Regulatory Elements

part of “Genomik der Genregulation”

Sonja Prohaska

Computational EvoDevo
University Leipzig

Leipzig, WS 2011/12

Organization of Transcription Factor Binding Sites

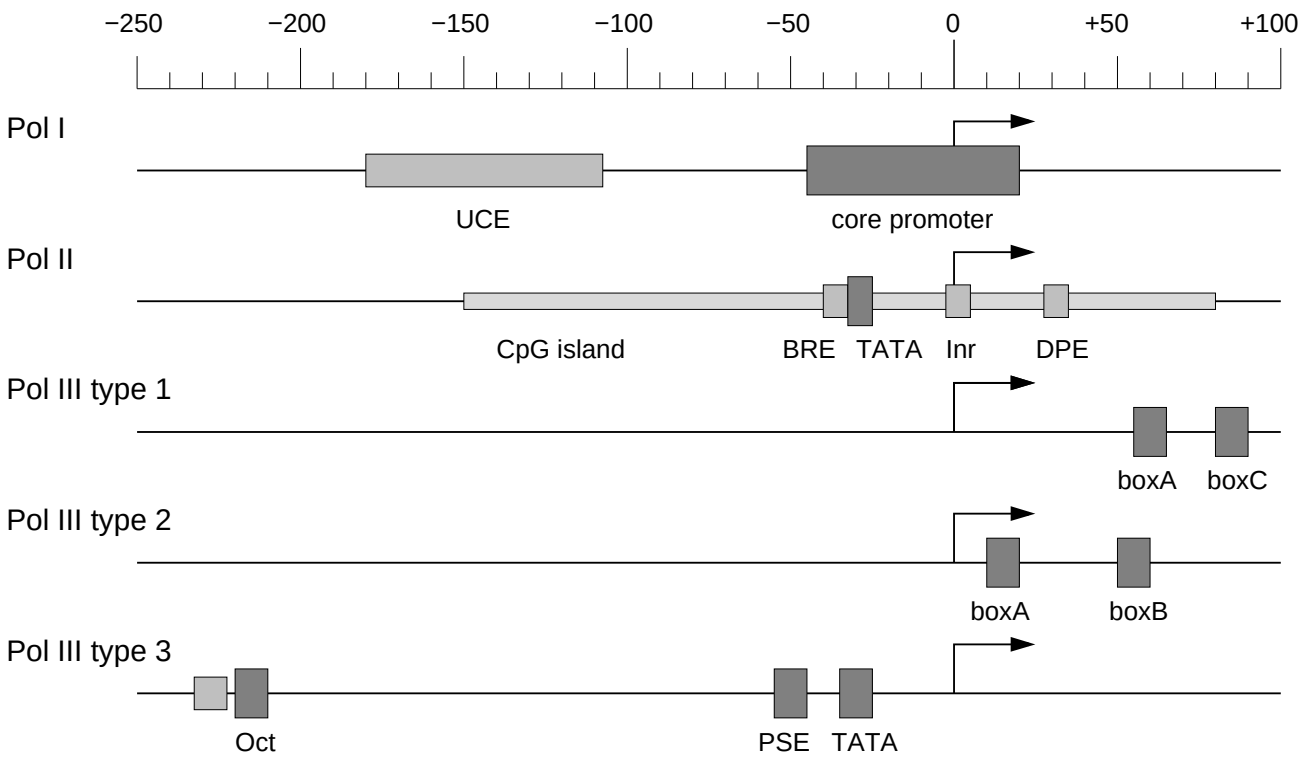
TFs are organized into *cis*-regulatory elements

- promoter regions: directly upstream and/or downstream of the transcription start site (TSS)
- proximal promoter: within 500nt upstream of the TSS
- enhancer elements: “anywhere” in the genome

Promoters

RNA polymerase	Promoter	Location relative to start site	Transcript	Function
Pol I	core element UCE (upstream control element)	-45 to +20 -180 to -107	pre-rRNA (28S, 18S, 5.8S)	components of the ribosome; translation
Pol II	TATA-Box Initiator CpG islands no	-25 to -35 -100	mRNA snRNA (U1-4) LINEs	protein coding genes components of the spliceosome; mRNA splicing Retrotransposon
Pol III	type 1: A-box, C-box type 2: A-box, B-box type 3: TATA-Box	+50 to +80 +10 to +60 -30 to -70	5S rRNA tRNA snRNA (U6) 7SL RNA	component of large ribosomal subunit translation components of the spliceosome; mRNA splicing component of the SRP (signal recognition particle); protein transport to ER (endoplasmatic reticulum)

Promoters

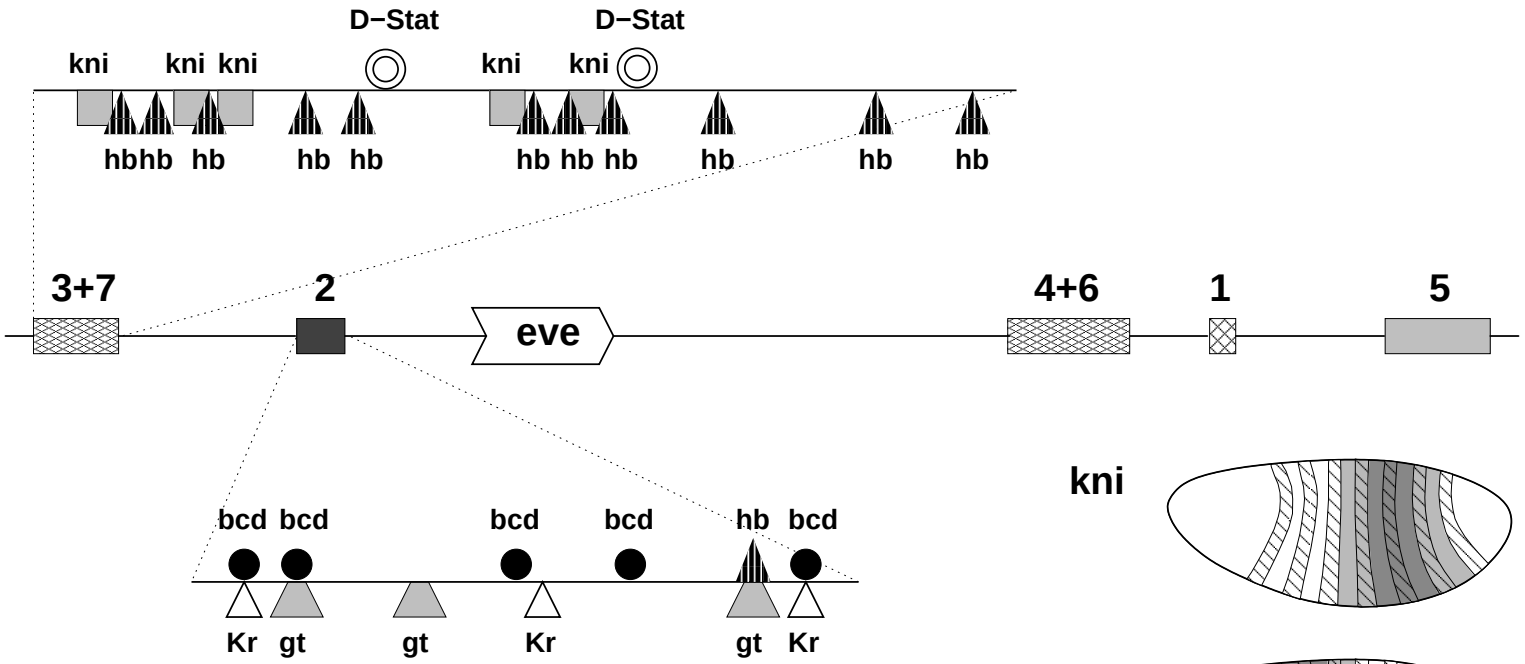


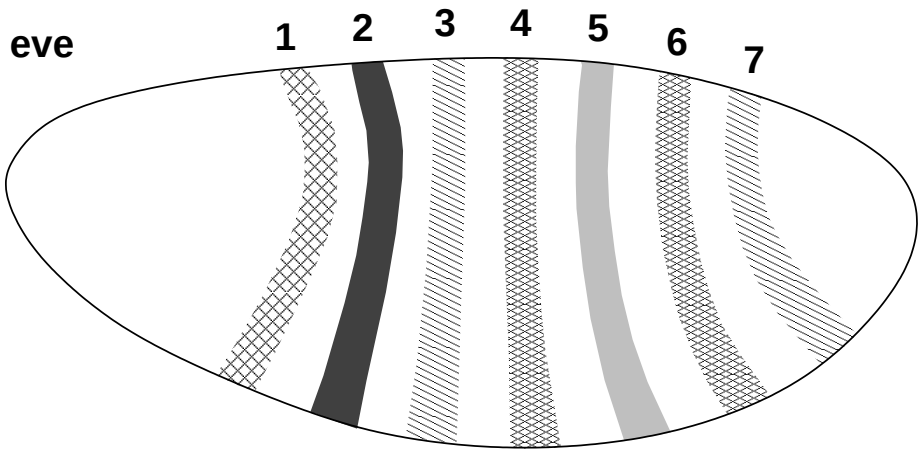
Core motifs of the different promoter types. Motifs in dark gray are less dispensable than motifs in light gray. Any specific promoter may contain just a subset or, in the worst case, none of these motifs. UCE = upstream control element, BRE = TFIIB recognition element, Inr = initiator element, DPE = downstream core promoter element, Oct = octamer binding site, PSE = proximal sequence element. The arrow indicates the transcription start site at +1.

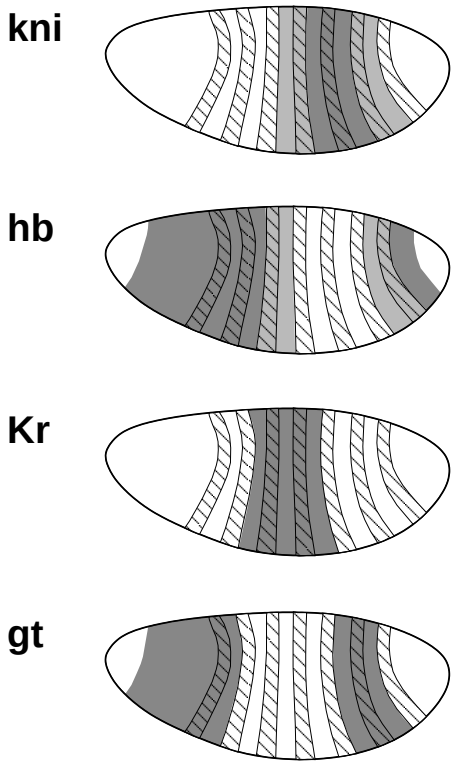
Regulatory Elements
cis-regulatory sequences - enhancer elements

- carry out a regulatory function
  that is a subfunction of a complex regulatory pattern
- execute this subfunction when linked to a reporter gene
- 100-300bp long
- may be separated from the target gene by several 100kb
- contain multiple sequence-specific binding motifs for transcription factors
  - short
  - gapless
  - more or less conserved

Example: *Drosophila melanogaster* eve gene





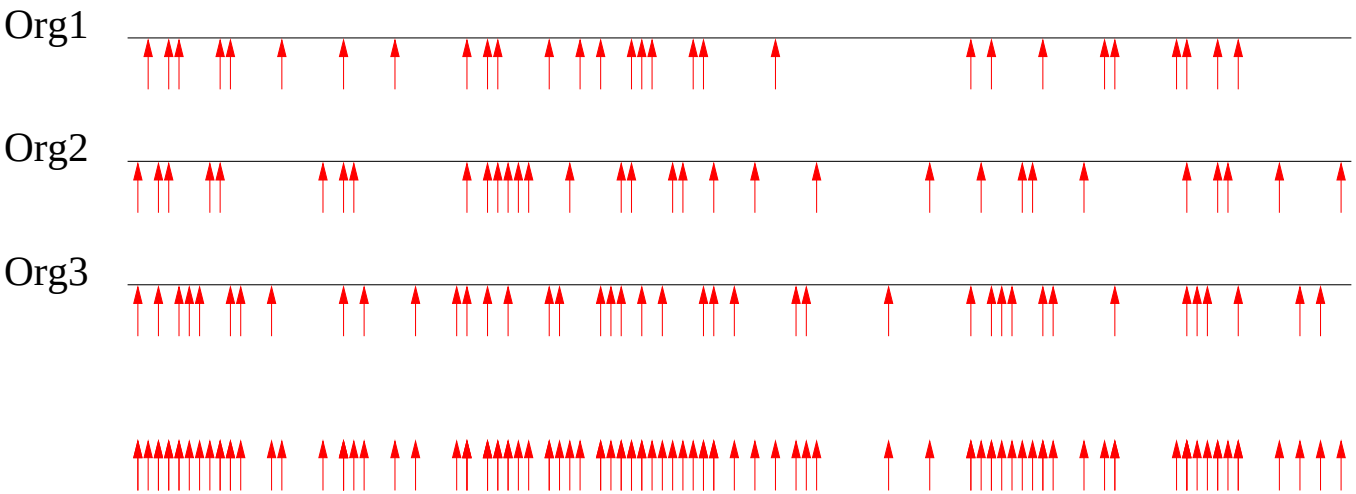


Identifying Regulatory Elements

computational approaches

- local (window of 100-200nt) overrepresentation of a particular TFBS (more observed BS than expected given a background model)
- mapping of genome-wide ChIP-seq data for the TF
- phylogenetic footprinting (for distant species)
- phylogenetic shadowing (for closely related species)

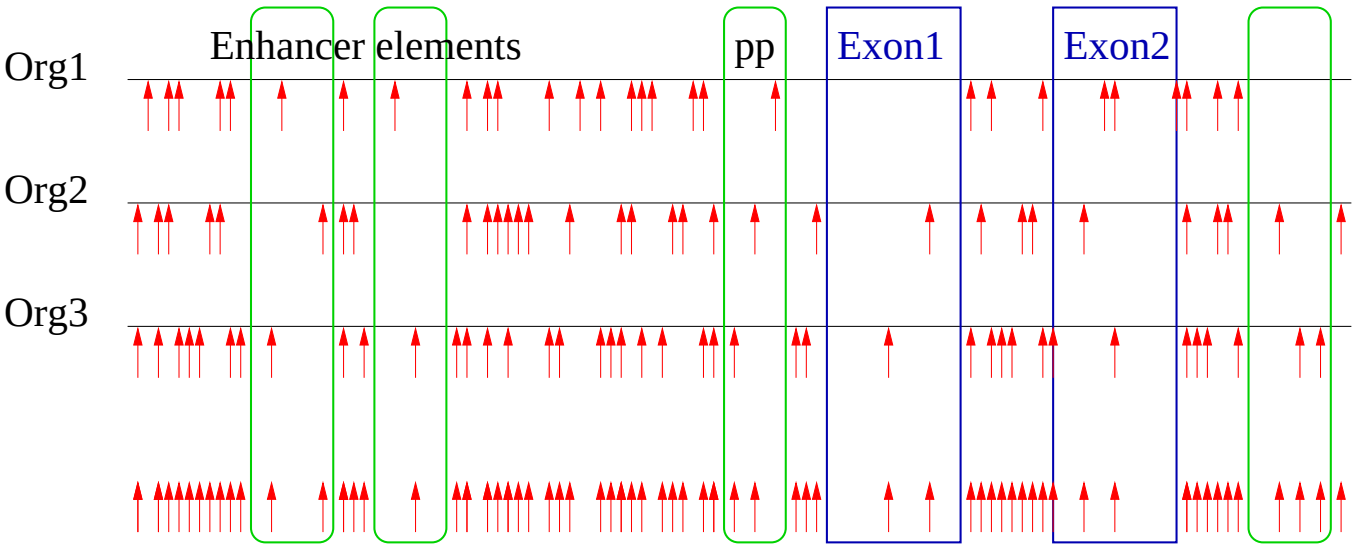
Phylogenetic Footprinting



divergent sites are indicated by red arrows

- functional DNA evolves much slower than non-functional DNA
- sequences of divergent species show some conservation
- conserved regions are thougt to be functional elements
- density and length of conserved regions decrease as evolutionary distances increase

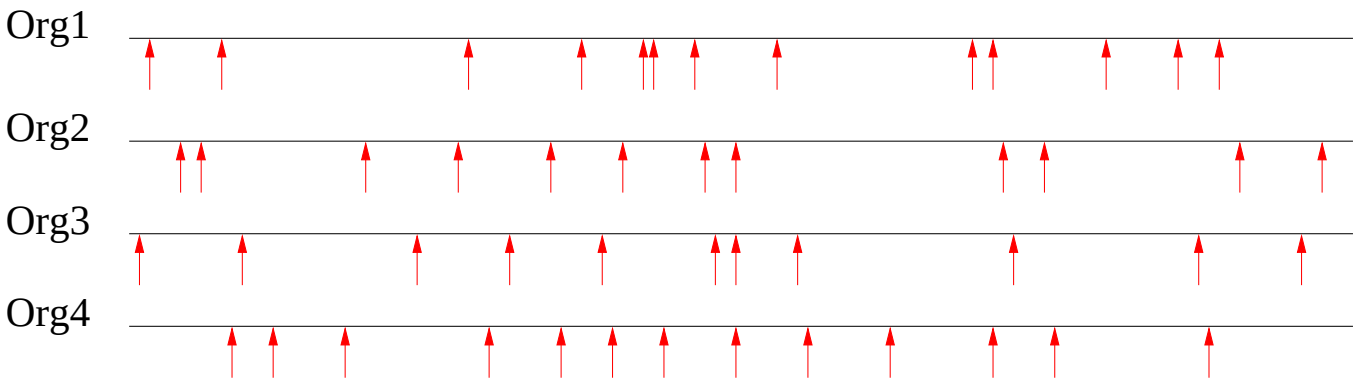
Phylogenetic Footprinting



divergent sites are indicated by red arrows

- functional DNA evolves much slower than non-functional DNA
- sequences of divergent species show some conservation
- conserved regions are thougt to be functional elements
- density and length of conserved regions decrease as evolutionary distances increase

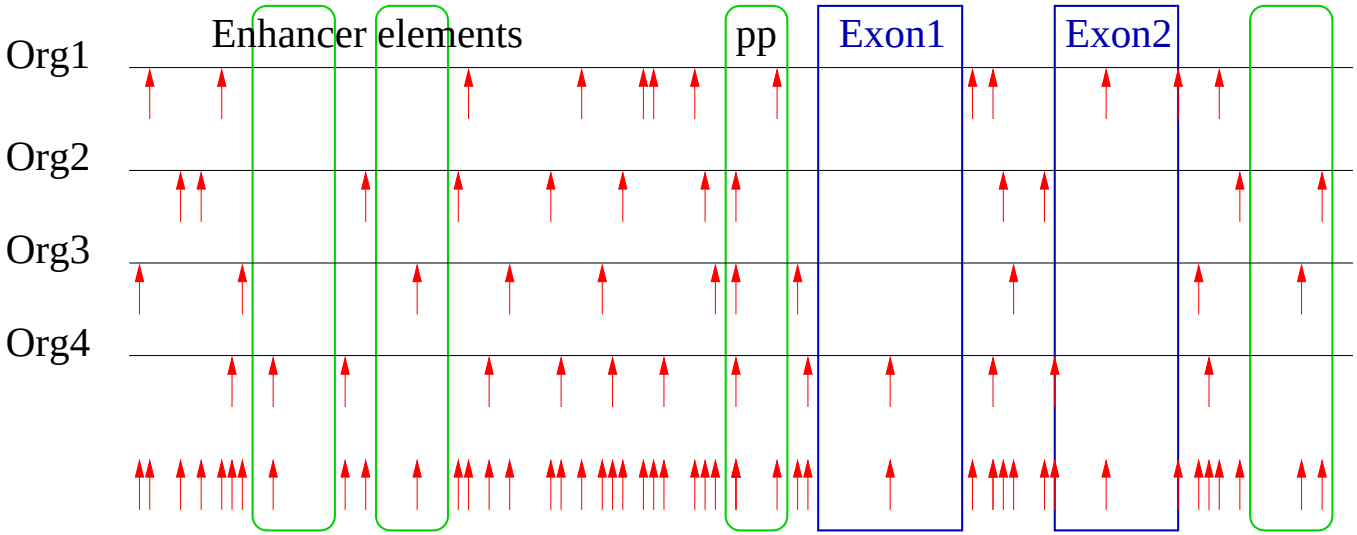
Phylogenetic Shadowing



divergent sites are indicated by red arrows

- sequences of closely related species show some differences
- differences from many sequences taken together reveal variable regions
- variability is assumed to be detrimental
- functional regions are said to lie in less variable regions

Phylogenetic Shadowing



divergent sites are indicated by red arrows

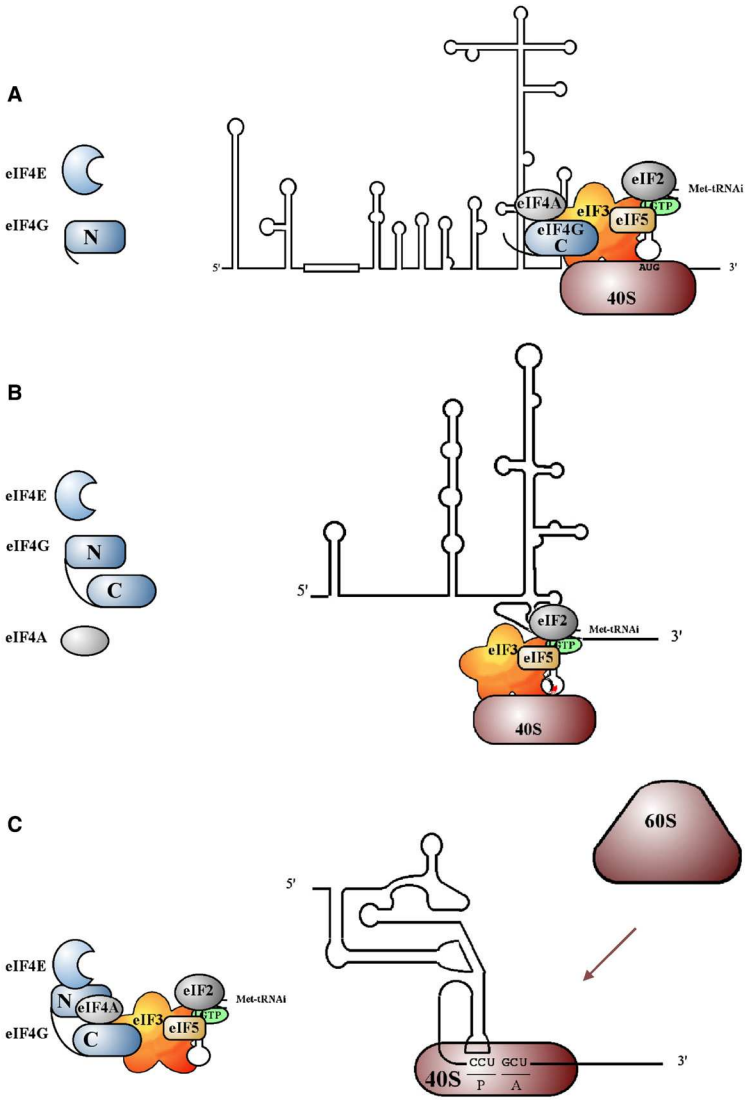
- sequences of closely related species show some differences
- differences from many sequences taken together reveal variable regions
- variability is assumed to be detrimental
- functional regions are said to lie in less variable regions

Role of RNA in gene regulation

- nascent transcripts can recruit specific RNA-binding proteins or RNAs to an actively transcribed genomic locus (e.g. TF binds RNA instead of DNA)
- non-coding sequences on mRNAs (UTRs) can sequence/structural elements regulating mRNA stability, and translation (e.g. upstream ORFs, IRES, miRNA target sites)
- small non-coding RNAs, like miRNAs (microRNAs), can a.o. regulate mRNA degradation or inhibition of translation)
- long non-coding RNAs can be involved in various regulatory mechanisms

internal ribosomal entry site – IRES

Factor requirement of different viral IRESes. (A) Most viral IRESes including Picornaviruses and Lentiviruses. All canonical initiation factors are required with the exception of eIF4E and the Nt of eIF4G. (B) Flaviviral/pestiviruses IRES. Ribosome entry does not require eIFs 4F/4A/4B/1/1A, but need eIF5, eIF2 and eIF3. (C) Dicistrovirus IGR IRESes. This 200 nt long IRES is able to contact directly a 40S subunit and to assemble an 80S elongation competent ribosome without any of the initiation factors.

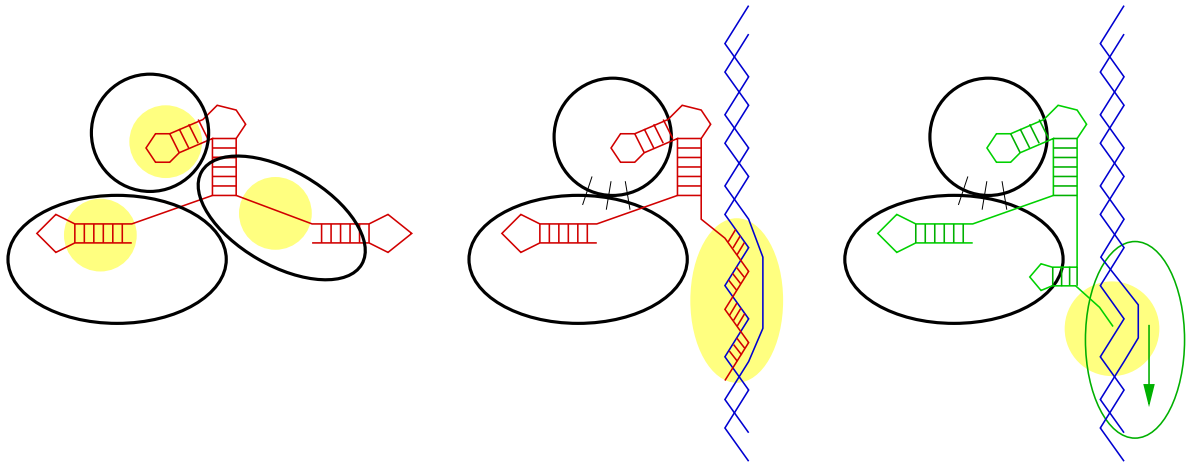


upstream ORFs

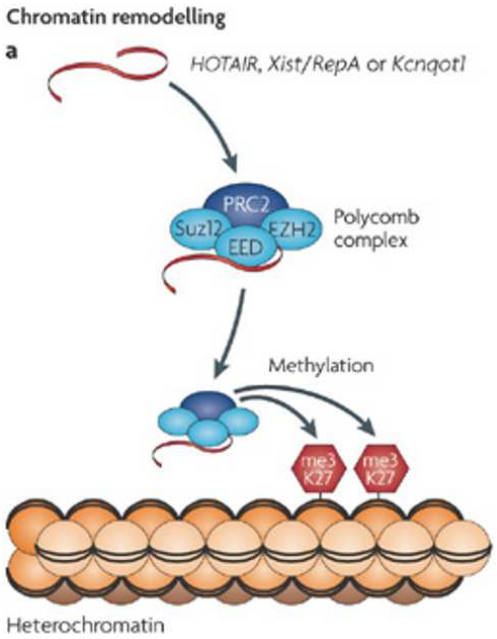
micro RNAs in left and right asymetric neural fate decision in *C.elegans*

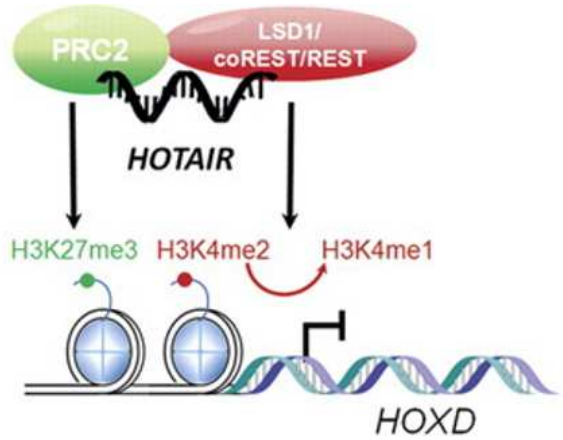
Role of Long ncRNAs

- the ncRNA is necessary to hold proteins of a protein complex together
  - the ncRNA specifies the set of proteins forming the protein complex
- the ncRNA links the protein complex to the genomic locus
- ncRNAs are not important only the act of transcription is



ncRNA hotair





- ncRNA hotair lies between hoxC and hoxC on the opposite strand
- it regulates expression of hoxD genes in *trans*
- it bind polycomb repressive complex 2 (PRC2) with its 5'-end which tri-methylates H3K27 (effect: repressive)
- it binds LSD1 with its 3'-end which demethylate mono- and di-methylated H3K4 and H3K9 (effect: repressive)