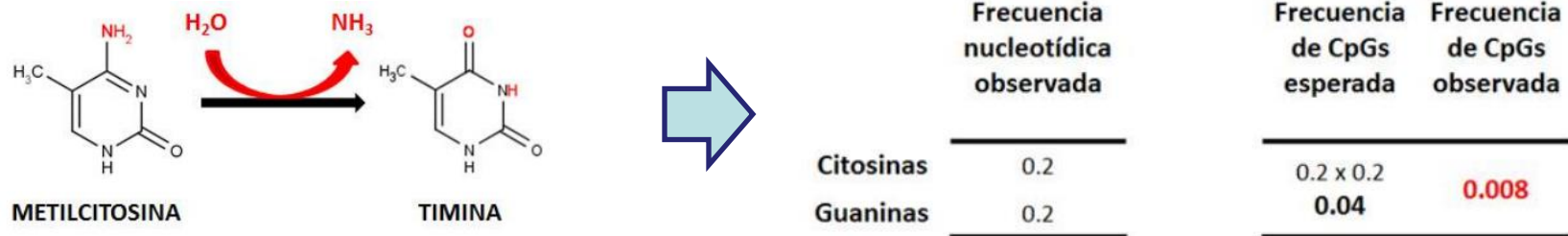


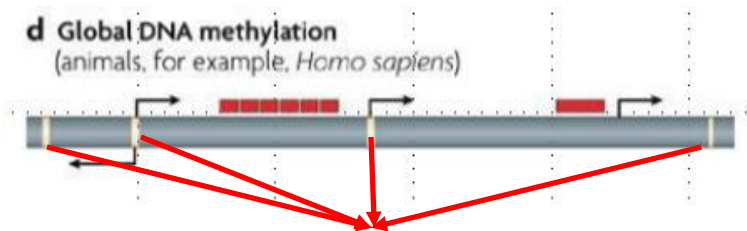
Islas CpG

¿Qué son las islas CpG?

Son la manifestación en la secuencia del genoma de los patrones de metilación



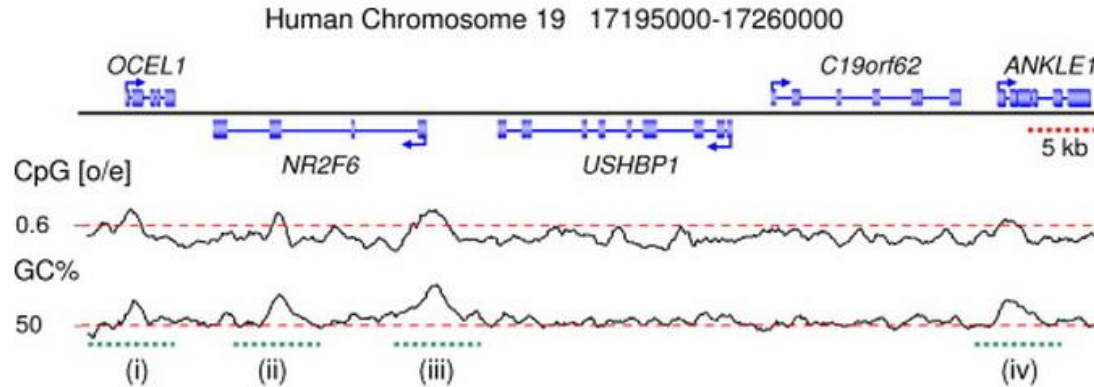
La desaminación espontánea de m5C (80% de CpGs en el genoma humano) provoca la mutación espontánea de contextos *CG* a *TG*
Generalmente, los CpGs se agrupan en regiones no metiladas



ISLAS CpG

CpG Islands in vertebrate genomes ☆

Propiedades de las islas CpG

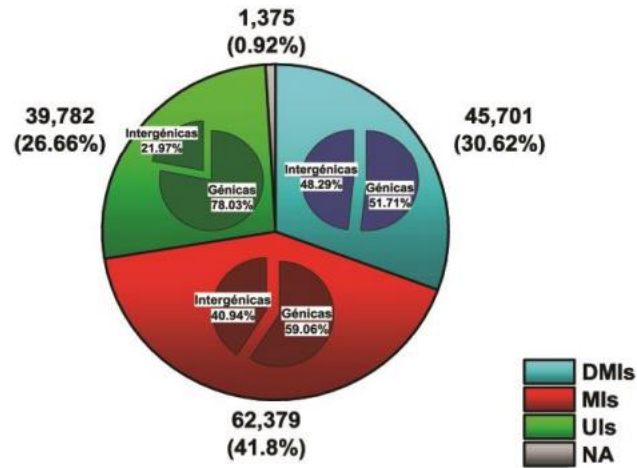


PROPIEDADES

- Ricas en G+C y CpGs con longitudes superiores a 200 nts
- Elevados ratios O/E (número de CpGs observado/esperado)
- Frecuentemente, localizadas en las regiones promotoras de los genes
- Alrededor del 70% de los genes presentan una isla CpG (Todos los 'housekeeping' y la mitad de los genes tejido-específicos)

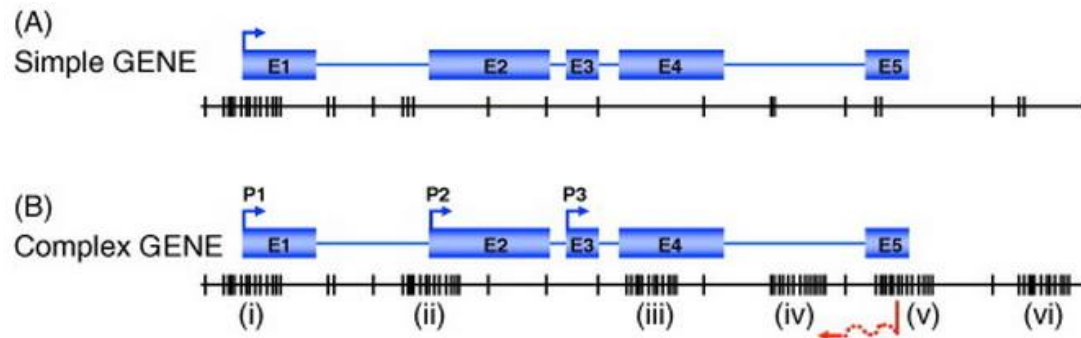
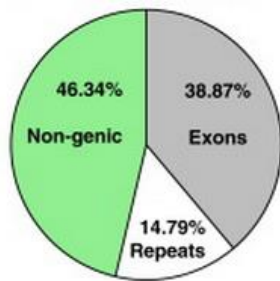
Anexos a la definición clásica

Estado de metilación



DMIs: Islas Diferencialmente metiladas
MIs: Islas Metiladas
UIs: Islas No Metiladas

Localización



Minireview

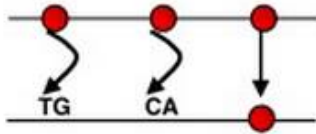
CpG islands – ‘A rough guide’

Edited by Miguel De la Rosa

Robert S. Illingworth , Adrian P. Bird 

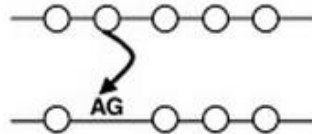
Dinámica evolutiva de las islas CpG

A 5mCpG-Genomic background



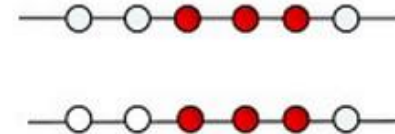
Rapid deamination
Low CpG content

B Bird's hypodeaminated regime



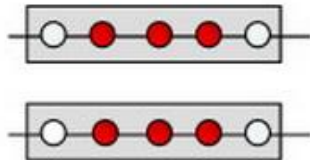
Low deamination
High CpG content

C CpG selection



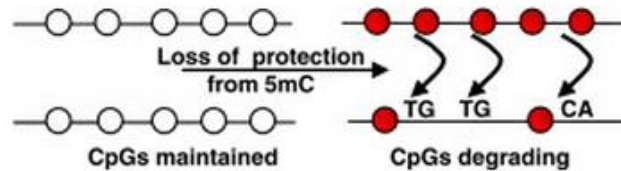
Low CpG loss rate
High CpG content

D Exon or general selection



Selection on non CpG + CpGs
High G+C, CpG content

E CpG decay



Recent increase in deamination rate
CpG content is decreasing

Theory

Primate CpG Islands Are Maintained by Heterogeneous Evolutionary Regimes Involving Minimal Selection

Densidad de CpGs y Metilación

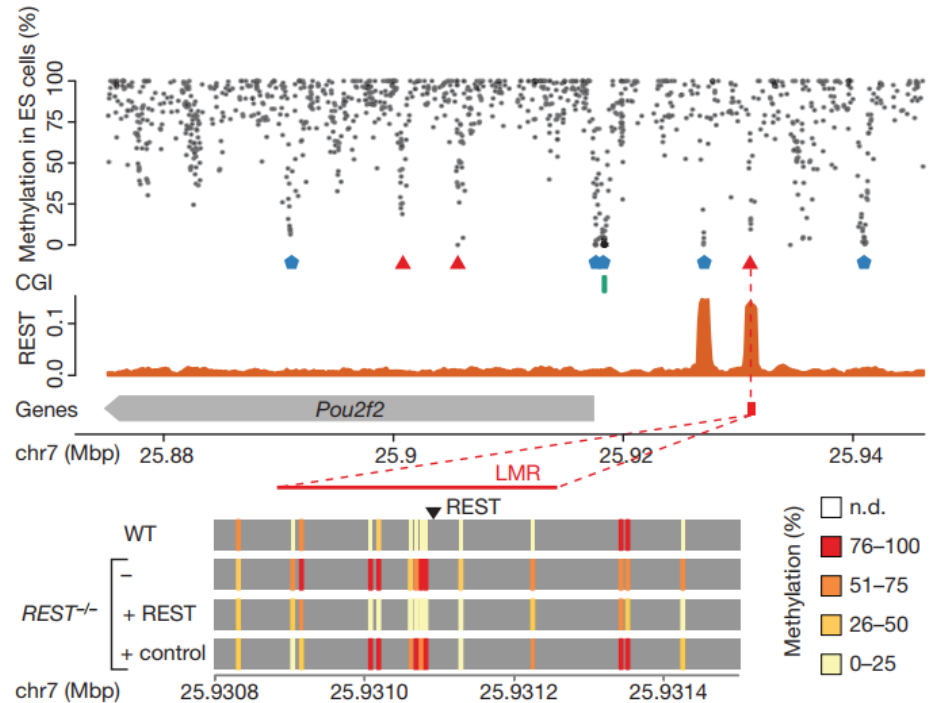
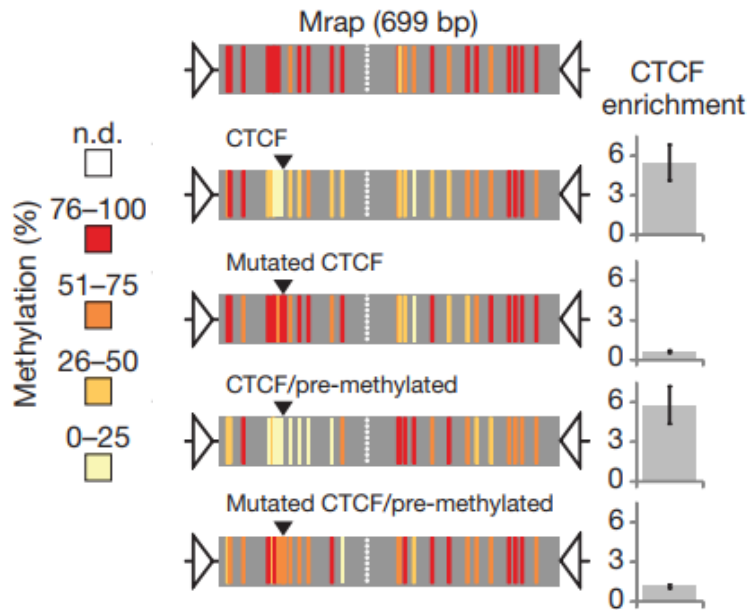


Figure 4 | REST is required for LMR formation at its binding sites.

En regiones con baja densidad de CpGs, los factores de unión al DNA y sus sitios de unión son necesarios y suficientes para generar regiones con bajos niveles de metilación.

DNA-binding factors shape the mouse methylome at distal regulatory regions

Michael B. Stadler, Rabih Murr, Lukas Burger, Robert Ivanek, Florian Lienert, Anne Schöler, Erik van Nimwegen, Christiane Wirbelauer, Edward J. Oakeley, Dimos Gaidatzis, Vijay K. Tiwari & Dirk Schübeler

Affiliations | Contributions | Corresponding author

Nature (2011) | doi:10.1038/nature10716

Received 28 June 2011 | Accepted 15 November 2011 | Published online 14 December 2011

Métodos computacionales de predicción de islas CpG

- **Métodos de ventana** (*Takai & Jones* y *CpGProD*)

Basados en umbrales de características composicionales dentro de ventanas móviles en la secuencia

- **Métodos de adición** (*CpGreporty UCSC*)

Basados en la identificación de regiones con una mayor proporción de CpGs que las regiones colindantes

- **Métodos de agrupación** (*CpGcluster*, *cgClusters* y *CGI_HMM*)

Basados en criterios estadísticos y composicionales de la secuencia

Métodos de ventana (Sliding Windows)

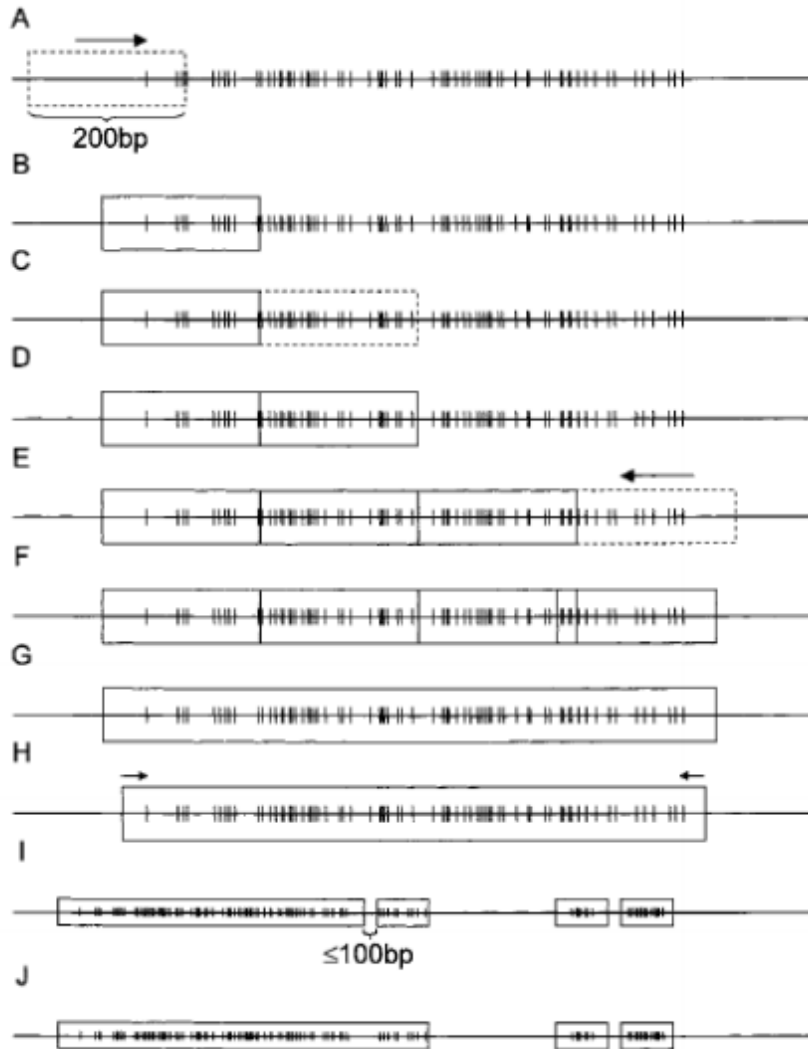


Table 1.
Overview of CpG island prediction algorithms.

| Database/prediction | Length | G + C | CpG[o/e] | RM ^a | Comments |
|---------------------|-----------------|-------|----------|-----------------|-----------------------------|
| ENSEMBL | ≥400 | ≥50% | ≥0.6 | N | Stringent length constraint |
| NCBI relaxed | ≥200 | ≥50% | ≥0.6 | N | Total CGIs = 307 193 |
| NCBI strict | ≥500 | ≥50% | ≥0.6 | N | Total CGIs = 24 163 |
| USCS ^b | >200 | ≥50% | >0.6 | Y | Total CGIs = 28 226 |
| EMBOSS | UD ^c | UD | UD | NA | Variable parameters |
| CpGProD | >500 | >50% | >0.6 | Y | Total CGIs = 76 793 |
| CpGcluster | NA | NA | NA | N | Clustering Total = 197 727 |

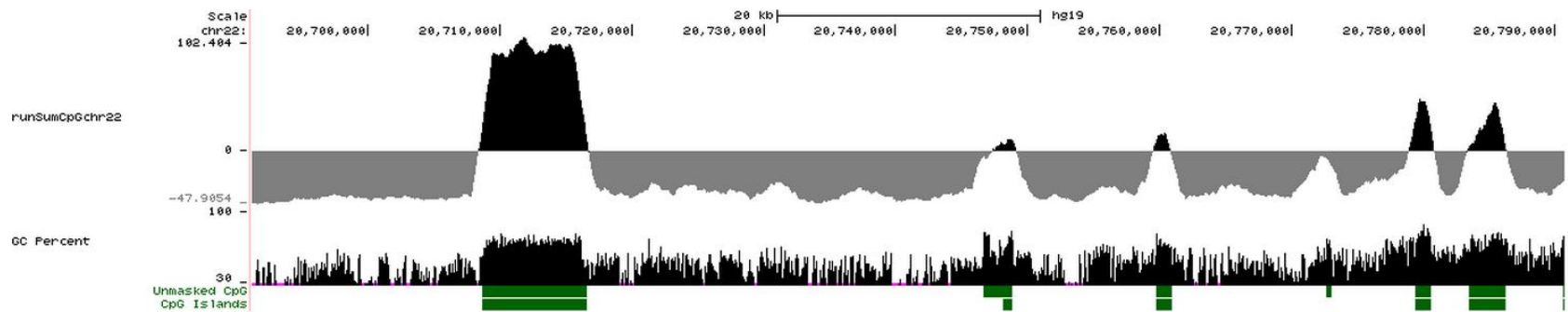
Dependencia de los umbrales seleccionados
Parámetros especie específicos
Matemáticamente incompletos

Comprehensive analysis of CpG islands in human chromosomes 21 and 22

Métodos de adición (Running Sum)

MÉTODO

1. Calcula score en torno a cada posición (+17 CpG / -1 noCpG)
2. Identifica regiones con scores máximos
3. Ventanas con scores máximos aplica umbrales de longitud, GC y ratio O/E (sólo UCSC)



**Dependencia de los umbrales seleccionados
Islas CpG en función de composición local**

Métodos de agrupación (Clustering)

Read the DNA sequence

Binary sequence: CpG = 1; noCpG = 0

0001000010100000010100010100001000101010000101

Determine the distance (d) of each CpG to the next CpG downstream in the DNA sequence:

10,5,5,3,1,8,23,34,21,12,2,5,8,6,9,...N-1

For example, for $d_m = 5$:

10, **5,5,3,1**, 8, 23, 34, 21, 12, **2,5**, 8, 6, 9, ...N-1

Statistically significant cluster \equiv CpG island

Research article

Highly accessed

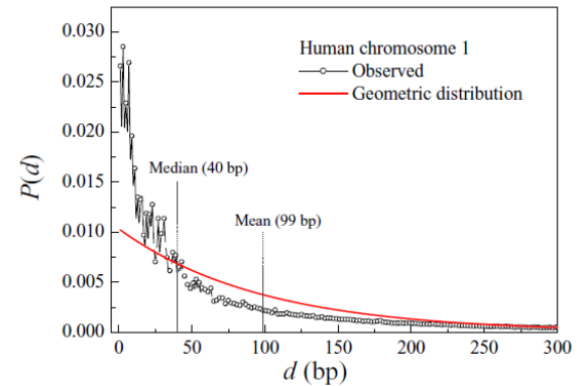
Open Access

CpGcluster: a distance-based algorithm for CpG-island detection

Michael Hackenberg¹, Christopher Previti^{1,5}, Pedro Luis Luque-Escamilla², Pedro Carpena³, José Martínez-Aroza⁴ and José L. Oliver^{1*}

¿Qué distancia?

$$P(d) = (1-p)^{d-1}p$$



¿Significación estadística?

$$P_{N,p}^{cum}(x \leq n_f) = \sum_{x=0}^{n_f} \binom{x-(N+1)-1}{(N-1)-1} \cdot p^{N-1} \cdot (1-p)^x$$

