

# Lezione XXII-XXIII

martedì 24-XI-2011



corso di genomica

Laurea Magistrale Biotecnologia Industriale

aula 8

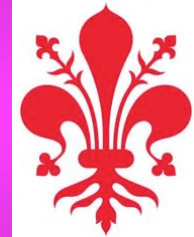
orario : Martedì ore 14.00 - 16.00

Giovedì ore 13.00 - 15.00

libro consigliato Zanichelli “Introduz. alla genomica” di  
Arthur M.Lesk

D. Frezza

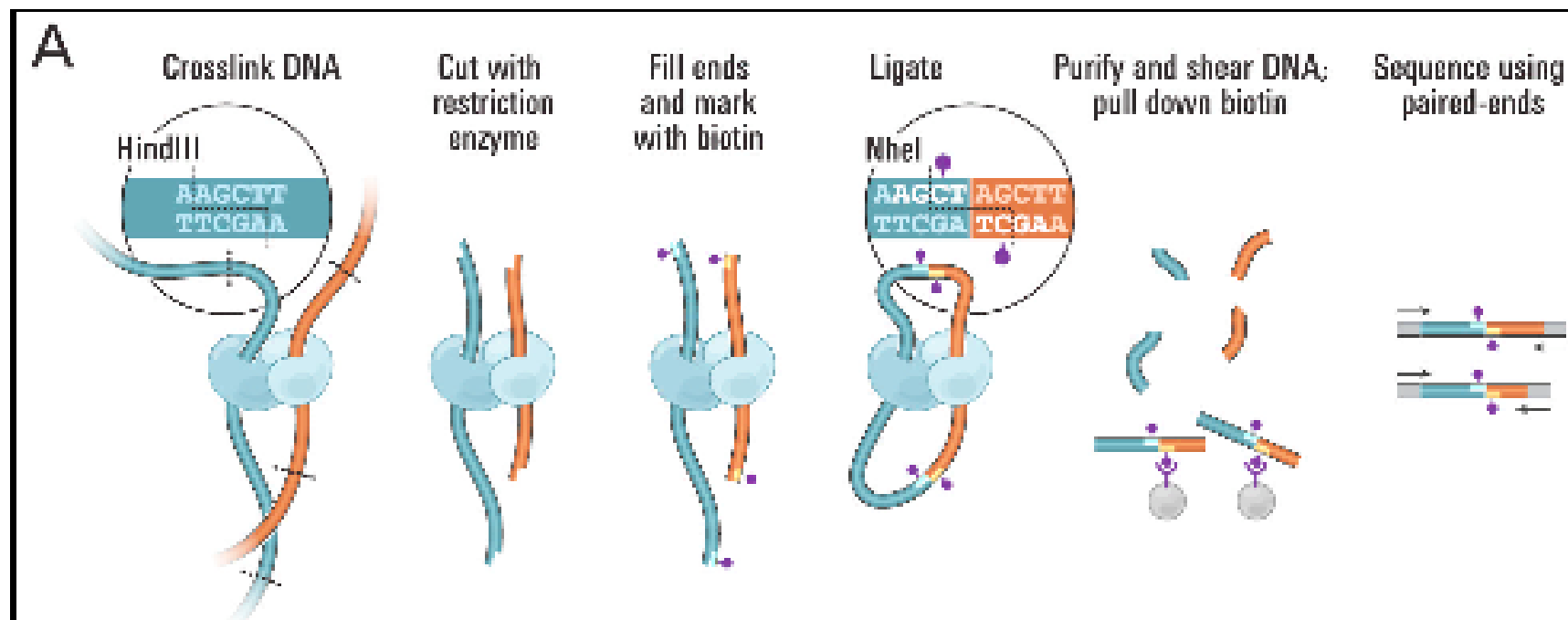
# Interazioni genomiche: metodo HiC



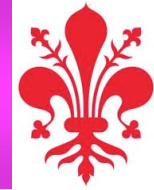
Comprehensive mapping of long range interactions reveals folding principles of the human genome.

Science 9 Ottobre 2009, 326 (5950) pp 289-293. Erez Lieberman-Aiden et al.

## The Hi-C method



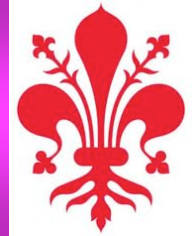
# Abstract Hi-c science Lieberman-Aiden



We describe Hi-C, a method that probes the three-dimensional architecture of whole genomes by coupling proximity-based ligation with massively parallel sequencing. We constructed spatial proximity maps of the human genome with Hi-C at a resolution of 1Mb. These maps confirm the presence of chromosome territories and the spatial proximity of small, gene rich chromosomes.

We identified an additional level of genome organization that is characterized by the spatial segregation of open and closed chromatin to form two genome-wide compartments. At the megabase scale, the chromatin conformation is consistent with a fractal globule, a knot-free conformation that enables maximally dense packing while preserving the ability to easily fold and unfold any genomic locus. The fractal globule is distinct from the more commonly used globular equilibrium model. Our results demonstrate the power of Hi-C to map the dynamic conformations of whole genomes.

# Procedura Hi-C I

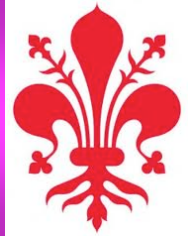


Long-range interactions between specific pairs of loci can be evaluated with Chromosome Conformation Capture (3C), using spatially constrained ligation followed by locus-specific PCR (6). Adaptations of 3C have extended the process with the use of inverse PCR (4C) (7,8) or multiplexed ligation-mediated amplification (5C) (9). Still, these techniques require choosing a set of target loci and do not allow unbiased genome-wide analysis.

Here we report a method named Hi-C that adapts the above approach to enable purification of ligation products followed by massively parallel sequencing. Hi-C allows unbiased identification of chromatin interactions across an entire genome.

La bibliografia sta in fondo all'articolo

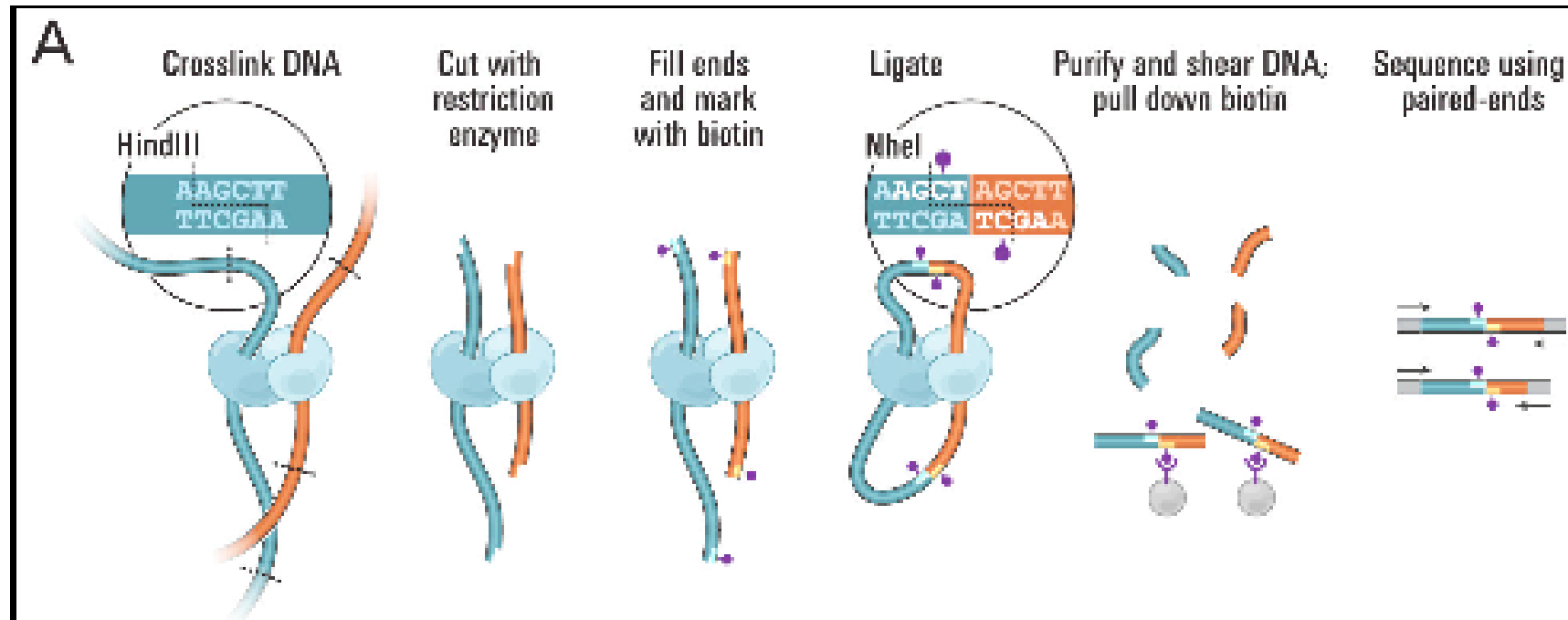
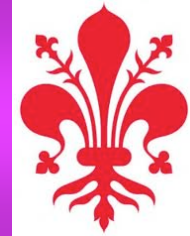
# Procedura Hi-C II



cells are crosslinked with formaldehyde; DNA is digested with a restriction enzyme that leaves a 5'- overhang; the 5'- end is filled, including a biotinylated residue; the resulting blunt-end fragments are ligated under dilute conditions that favor ligation events between the cross-linked DNA fragments. The resulting DNA sample contains ligation products consisting of fragments that were originally in close spatial proximity in the nucleus, marked with biotin at the junction. A Hi-C library is created by shearing the DNA and selecting the biotin- containing fragments with streptavidin beads. The library is then analyzed using massively parallel DNA sequencing, producing a catalog of interacting fragments (Fig. 1A, SOM).

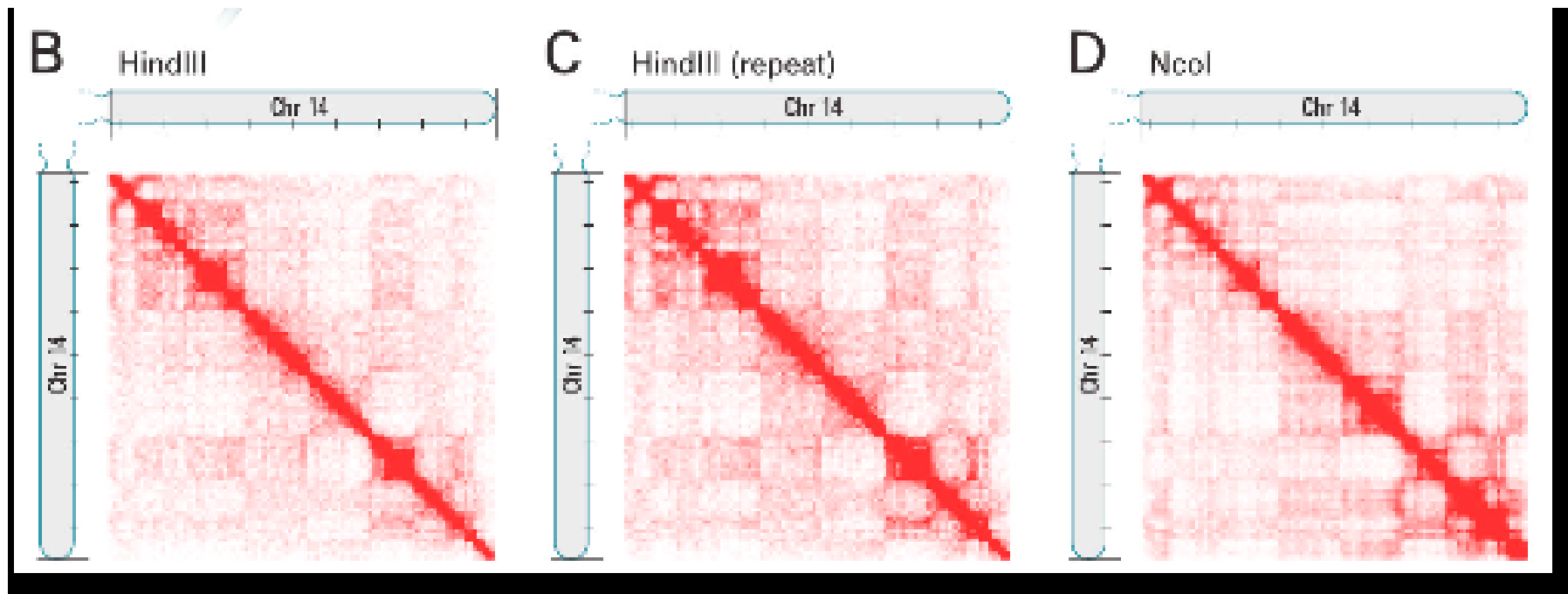
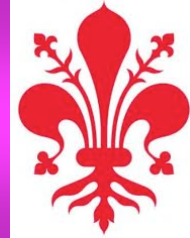
The library was created from a normal lymphoblastoid cell line and sequenced on a two lanes of Illumina Genome Analyzer, generating 8.4 millions of reads pairs uniquely aligned to human genome ref. seq. of these 6.7 millions corresponded to long range contacts between segments greater than >20 kb apart.

# Fig 1 procedura



Cross link with formaldehyde resulting in covalent links between spatially adjacent chromatin segments mediated by proteins interactions. Chromatin digestion, fill up of sticky ends with one fluorescent nucleotide (purple dot) Ligation (in diluted condition to avoid non linked ligations) and HindIII transformed in NheI. DNA purified and sheared, biotinylated junctions isolated by streptavidin beads and identified by paired end sequencing.

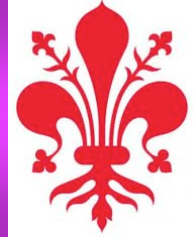
# Mapping chromosome 14 interactions



Hi-C produces a genome-wide contact matrix. The submatrix shown here corresponds to intrachromosomal interactions on chromosome 14. Each pixel represents all interactions between a 1Mb locus and another 1Mb locus; intensity corresponds to the total number of reads (0-50). Tick marks appear every 10Mb.

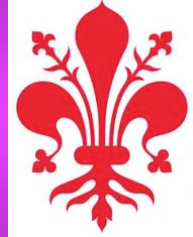
C, D We compared the original experiment to a biological repeat using the same restriction enzyme (C, range: 0-50 reads) and to results with a different restriction enzyme (D, range: 0- 100 reads, NcoI).

# Primo risultato



una libray Hi-C di una linea umana con cariotipo normale B linfoblastoide immortalizzata con EBV è stata sequenziata con uno strumento a due corsie Illumina Genome Analyzer ottenendo  $8.4 \times 10^6$  di paia di “reads” (sequenze utili) con allineamento unico sul genoma umano di riferimento; di queste  $6.7 \times 10^6$  corrispondevano a contatti tra segmenti separati a distanza  $>20$  Kb.  
Conclusione: il metodo funziona

# secondo esperimento



Hanno costruito una matrice di contatti dell'intero genoma  $M$  dividendo il genoma in regioni da 1 Mb (loci) e definendo l'entry della matrice  $m_{ij}$  come il numero di prodotti ligati tra locus  $i$  e  $j$  (SOM) (=rappresentatività). Questa matrice riflette una media d'insieme delle interazioni presenti nel campione originale di cellule. Si visualizza come heat-map con l'intensità che indica la frequenza dei contatti. Fig 1B

## analisi di ripetibilità con altro enzima di restrizione



Esperimento ripetuto digerendo i frammenti con NcoI, la library era estremamente simile a quella con HindIII vista col test Pearson's  $r=0,990$  e con NcoI  $r=0,814$ . Il p trascurabile  $<10^{-300}$  in entrambe i casi.

## analisi di probabilità di contatto

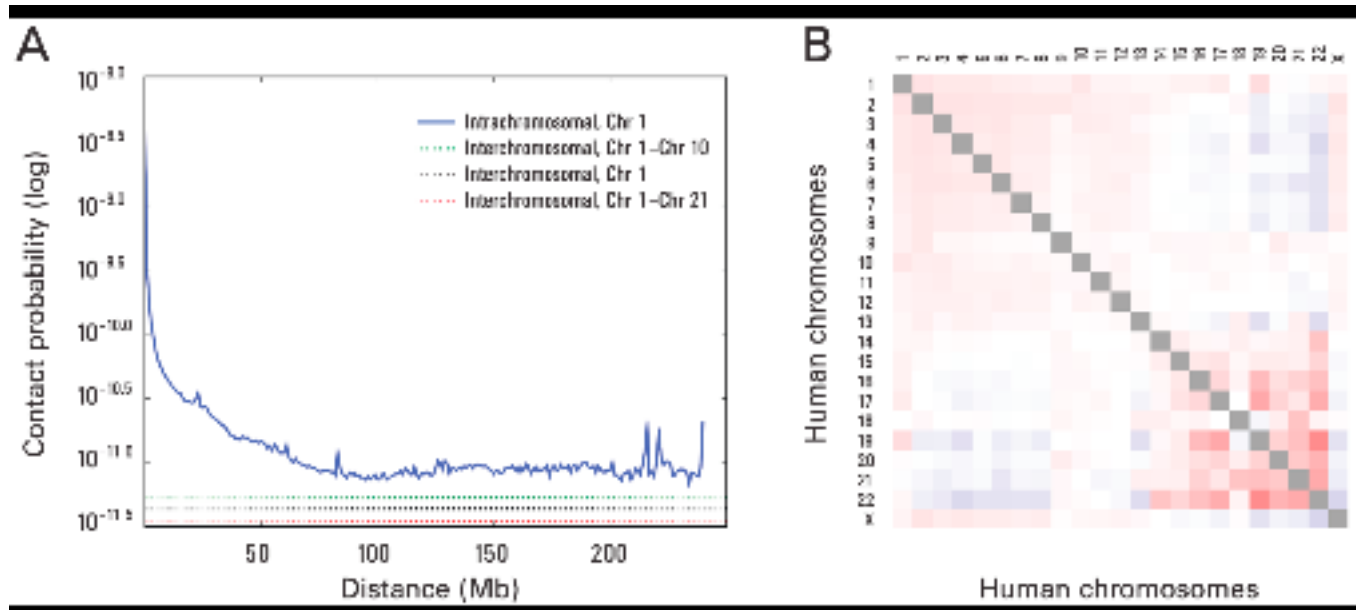


La probabilità di contatto tra una coppia di loci  $ln(s)$  separati dalla distanza genomica  $s$  (in bp) sul cromosoma  $n$ .

$ln(s)$  decresce in maniera monotonica in ogni cromosoma suggerendo un comportamento di un polimero in cui la distanza tridimensionale tra i loci cresce con la distanza genomica.

Queste evidenze sono in accordo con i risultati con la 3C e fluorescence in situ hybridization FISH. A distanze maggiori di 200 Mb  $ln(s)$  è molto maggiore della probabilità di contatto medio tra differenti cromosomi (Fig 2A) e ciò implica l'esistenza di territori cromosomici.

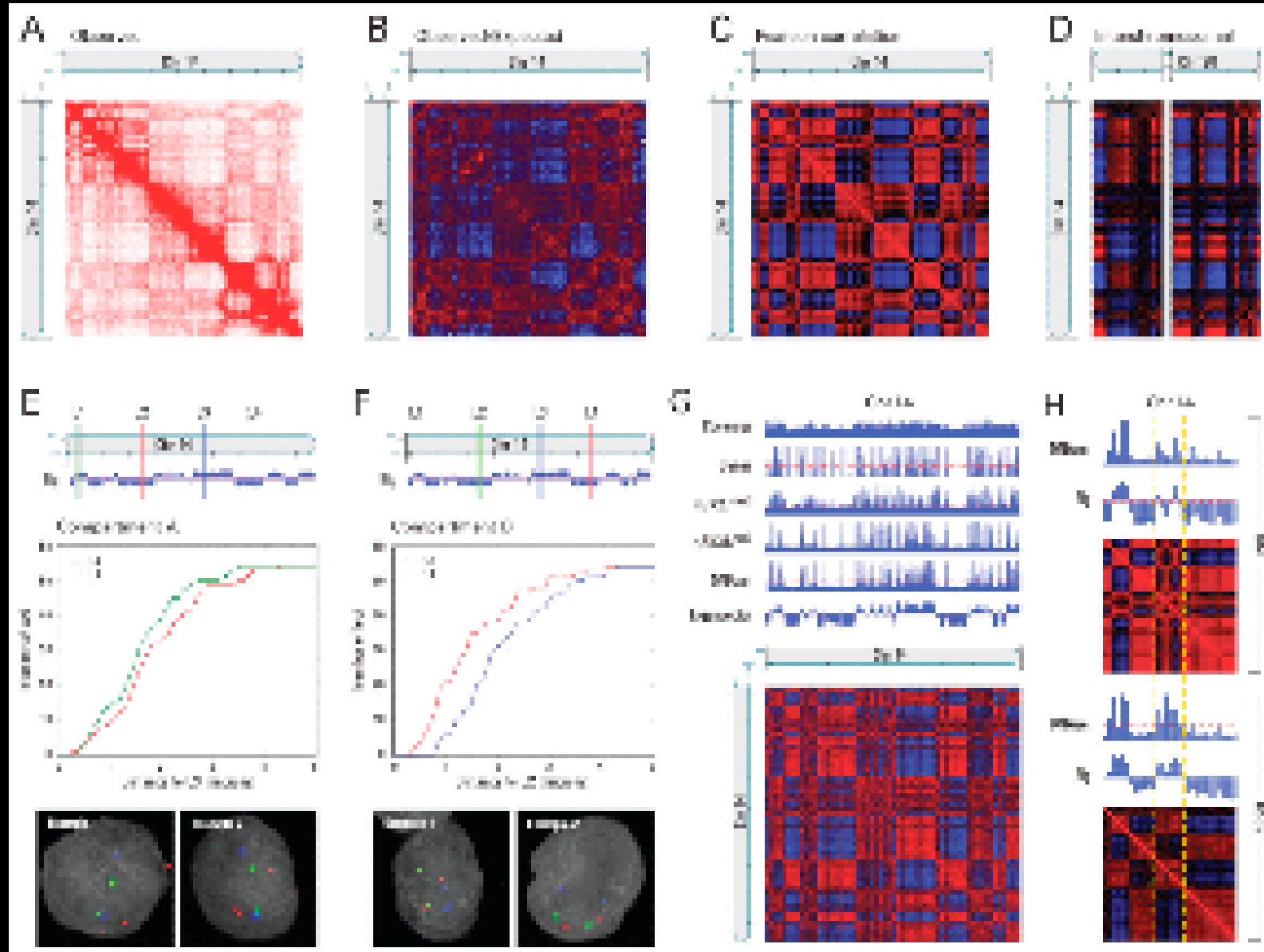
## fig 2 chromosome territories organization



observed / expected

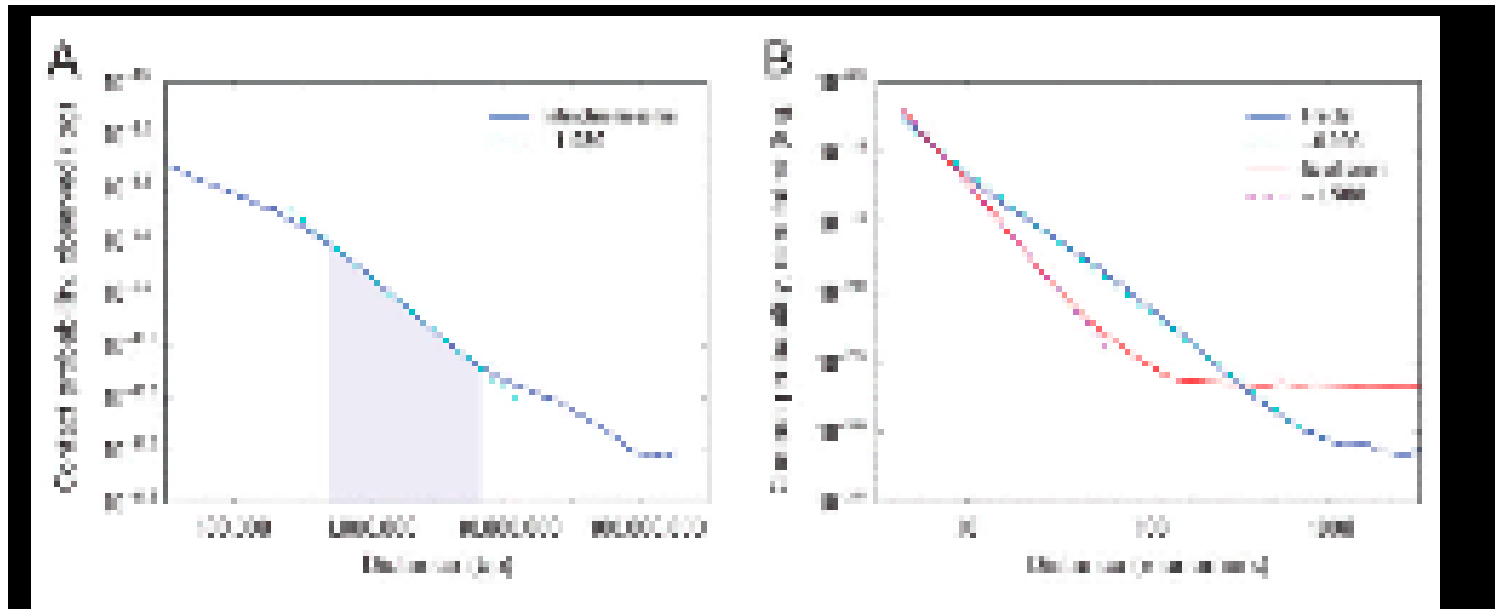
The presence and organization of chromosome territories. **(A)** Probability of contact decreases as a function of genomic distance on chromosome 1, eventually reaching a plateau at ~90Mb (blue). The level of interchromosomal contact (black dashes) differs for different pairs of chromosomes; loci on chromosome 1 are most likely to interact with loci on chromosome 10 (green dashes) and least likely to interact with loci on chromosome 21 (red dashes). Interchromosomal interactions are depleted relative to intrachromosomal interactions. **(B)** Observed/expected number of interchromosomal contacts between all pairs of chromosomes. Red indicates enrichment, and blue indicates depletion (up to twofold). Small, gene-rich chromosomes tend to interact more with one another.

Figura 3B



# Fig 4 AB

Fino a 200 “reads” letture di sequenza



Purtroppo la qualità è scarsa, siete pregati di guardare la figura direttamente sul pdf dell'articolo

# figura 4 modello

