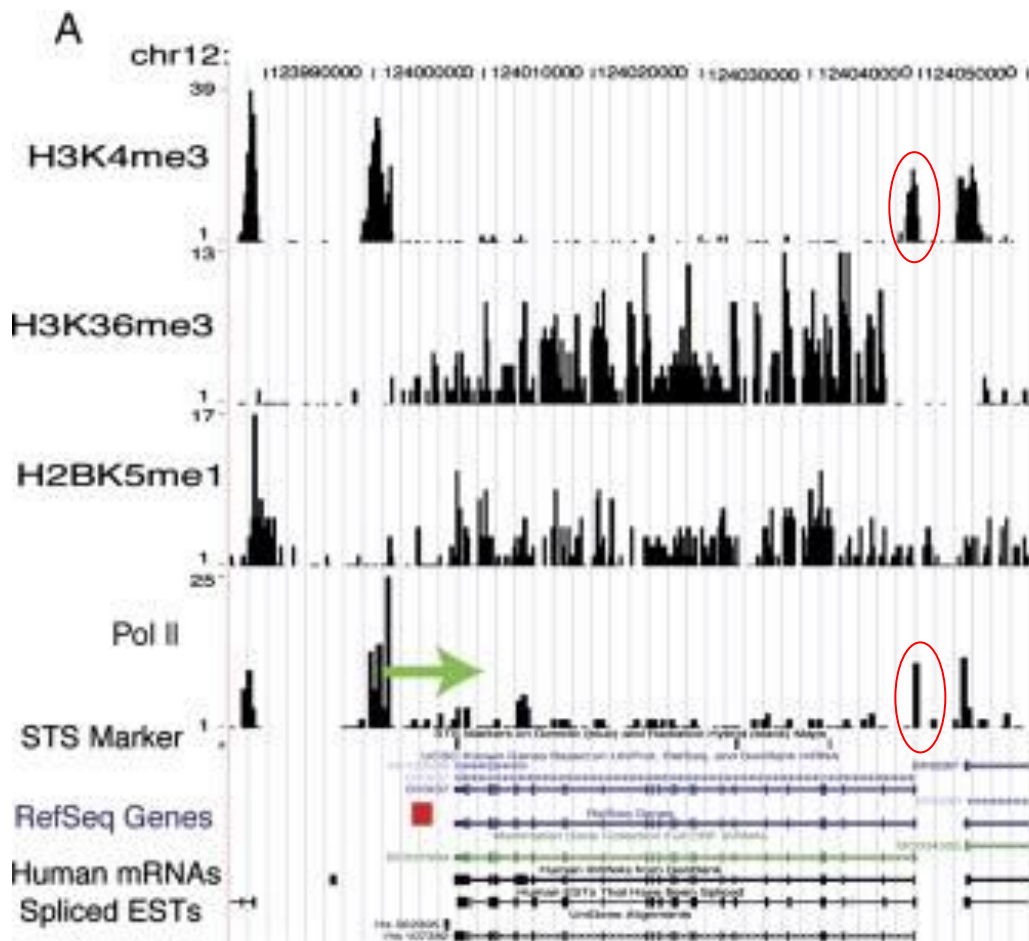


Figura 7A: un potenziale nuovo TSS a valle del gene DHX37. Sono mostrati i segnali di H3K4me3, H3K36me3, H2BK5me1 e Pol II. La potenziale direzione del trascritto è indicata dalla freccia vicino al segnale di Pol II. La regione testata per l'esistenza del trascritto è evidenziata dal rettangolo rosso. I dati sono mostrati come custom tracks sul browser genomico UCSC [18].

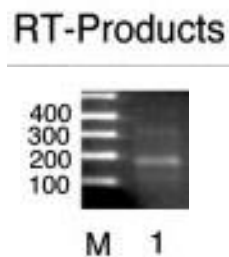
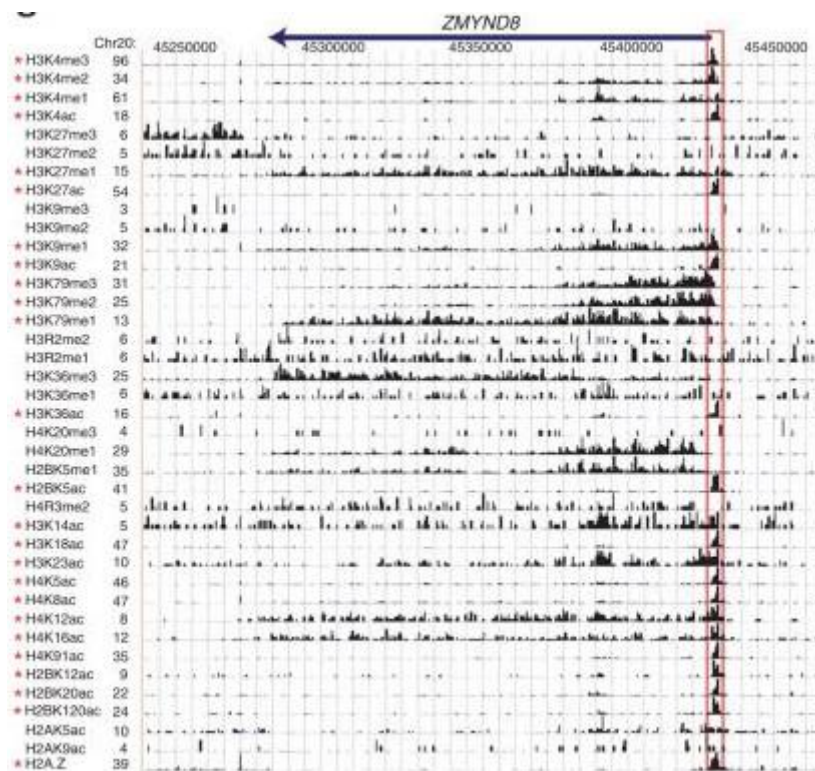


Figura 7B: analisi RT-PCR di sospetta unità trascrizionale. La lettera M indica la dimensione del DNA marker [19].

## ACETILAZIONE

Per esaminare la distribuzione delle acetilazioni istoniche a livello di differenti regioni funzionali, sono state generate diverse mappe, per gruppi di 1000 geni clusterizzati in base al loro livello d'espressione, per regioni comprendenti il TSS, per regioni comprendenti l'intero gene e per regioni estese 5 Kbp sia a monte che a valle rispetto al TSS. Si è così verificato come tutte le acetilazioni siano generalmente correlate positivamente con l'espressione genica, coerentemente con il loro coinvolgimento nell'attivazione trascrizionale. Tuttavia, dai dati ricavati, si deduce che differenti acetilazioni potrebbero caratterizzare differenti regioni geniche. Ad esempio H2AK9ac, H2BK5ac, H3K9ac, H3K18ac, H3K27ac, H3K36ac e H4K91ac sono principalmente localizzate nelle regioni circostanti il TSS, mentre H2BK12ac, H2BK20ac, H2BK120ac, H3K4ac, H4K5ac, H4K8ac, H4K12ac e H4K16ac sono elevate nelle regioni promotrici e trascritte. Questi risultati sono coerenti con quelli di precedenti studi che hanno individuato che l'istone acetil transferasi (*HAT*, *histone acetyl transferases*) può essere associata a differenti regioni geniche (ad esempio PCAF è associata con l'elongation-component della RNA Pol II). I patterns di distribuzione di queste acetilazioni istoniche sono meglio spiegate prendendo in considerazione il gene *ZMYND8*, anche chiamato *PRKCPB1*, espresso nelle cellule umane T CD4<sup>+</sup> (*Figura 8A*). Le regioni promotrici (evidenziate nel rettangolo rosso in *Figura 8A*), definite come regioni circostanti 2Kb il TSS, sono associate a 25 modificazioni.



*Figura 8A: patterns di modificazione cromatinica sul gene ZMYND8 (PRKCPB1). Le modificazioni significative, individuate nelle regioni a -1Kb e +1Kb dal TSS, sono indicate dall'asterisco sulla sinistra [20].*

Dalla letteratura è noto che le modificazioni istoniche sono associate a elementi regolatori critici per l'attivazione o la repressione trascrizione. Allo scopo di identificare cooperazioni tra diverse modificazioni istoniche, sono stati esaminati i segnali di modificazione alle regioni promotrici. È stata così identificata una particolare combinazione, nominata “backbone”, consistente di 17 modificazioni (H2A.Z, H2BK5ac, H2BK12ac, H2BK20ac, H2BK120ac, H3K4ac, H3K4me1, H3K4me2, H3K4me3, H3K9ac, H3K9me1, H3K18ac, H3K27ac, H3K36ac, H4K5ac, H4K8ac e H4K91ac) associata a 3286 promotori. Per esaminare se queste modificazioni tendono a coesistere, sono stati contati i promotori mancanti di una di queste modificazioni. In particolare, solamente un piccolo numero di promotori è stato identificato come mancante di una di queste diciassette modificazioni, rispetto ai 3286 promotori che possedevano tutte e diciassette le modificazioni (Figura 8B).

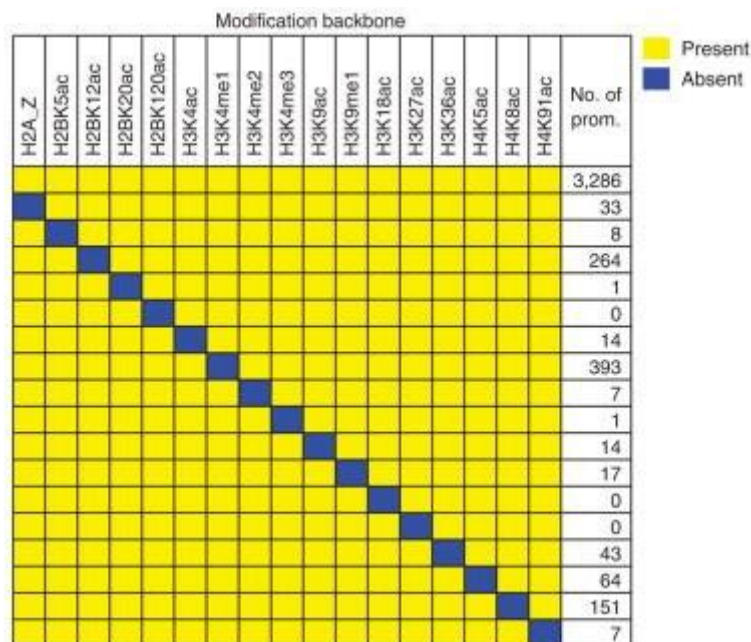


Figura 8B: modificazione “backbone”: gruppo di 17 modificazioni che tendono a coesistere. Sulla destra è indicato il numero dei promotori associati a tutte le 17 modificazioni (rettangolo giallo) e il numero dei promotori mancanti di una di queste 17 modificazioni (rettangolo blu) [21].

La tendenza di queste diciassette modificazioni a coesistere suggerisce che la maggior parte di esse siano correlate le une con le altre a livello di singolo nucleosoma. Sebbene i geni associati con queste modificazioni tendano ad avere più alti livelli di espressione, le modificazioni istoniche di per sé non determinano l'espressione genica, ma potrebbero funzionare cooperativamente per preparare la cromatina all'attivazione trascrizionale mediata dai fattori di trascrizione.

## CONCLUSIONI

Ampi studi hanno stabilito che l'acetilazione istonica è principalmente associata con l'attivazione genica, mentre la metilazione, a seconda della sua posizione, del suo stato, e del residuo coinvolto, è associata sia alla repressione che all'attivazione genica. Ad esempio, H3K79me1 è modestamente correlata con l'attivazione genica, H3K79me2 è associata con la repressione genica e H3K79me3 non mostra alcuna preferenza verso geni attivi o silenti; alti livelli di H3K36me3, H3K27me1, H3K9me1, H4K20me1, e H2BK5me1 marcano le regioni trascritte; H3K36me3 mostra elevati picchi vicino l'estremità 3', mentre H3K9me1 e H2BK5me1 mostrano alti livelli del segnale vicino all'estremità 5' e H3K27me1 ha una distribuzione uniforme in tutte le regioni trascritte.

Importante tema di ricerca nell'era postgenomica è quello di identificare elementi regolatori funzionali in genomi di eucarioti superiori. A tal scopo è stata utilizzata la genomica comparata per identificare sequenze altamente conservate non codificanti, come gli enhancers. Tuttavia questo metodo presenta una serie di limitazioni, tra cui l'impossibilità di identificare enhancers specie-specifici e l'incapacità di fornire informazioni su potenziali enhancer in particolari tessuti. L'acetilazione istonica è stata usata con successo per predire un ampio set di enhancer funzionali, conservati e non conservati, in cellule umane T. In contrasto, recenti analisi su larga scala, hanno associato i segnali di H3K4me3 e tutti e tre gli stati di metilazione di H3K4 con elementi enhancer. Sebbene sia gli enhancer che i promotori siano associati a tutti e tre gli stati di metilazione di H3K4, con la variante istonica H2A.Z e con H3K9me1, promotori attivi sono caratterizzati da alti livelli di H3K27me1, H3K36me3, H3K9me1, H4K20me1, e H2BK5me1 a valle del TSSs (si veda [Figura 5E](#)). Questo particolare pattern di promotori attivi può essere utilizzato per individuare nuove unità trascrizionali o confermare TSS di geni sospetti (fare riferimento a [Figura 7A](#)).

Si è dimostrato inoltre come il sequenziamento diretto del ChIP DNA utilizzando tecniche next-generation sequencing (in questo studio è stato utilizzato il sequenziatore Solexa 1G Genome Analyzer per la tecnica ChIP-Seq) sia un efficiente metodo per la mappatura genome-wide della distribuzione delle modifiche istoniche e dei siti target delle proteine cromatiniche. Tale tecnica è particolarmente adatta per studiare la posizione delle proteine e le modificazioni della cromatina in ampi genomi, come, ad esempio, quello umano. Questo approccio produce dati di elevata qualità e sarà senza dubbio utilizzato negli anni a venire per comprendere meglio il contributo dell'espressione genica e dell'ereditarietà epigenetica nel genoma umano.

## BIBLIOGRAFIA

1. Artem Barski, Suresh Cuddapah, Kairong Cui, Tae-Young Roh, Dustin E. Schones, Zhibin Wang, Gang Wei, Iouri Chepelev and Keji Zhao. High-Resolution Profiling of Histone Methylations in the Human Genome. *Cell* 129, 823-837 (2007).
2. Zhibin Wang, Chongzhi Zang, Jeffrey A Rosenfeld, Dustin E Schones, Artem Barski, Suresh Cuddapah, Kairong Cui, Tae-Young Roh, Weiqun Peng, Michael Q Zhang & Keji Zhao. Combinatorial patterns of histone acetylations and methylations in the human genome. *Nature Genetics*, 897-902 (2007).
3. Dustin E. Schones and Keji Zhao. Genome-wide approaches to studying chromatin modifications. *Nature Genetics*, 179-191 (2008).
4. David L. Nelson e Michael M.Cox. *Lehninger Principles in Biochemistry* 5<sup>th</sup> Edition (2010).
5. Donald Voet and Judith Voet. *Biochemistry* 2<sup>nd</sup> Edition (1995).
6. Jeffrey C. Hansen. Conformational Dynamics of the Chromatin Fiber in Solution: Determinants, Mechanism, and Functions. *Annual Review of Biophysics and Biomolecular Structure* 31, 361-392 (2002).
7. Bradley E. Bernstein, Alexander Meissner and Eric S. Lander. The Mammalian Epigenome. *Cell* 128, 669-681 (2007).