

lezione 21-22

venerdì 4 Dicembre 2009

corso di genomica

a.a. 2009/10

aula 6a ore 14.00-16.00

corso di laurea specialistica
magistrale Biotecnologia

lezione 11 Dicembre sequenziamento shot-gun metodo
pyrofosfato 454 e 480 Roche. Dr.L.Rodriguez :[rinviata](#)

lezione 15 Dicembre Programmi informatici per confronti
genomici. Dr.P. D'addabbo

Genomic views of distant-acting enhancers

Axel Visel, Edward M. Rubin & Len A. Pennacchio; Nature vol 461, 10 Sept. 09; 199-205

In contrast to protein-coding sequences, the significance of variation in non-coding DNA in human disease has been minimally explored. A great number of recent genome-wide association studies suggest that noncoding variation is a significant risk factor for common disorders, but the mechanisms by which this variation contributes to disease remain largely obscure. Distant-acting transcriptional enhancers — a major category of functional non-coding DNA — are involved in many developmental and disease-relevant processes. Genomewide approaches to their discovery and functional characterization are now available and provide a growing knowledge base for the systematic exploration of their role in human biology and disease susceptibility.

functional non coding DNA

first associations from WGA and human pathologies of variants of non coding seq.

es: 58kb linkage diseq.w/o prot.coding genes crms 9p21
risk for coronary artery disease far 60kb from the first coding gene

meta-analysis of ~1200 SNPs most significantly assoc in WGA
analyses published within March 5th 2009

con sovrastima di LD (per non perdere info.) 472 /1170 regioni
non hanno sovrapposizione con esoni o SNPs linked

più di un terzo di seq. non codif. hanno varianti assoc. a
patologie o ai fenotipi della ricerca WGA (qui WGAS)

spiegazione: elementare Watson

andiamo a finire sullo stesso carugio:

one possibility that could explain these GWAS hits is that the non-coding intervals contain enhancers, a category of gene regulatory sequence that can act over long distances. A simplified view of the current understanding of the role of enhancers in regulating genes is summarized in Fig. 1.

ma oltre a questo stanno tutte le altre cose che abbiamo già dette a proposito della conformazione-architettura variabile della cromatina, posizionamento ed epigenetica che sono fortemente influenzate dalle varianti alleliche

ci sarà un secondo codice genetico?

insulators and silencers

isolatori

e

silenziatori

The docking of RNA polymerase II to proximal promoter sequences and transcription initiation are fairly well characterized; by contrast, the mechanisms by which insulator and silencer elements buffer or repress gene regulation, respectively, are less well understood.

Transcriptional enhancers are regulatory sequences that can be located upstream of, downstream of or within their target gene and can modulate expression independently of their orientation.

In vertebrates, enhancer sequences comprise clustered aggregations of transcription-factor-binding sites. When appropriate occupancy of transcription-factor-binding sites is achieved, recruitment of transcriptional coactivators and chromatin remodelling proteins occurs.

rimodellamento - looping

l'aggregato cromatina proteine risultante (probabilmente liberando i nucleosomi) facilita il looping e attivazione del gene mediata dal promotore

molti geni regolati da serie di enhancers ognuno con compiti diversi per il pattern di espressione dei mRNA (i cui effetti additivi visti anche su geni eterologhi e geni reporter).

GWAS non possono mostrare i meccanismi molecolare (solo identificare le regioni)

Studi successivi dei singoli loci mostrano cambiamenti di funzione dei diversi alleli di enhancer attivi a distanza e associazione a patologie

- analisi delle strategie per trovare i siti e le funzioni su tutto il genoma

ipotesi su malattie e regioni non codificanti

It is often less recognized that in addition to the integrity of the protein-coding sequences, human health critically also depends on the spatially, temporally, and quantitatively correct expression of those genes

Genetic disease could equally be caused by disruption of the regulatory mechanisms that ensure proper gene expression. The term “position effect” is used in those situations where the expression level of a gene is deleteriously affected by an alteration in its chromosomal environment, while maintaining an intact transcription unit.

WGAS found many regions without codifying exons.

metodi iniziali

ci possono essere malattie associate a regioni regolative?

prime evidenze nelle talassemie scoperte nei riarrangiamenti che mostravano solo mutazioni in regioni non codificanti

(c'erano già evidenze di traslocazioni nei linfomi di Burkitt con enhancer delle Ig14q32 e c-myc 8q24 = 8/14) ricordatevi dei chrms territories e del perchè delle traslocazioni ricorrenti.

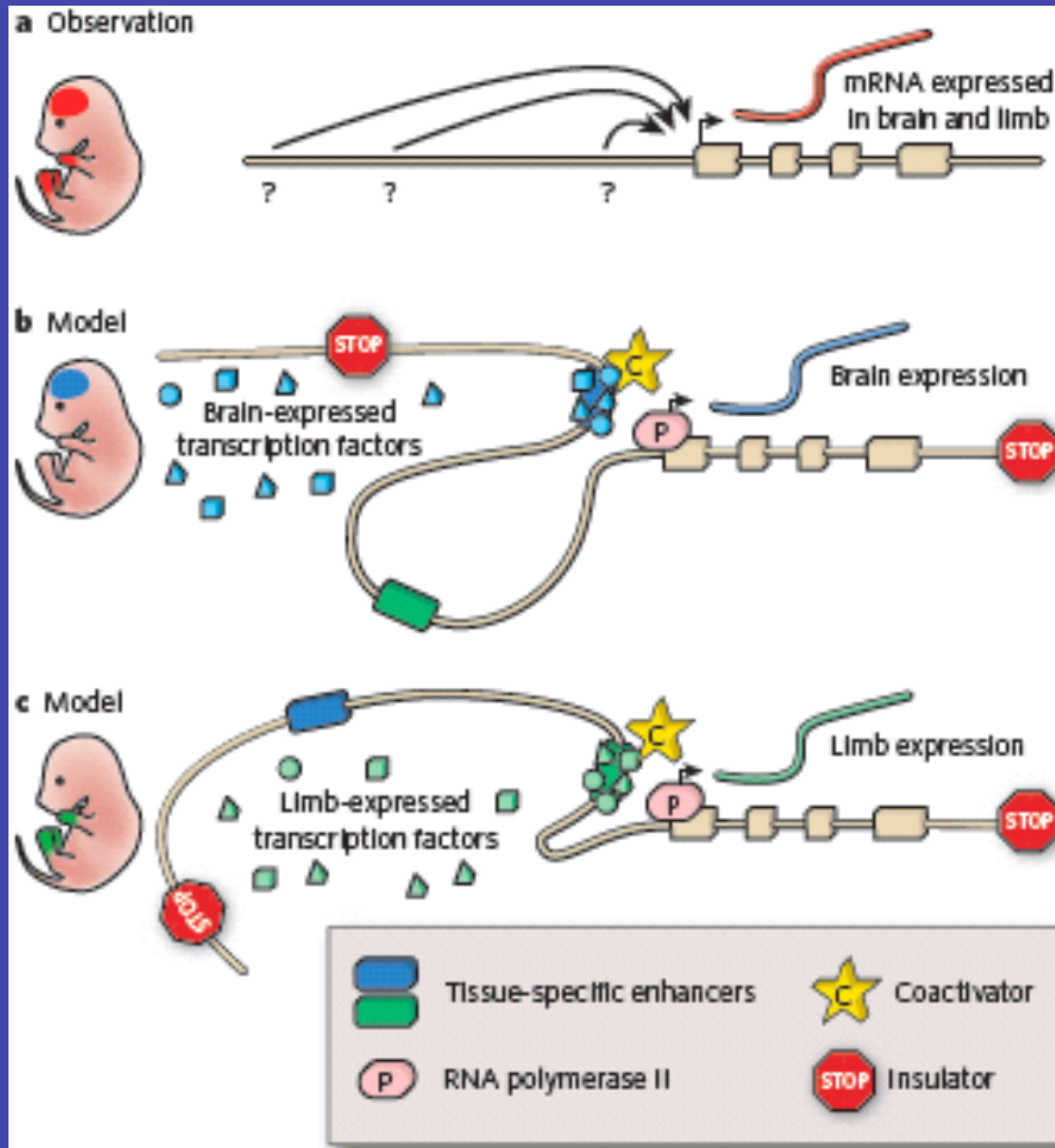
Esempio classico è quello mostrato in Figura 1

enhancers in human diseases

tre esempi:

- delez. e riarrang. degli enhancers della β globina HBB (hum beta globin)
- mutaz. dell'enhancer Sonic Hedgehog HGG limb (provoca polidattilia)
- suscettibilità alla malattia di Hirschsprung associata alla variante allelica dell' enhancer di RET (protoncogene)

models of enhancer long distance functioning



sito diverso
di attacco dei fattori
di trascrizione

- del SNC (blu)
- degli abbozzi degli arti (verde)

legend to figure 1

Overview of gene regulation by distant-acting enhancers.

l'azione regolativa di un promotore per molti geni non è sufficiente per un pattern di espressione complesso: espressione di un gene "G" in due tessuti diversi durante le fasi di sviluppo embrionale (ZRS o MFCS1 di SHH)

"Amano and colleagues report that expression of Sonic hedgehog (Shh) protein in the posterior mesenchyme of the mouse limb bud correlates with a long-range chromatin interaction with enhancer MFCS1 and looping of the Shh locus from its chromosome territory (CT)."

a) in rosso espressione di G negli abbozzi degli arti e nel SNC

b) modello: enhancer blu può legare una combinazione di alcuni fatt. di trascriz. se al completo l'enhancer si attiva (ricollocano l'enhancer fisicamente vicino al promotore (con un "loop"))

c) enhancer verde legato da una combinazione diversa di fattori di trascrizione specifici per il tessuto "c", solo quando si legano parte l'attività tessuto specifica di questo secondo enhancer

parte II della leggenda alla fig.1

come è possibile che due enhancers regolino lo stesso promotore ?

- a) tramite fattori di trascrizione tessuto specifici
- b) tramite “insulators” che limitano l’attività a domini cromosomici definiti
- c) oltre agli enhancer (regolazione positiva) ci sono repressori “repressors” e silenziatori “silencers” che partecipano alla regolazione trascrizionale (negativa).

distanza degli enhancers

nel caso di MFCS1 o ZRS sta 1 mega base da SHH dentro un introne di un gene limitrofo,

- inizialmente considerato rilevante per gli abbozzi degli arti e fu chiamato LMBR1 (limb region)

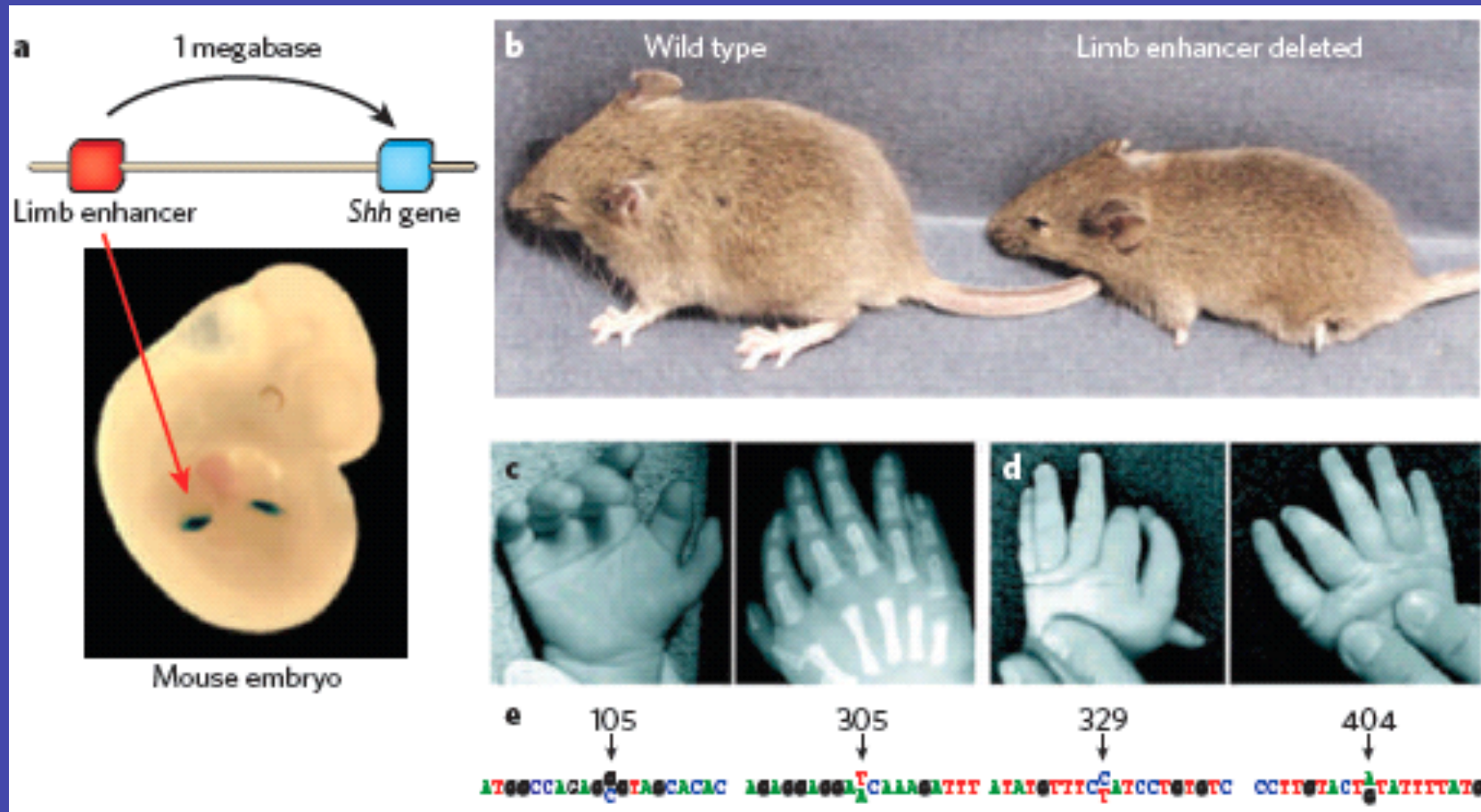
- studi successivi di topo e uomo: risequenziamento (single nucleotide variations) associazione con polidattilia preassiale, segrega in famiglie con anomalie degli arti

- topi transgenici per sostituzione di nucleotide singolo hanno come nei gatti un dito preassiale soprannumerario

- delezione dell'enhancer provoca arti tronchi

questi studi hanno mostrato per la prima volta sperimentalmente l'importanza di sequenze non codificanti che agiscono a grande distanza da cui studi sull'uomo

polidattilia o arti tronchi



Hirschsprung disease

altro esempio di mutazioni in regioni non codificanti

Although multigenic, Hirschsprung's disease risk is strongly linked to coding mutations in the *RET* proto-oncogene. However, family-based studies have also revealed evidence for Hirschsprung's disease linked to the *RET* locus in people lacking any accompanying functional *RET* coding mutations. Through the use of multispecies comparisons of orthologous genomic intervals that include and flank *RET*, coupled with *in vitro* and *in vivo* functional studies, an enhancer sequence located in intron 1 of *RET* was identified and found to contain a common variant contributing more than a 20-fold increased risk for Hirschsprung's disease than rarer alleles in this element. It is not sufficient since the disease is multifactorial and with complex aetiology.

enhancers in malattie rare o comuni ?

le patologie associate a varianti o polimorfismi di enhancers sono frequenti o rare ?

- a rapidly growing number of examples in which noncoding SNPs linked to disease traits through GWAS were found to affect the expression levels of nearby genes, suggesting that variation in regulatory sequences may commonly contribute to a wide range of disorders.

The results of the recent GWAS, coupled with the role of gene regulation in normal human biology, provide a strong incentive for defining the distant-acting-enhancer architecture of the human genome.

come trovare gli enhancers e regulatory regions RR

Gene centric studies gave limited results

Large scale studies through comparative genomics

- assumption: sequences of gene regulatory elements, like those of protein-coding genes, are under negative evolutionary selection, because most changes in functional sequences have deleterious consequences.

Statistical measures of evolutionary sequence constraint would provide a way to identify potential enhancer sequences within the vast amount of noncoding sequence in the human genome.

Retrospective studies on conservation of well defined orthologous enhancer encouraged blind studies of regions flanking genes of interest towards focusing specifically on non-coding sequences constrained across vertebrate species, culminating in whole-genome studies in which conservation level alone guided experimentation.

strategie ed osservazioni

Large-scale transgenic mouse and fish studies evidenced many of these non-coding sequences that had been conserved for hundreds of millions of years of evolution and were enhancers that drove expression in highly specific anatomical structures during embryonic development.

Blocks of 200 base pairs or more that are perfectly conserved between humans, mice and rats, were also found to be highly enriched for tissue-specific enhancers, suggesting that the success rate of comparative approaches for enhancer identification depends on scoring criteria, rather than just evolutionary distance.

studi su topi transgenici in larga scala

- nuovi approcci statistici per studiare le regioni evolutivamente conservate per predire enhancers
- informazioni da confronti tra specie anche vicine
- studio su topi transgenici comprendente anche regioni non codificanti
- evidenze uomo-topo che molti enhancers possono essere altamente conservati mentre altri lo sono di meno
- la distribuzione non è random, è associata in maniera “bias” a regioni attive durante lo sviluppo, conferme anche in vivo su saggi che mostrano attività nello sviluppo e tessuto specificità Figura 2.

ci sono limiti in questi studi?

le analisi throughput studiano le compatibilità con le ipotesi imposte di una serie di dati analizzati anche in corso di analisi

- la conservazione di seq non codificanti non necessariamente indica che si tratti di enhancers
- ci sono molti altri tipi di elementi non codificanti funzionali
- la conservazione di una regione con funzione enhancer non può predire luogo e tempo di espressione nello sviluppo o nell'animale adulto
- per tutti gli enhancers candidati vanno decifrate le proprietà regolatorie di ogni elemento di quali geni
- fattibilità di studi a larga scala che comprendano i sets di dati ottenuti su analisi di GW

fenotipi transgenici

ad un topo transgenico non sempre corrisponde 1 fenotipo

- Ahituv, N. et al. Deletion of ultraconserved elements yields viable mice. PLoS Biol. 5, e234 (2007). This paper shows that deletion of several ultraconserved non-coding sequences in mice may not result in obvious phenotypes, demonstrating that even extreme evolutionary constraint does not necessarily indicate that a non-coding sequence is required for viability.
- The ENCODE Project Consortium. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. Nature 447, 799–816 (2007).
- Heintzman, N. D. et al. Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. Nature Genet. 39, 311–318 (2007). This paper identifies a histone H3K4 differential methylation signature that distinguishes promoters from enhancers, providing a chromatin-based tool for genome-wide enhancer prediction.
- Wei, C. L. et al. A global map of p53 transcription-factor binding sites in the human genome. Cell 124, 207–219 (2006). Describes mapping of protein–DNA interactions by ChIP coupled with conventional capillary-based sequencing of concatenated paired-end tags (ChIP-PET), a conceptual predecessor of the ChIP-seq approach.

la forza delle interazioni

la plasticità dei sistemi biologici - sistemi rigidi e sistemi flessibili

interpretazione ed ipotesi esplicativa:

- assenza di fenotipo per la capacità di altri sistemi a supplire per le funzioni mancanti generate dalla transgenia di quel gene o enhancers in questi casi specifici

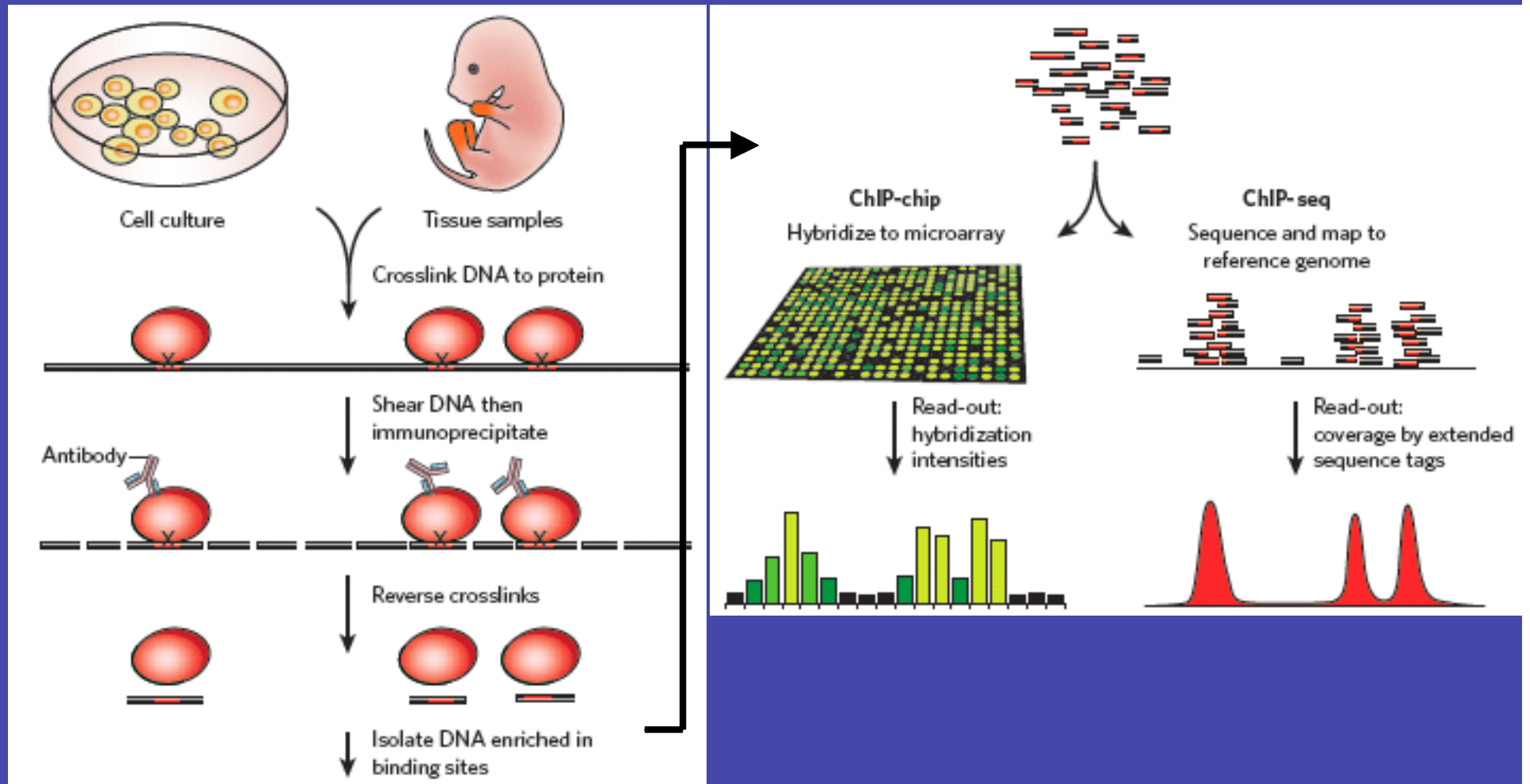
- possibile spiegazione con la ridondanza e magari una riduzione di funzione leggera che porta in molte generazioni ad una diminuzione di fitness in un ambiente reale non di laboratorio.

- conclusioni: nonostante l'alta conservazione evolutiva di una regione NC predice una regione regolativa, la costrizione evolutiva può non essere correlata con la predizione di un fenotipo gravemente alterato.

nuovi metodi alternativi ai confronti genomici

- New technologies were developed that allow traditional chromatin immunoprecipitation (ChIP) techniques to be applied on the scale of whole vertebrate genomes. The initial in-depth studies of 1% of the genome in the Encyclopedia of DNA Elements (ENCODE) pilot projects were largely based on data sets generated by the ChIP-chip technique (Box 1) and revealed the molecular properties of a variety of regulatory elements.
- With respect to enhancer identification, a particularly relevant insight was the identification of specific histone methylation signatures found at enhancers. In contrast to promoters, which are marked by trimethylation of histone H3 at lysine residue 4 (H3K4me3), active enhancers are marked by monomethylation at this position (H3K4me1)⁴⁰. Mapping these marks in the ENCODE regions and, more recently, throughout the entire genome⁴¹ revealed tens of thousands of elements that were predicted to be active enhancers
- predicted enhancers were frequently associated with the transcriptional coactivators p300 and/or TRAP220 (also known as MED1), such coactivators might be useful general markers for mapping enhancers.

ChIP on chip



a number of chromatin marks and transcription factors both in humans and mice have become available. These data sets allowed the identification of not only the H3K4me1 and H3K4me3 signatures but also additional chromatin marks present at predicted or validated enhancers, and provided a refined view of their correlation to enhancer activities^{44,51,55}.

categorie di seq funzionali non cod identificate

Table 1 | Selected major categories of non-coding functional element

Category	Function	Selected associated chromatin marks*
Promoter	Region that is located immediately upstream of a protein-coding gene, and binds to RNA polymerase II; where transcription is initiated	RNA polymerase II ⁴⁴ , H3K4me3 (ref. 40) (active promoters)
Enhancer	Region that activates transcription, often in a temporally and spatially restricted manner, by acting on a promoter. Enhancers can be located far from target promoters and are orientation independent	p300 (refs 40, 56), H3K4me1 (ref. 40)
Insulator	Separates active from inactive chromatin domains and interferes with enhancer activity when placed between an enhancer and promoter	CTCF ^{44,53}
Repressor/silencer	Negative regulators of gene expression	REST ⁴⁵ , SUZ12 (refs 69, 70)

*Many additional chromatin marks were found to correlate with one or several of these categories of regulatory element. Detailed descriptions of these markers and their respective binding characteristics at different types of regulatory sequence element can be found in refs 40, 41, 44, 51 and 55.