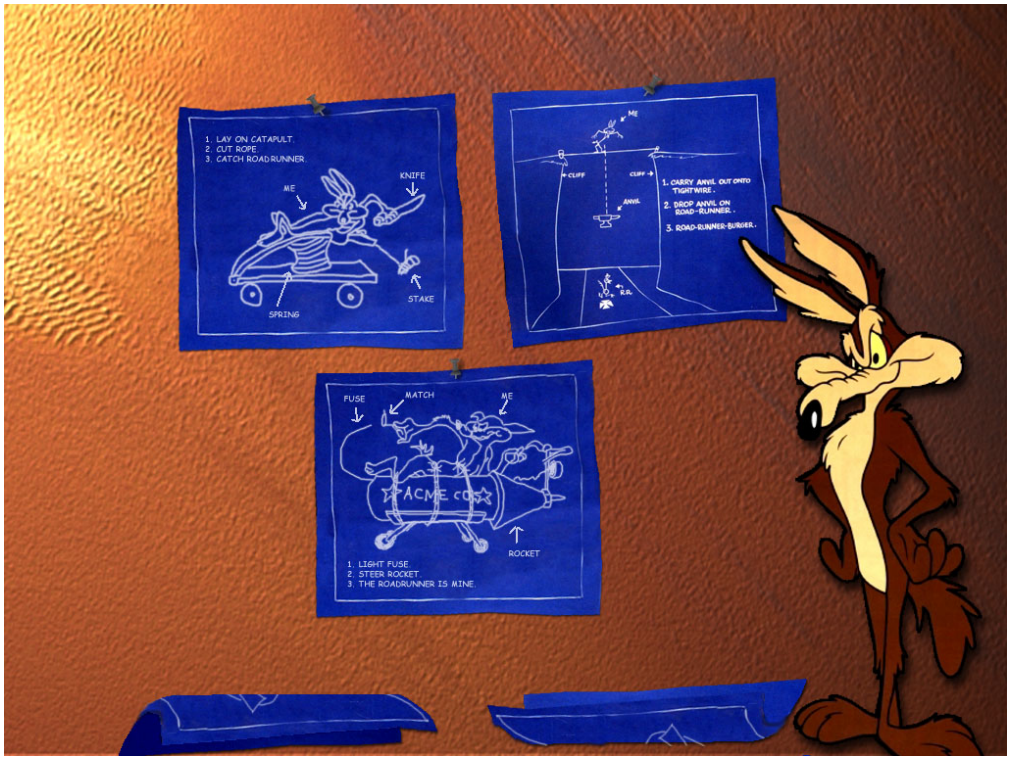
~~Tomorrow~~……today

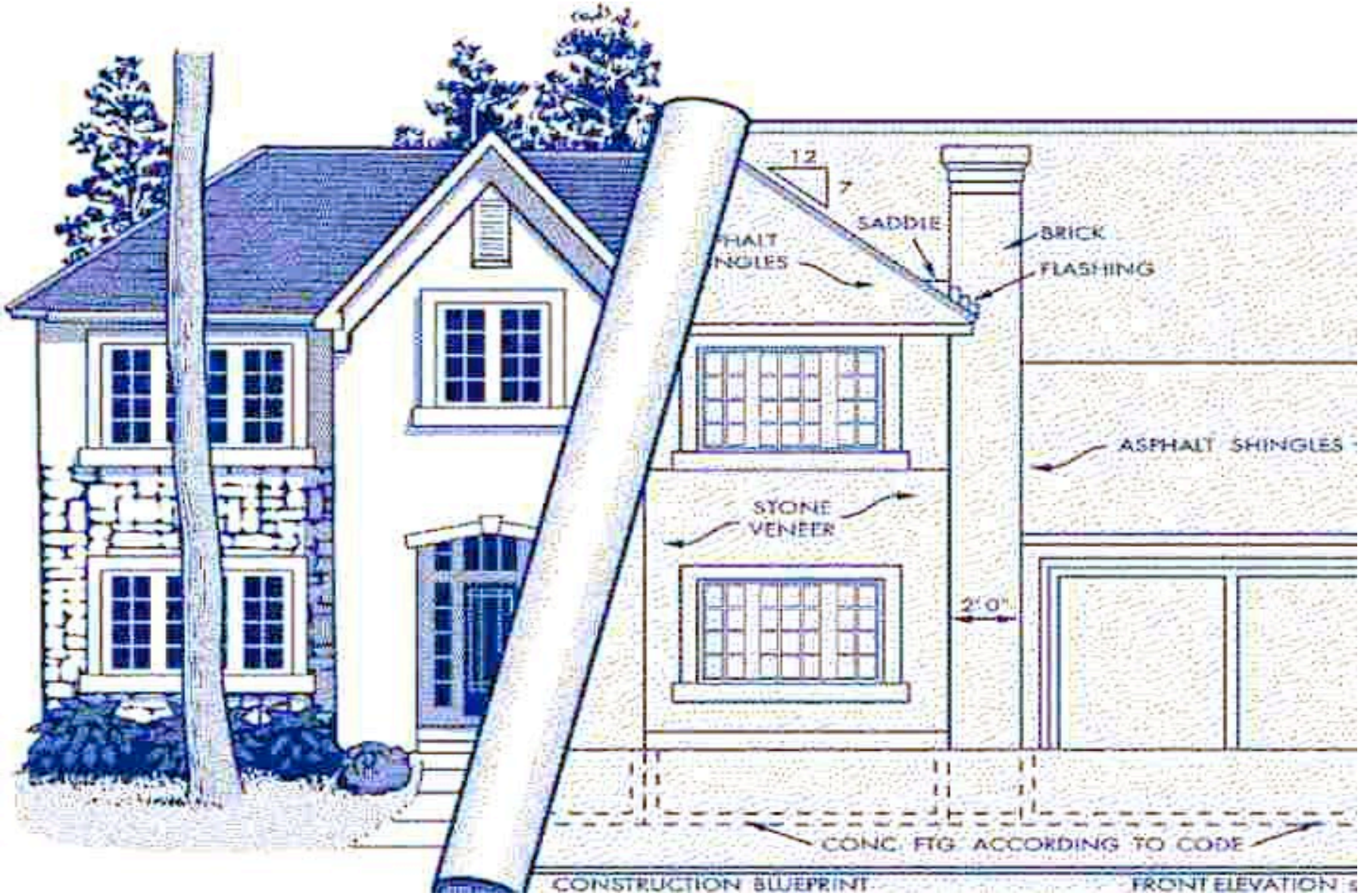
Junk-DNA is not junk at all

The DNA sequence is just the blueprint



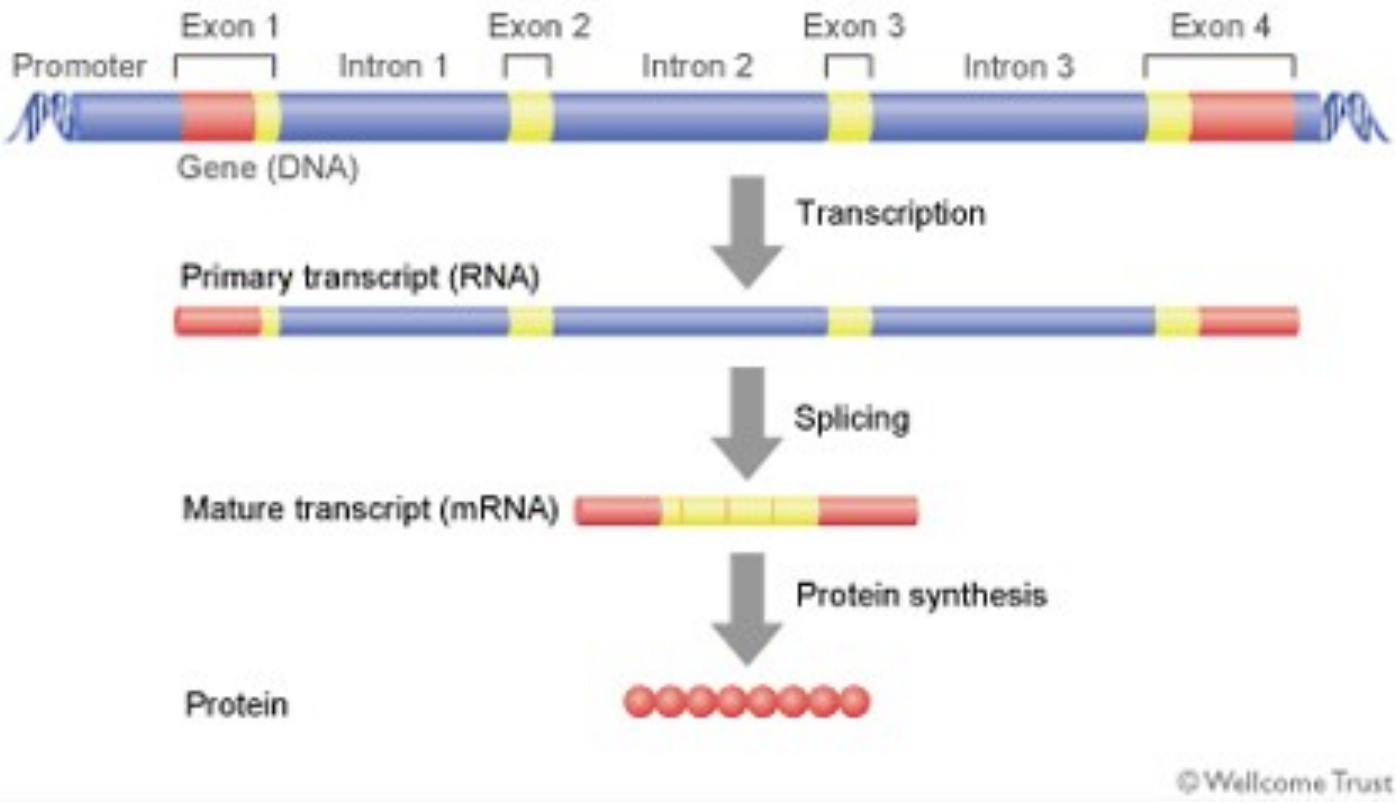
The genome is the blueprint

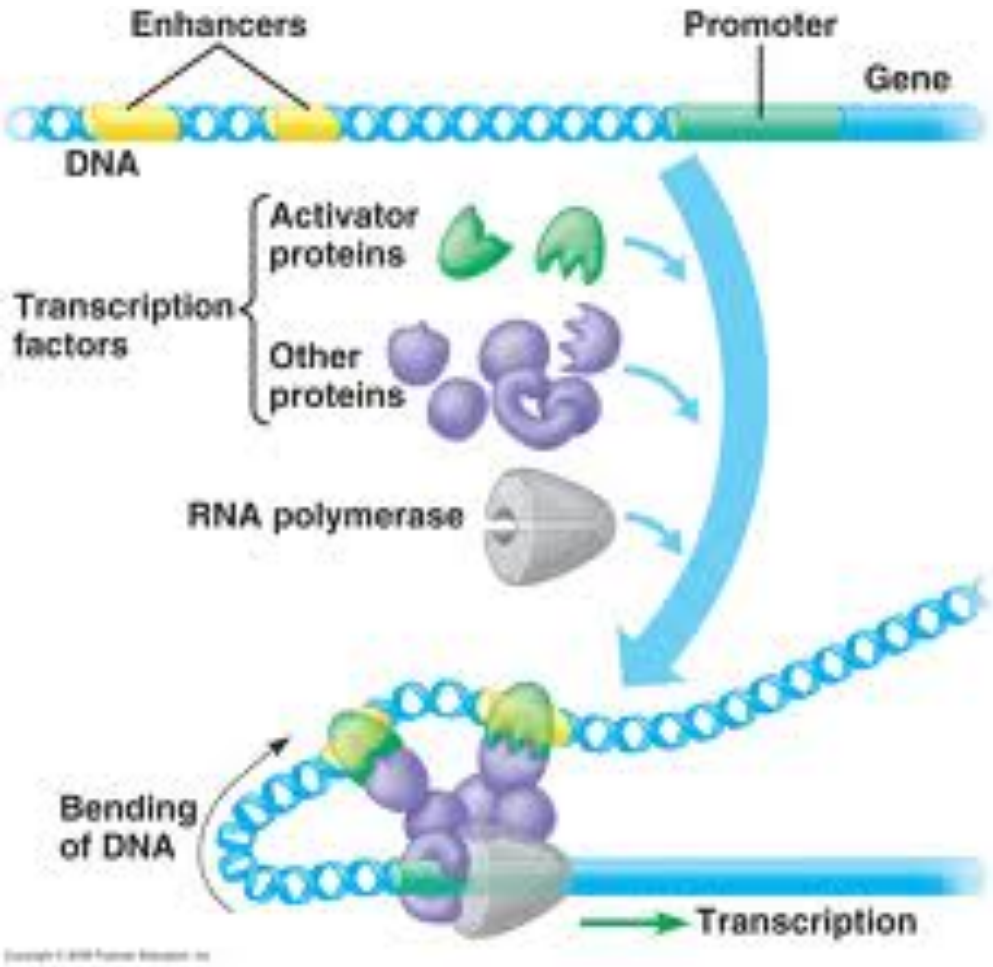
genome + epigenome



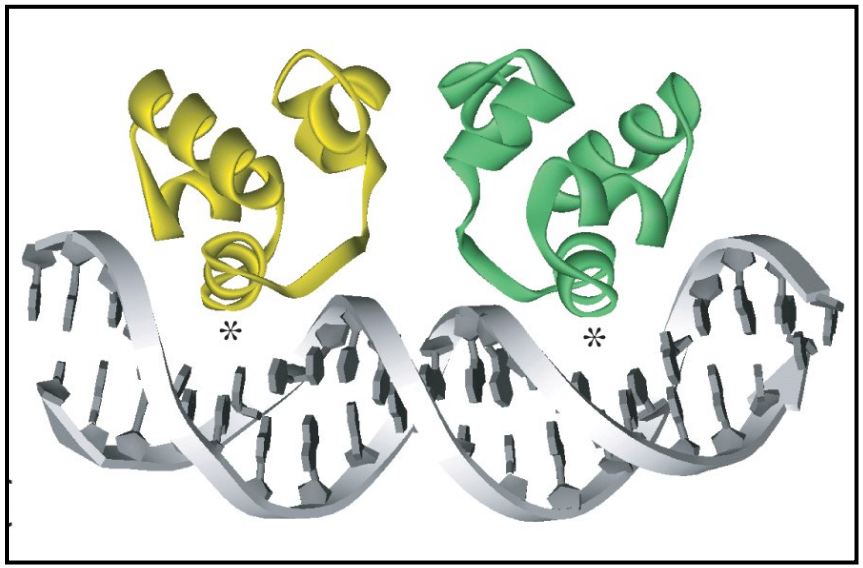
genome

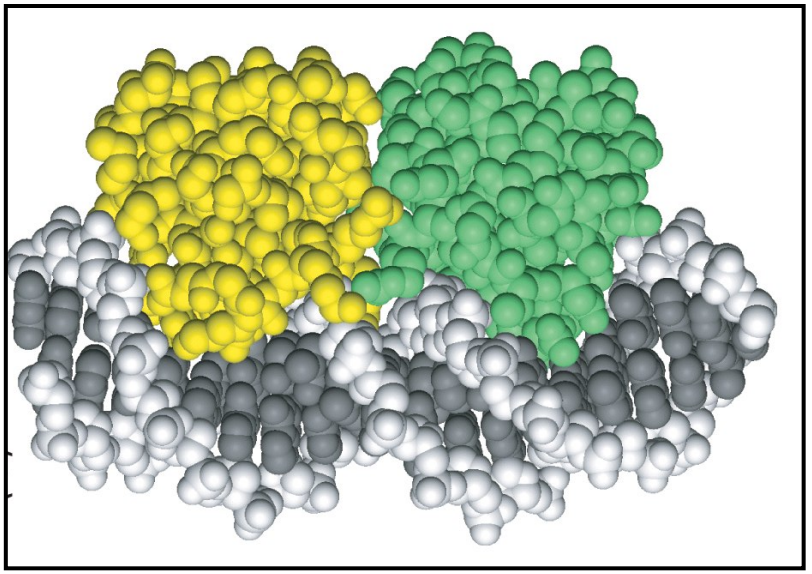
Structure of a Gene




Enhancers
Promoter
Gene
DNA
Activator proteins
Transcription factors
Other proteins
RNA polymerase
Bending of DNA
Transcription

Control of Gene Expression—Transcription Factors





Transcription factors (TFs) are proteins that bind to the DNA and help to control gene expression. We call the sequences to which they bind ***transcription factor binding sites (TFBSs)***, which are a type of *cis*-regulatory sequence.

Epigenetics

Επί:over, above

Epigenetics

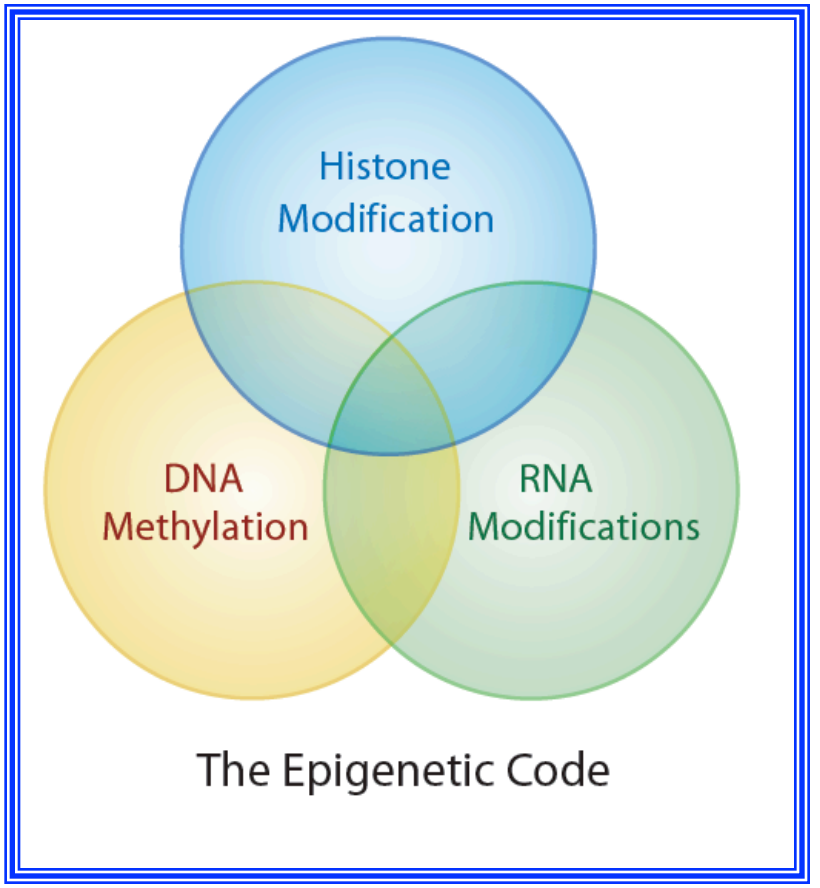
- The study of reversible heritable changes in gene function that occur without a change in the sequence of nuclear DNA

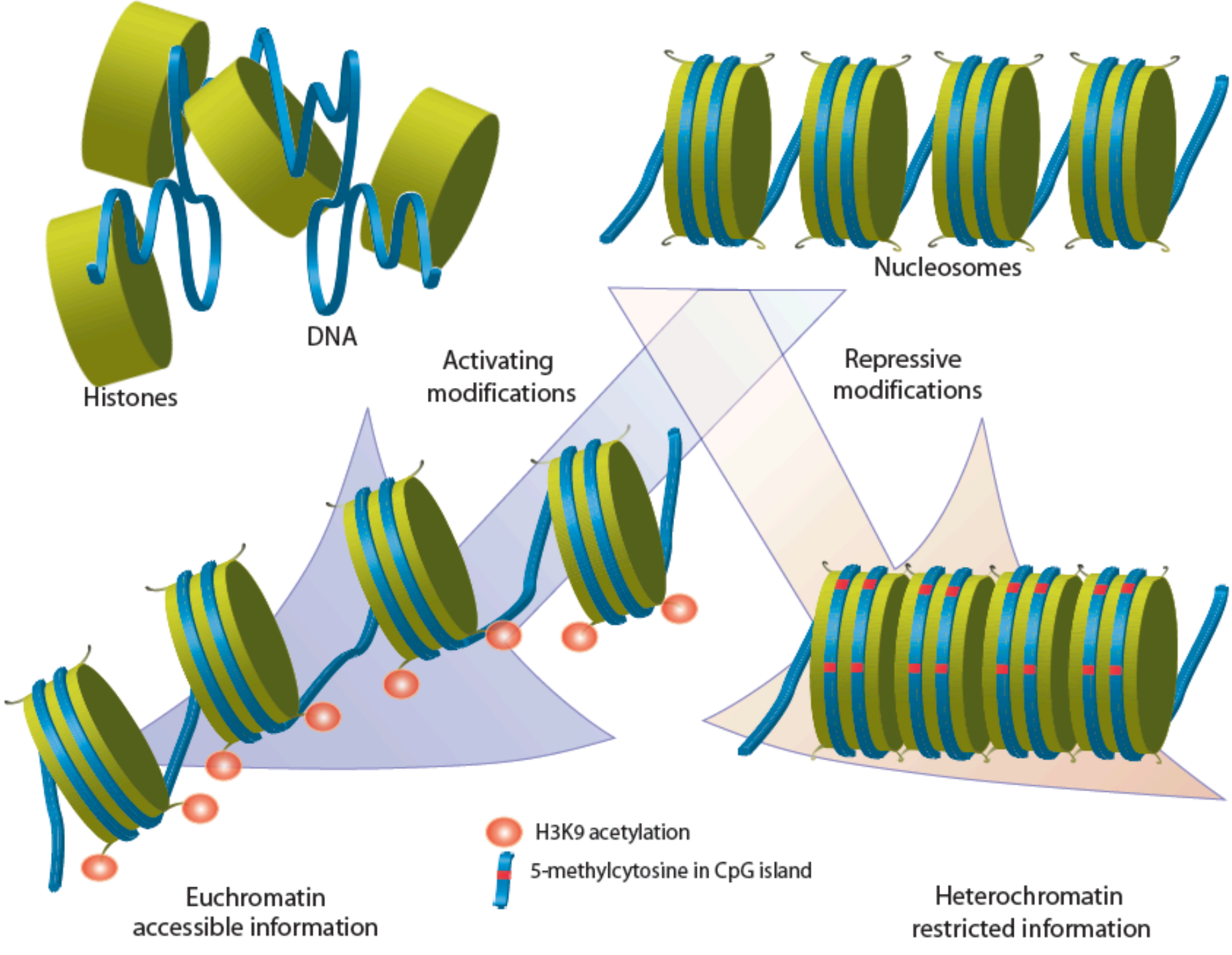
- Gene-regulatory information that is not expressed in DNA sequences is transmitted from one generation (of cells or organisms) to the next

The epigenetic code: no changes of DNA sequence

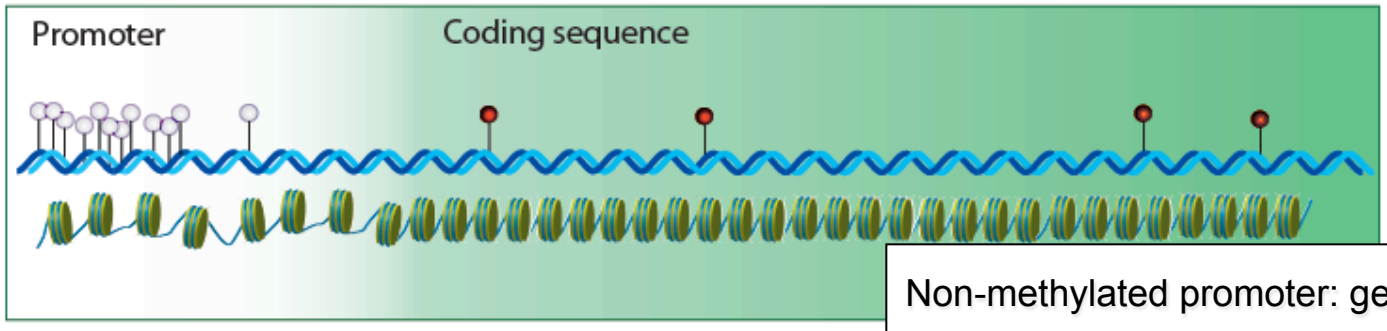
Epigenetic modifications include:

- ➢Reversible chemical modifications of histone proteins:
  - •Acetylation
  - •Methylation
  - •Phosphorylation
  - •……
- ➢DNA methylation
- ➢RNA modifications

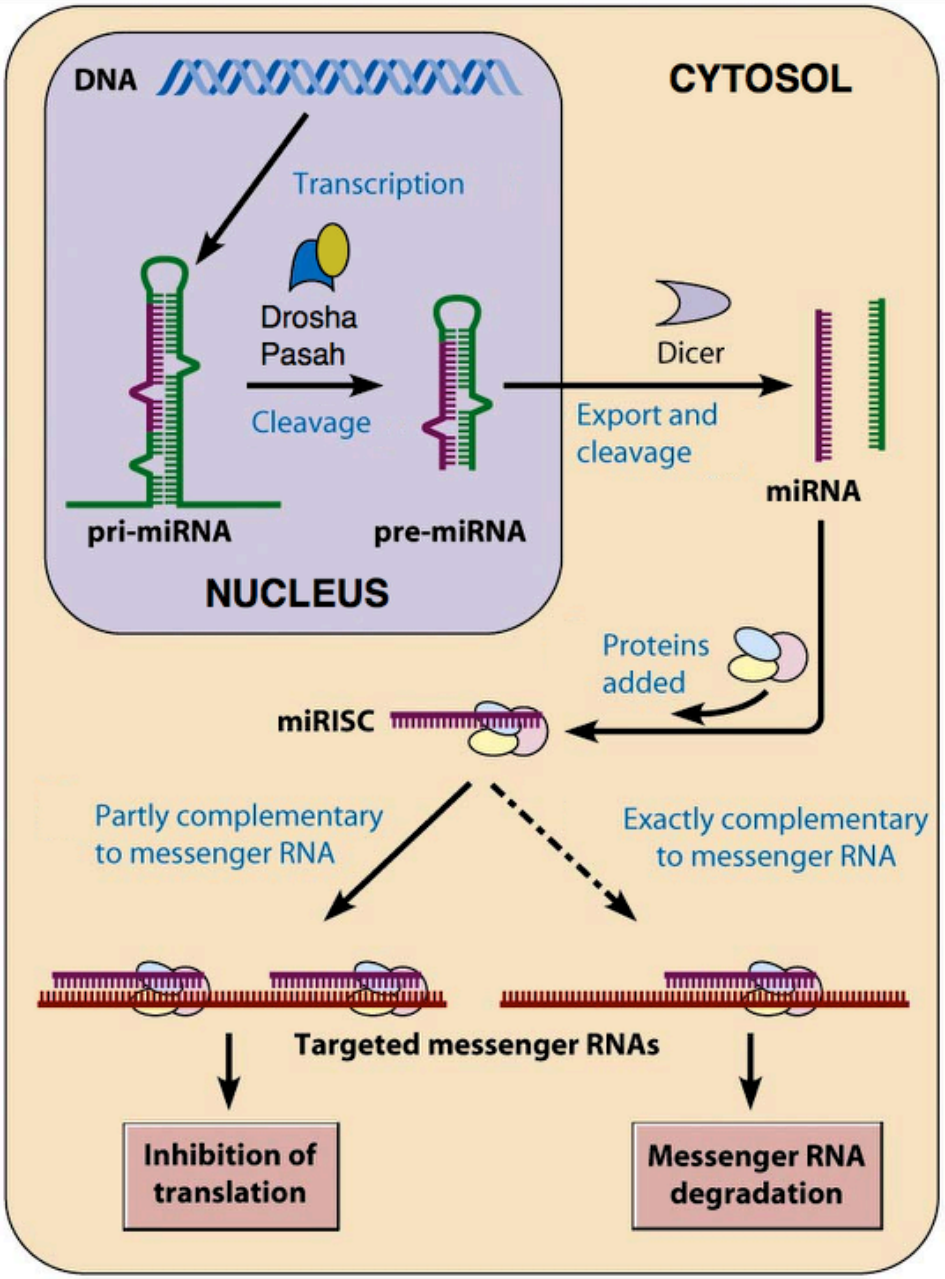



Histones
DNA
Nucleosomes
Activating modifications
Repressive modifications
H3K9 acetylation
5-methylcytosine in CpG island
Euchromatin accessible information
Heterochromatin restricted information

DNA methylation is associated with gene expression



- **Transcription factors do not interact with methylated promoters**
- **Specific proteins can bind to methylated DNA and recruit other proteins that induce DNA compaction**


DNA
CYTOSOL
Transcription
Drosha
Pasah
Cleavage
pri-miRNA
pre-miRNA
NUCLEUS
Dicer
Export and
cleavage
miRNA
Proteins
added
miRISC
Partly complementary
to messenger RNA
Exactly complementary
to messenger RNA
Targeted messenger RNAs
Inhibition of
translation
Messenger RNA
degradation

THE ENCODE PROJECT

Understanding the genome

We know the sequence—but can we understand it?

Anna Pavlovna's drawing room was gradually filling. The highest Petersburg society was assembled there: people differing widely in age and character but alike in the social circle to which they belonged. Prince Vasili's daughter, the beautiful Helene, came to take her father to the ambassador's entertainment; she wore a ball dress and her badge as maid of honor. The youthful little Princess Bolkonskaya, known as la femme la plus seduisante de Petersbourg, was also there. She had been married during the previous winter, and being pregnant did not go to any large gatherings, but only to small receptions. Prince Vasili's son, Hippolyte, had come with Mortemart, whom he introduced. The Abbe Morio and many others had also come.

To each new arrival Anna Pavlovna said, "You have not yet seen my aunt," or "You do not know my aunt?" and very gravely conducted him or her to a little old lady, wearing large bows of ribbon in her cap, who had come sailing in from another room as soon as the guests began to arrive; and slowly turning her eyes from the visitor to her aunt, Anna Pavlovna mentioned each one's name and then left them.

--Tolstoy, War and Peace

Understanding the genome

We don't know the language:

Гостиная Анны Павловны начала понемногу наполняться. Приехала высшая знать Петербурга, люди самые разнородные по возрастам и характерам, но одинаковые по обществу, в каком все жили; приехала дочь князя Василия, красавица Элен, заехавшая за отцом, чтобы с ним вместе ехать на праздник посланника. Она была в шифре и бальном платье. Приехала и известная, как la femme la plus séduisante de Pétersbourg 1, молодая, маленькая княгиня Болконская, прошлую зиму вышедшая замуж и теперь не выезжавшая в *большой* свет по причине своей беременности, но ездившая еще на небольшие вечера. Приехал князь Ипполит, сын князя Василия, с Мортемаром, которого он представил; приехал и аббат Морио и многие другие.

— Вы не видали еще, — или: — вы не знаком — говорила Анна Павловна приезжавшим гостям и весьма серьезно подводила их к маленькой старушке в высоких бантах, выплывшей из другой комнаты, как скоро стали приезжать гости, называла их по имени, медленно переводя глаза с гостя на ma tante, и потом отходила.

Все гости совершали обряд приветствования никому не известной, никому не интересной и не нужной тетушки. Анна Павловна с грустным, торжественным участием следила за их приветствиями, молчаливо одобряя их. Ma tante каждому говорила в одних и тех же выражениях о его здоровье, о своем здоровье и о здоровье ее величества, которое нынче было, слава Богу, лучше. Все подходившие, из приличия не выказывая поспешности, с чувством облегчения исполненной тяжелой обязанности отходили от старушки, чтоб уж весь вечер ни

--Tolstoy, War and Peace

Understanding the genome

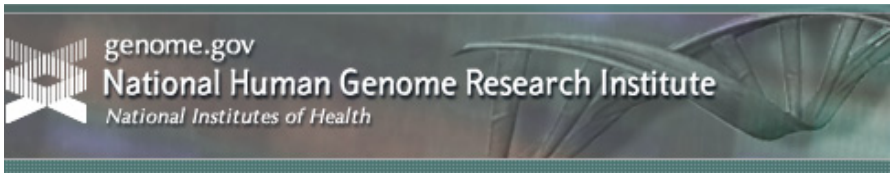
Even if we did, we don't know the grammar, or punctuation:

annapavlovnasdrawingroomwasgraduallyfillingthehighestpetersburgsocietywasass
embledtherepeopledifferingwidelyinageandcharacterbutalikeinthesocialcircletowhic
htheybelongedprincevasilisdaughterthebeautifulhelenecametotakeherfathertothea
mbassadorsentertainmentsheworeaballdressandherbadgeasmaidofhonortheyouthf
ullittleprincessbolkonskayaknownaslafemmelaplusseduisantedepetersbourgwasals
othereshehadbeenmarrieddurингthepreviouswinterandbeingpregnantdidnotgotoan
ylargegatheringsbutonlytosmallreceptionsprincevasilissonhippolytehadcomewithm
ortemartwhomheintroducedtheabbemorioandmanyothershadalsocometoeachnewa
rrivalannapavlovnasaidyouhavenotyetseenmyauntoryoudonotknowmyauntandvery
gravelyconductedhimorhertoalittleoldladywearinglargebowsofribboninhercapwhoha
dcomesailinginfromanotherroomassoonastheguestsbegantoarriveandslowlyturning
hereyesfromthevisitortoherauntannapavlovnamentionedeachonesnameandthenleft
themeachvisitorperformedtheceremonyofgreetingthisoldauntwhomnotoneofthemkn
ewnotoneofthemwantedtoknowandnotoneofthemcaredaboutannapavlovnaobserve
dthesegreetingswithmournfulandsolemninterestandsilentapprovaltheauntspoketoe
achofthemin thesamewordsabouttheirhealthandherownandthehealthofhermajestyw
hothankgodwasbettertodayandeachvisitorthoughpolitenesspreventedhisshowingim
patienceleftttheoldwomanwithasenseofreliefathavingperformedavexatiousdutyan

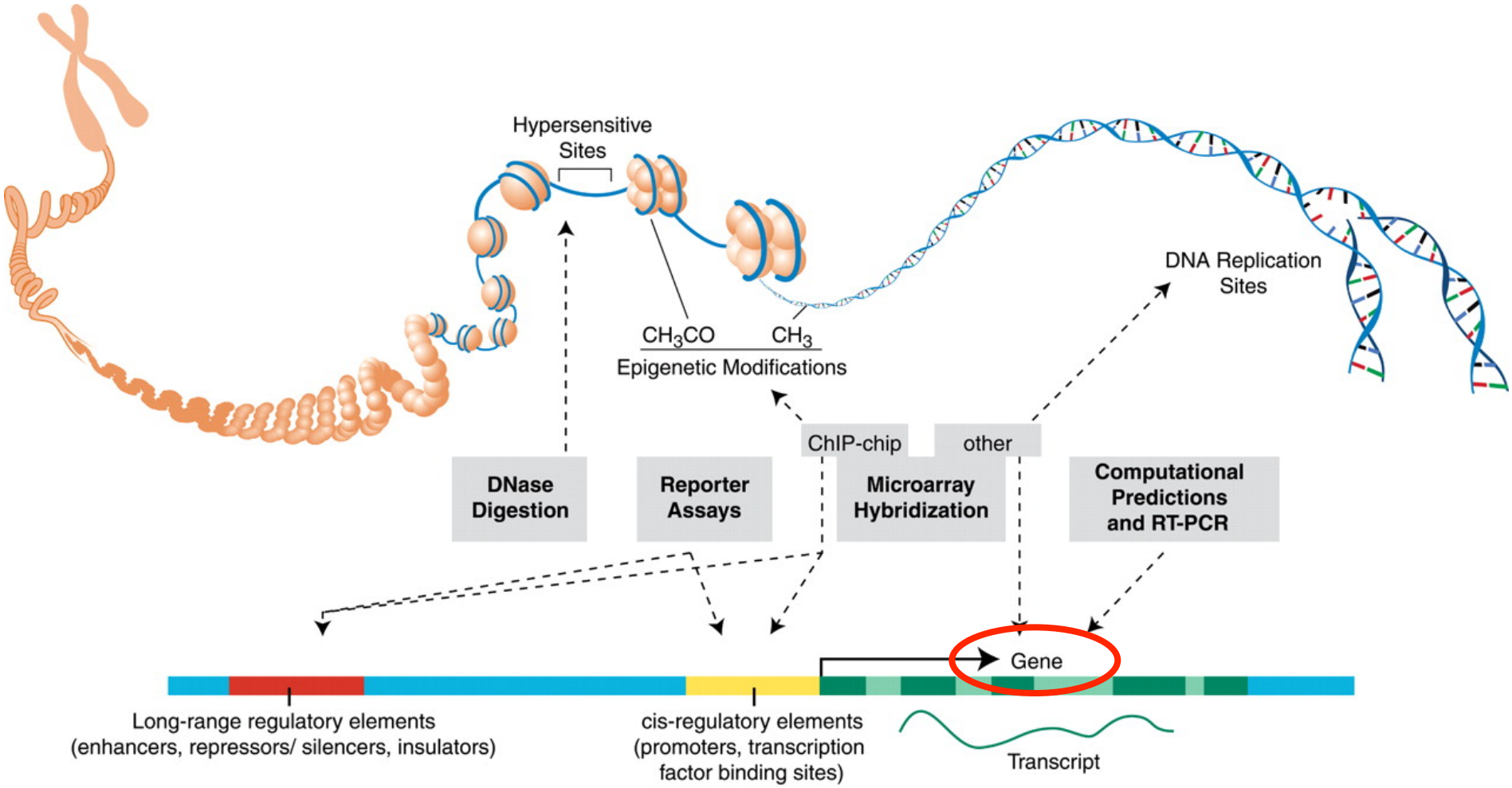
--Tolstoy, War and Peace

With the complete human genome sequence now in hand, we face the enormous challenge of interpreting it and learning how to use that information to understand the biology of human health and disease.

The **ENCyclopedia Of DNA Elements (ENCODE) Project** is predicated on the belief that a comprehensive catalog of the structural and functional components encoded in the human genome sequence will be critical for understanding human biology well enough to address those fundamental aims of biomedical research.



Functional genomic elements being identified by the ENCODE pilot phase



The ENCODE Project Consortium Science 306, 636 -640 (2004)

Published by AAAS

Science
AAAS

The competing endogenous RNA (ceRNA)

Only 1.5-3% of the genome is coding for proteins, whereas it is becoming clear that approximately 50% of the genome is transcribed into RNA material and that the information carrying capacity of the genome is much more vast, extending beyond the protein coding genome

The non-coding revolution

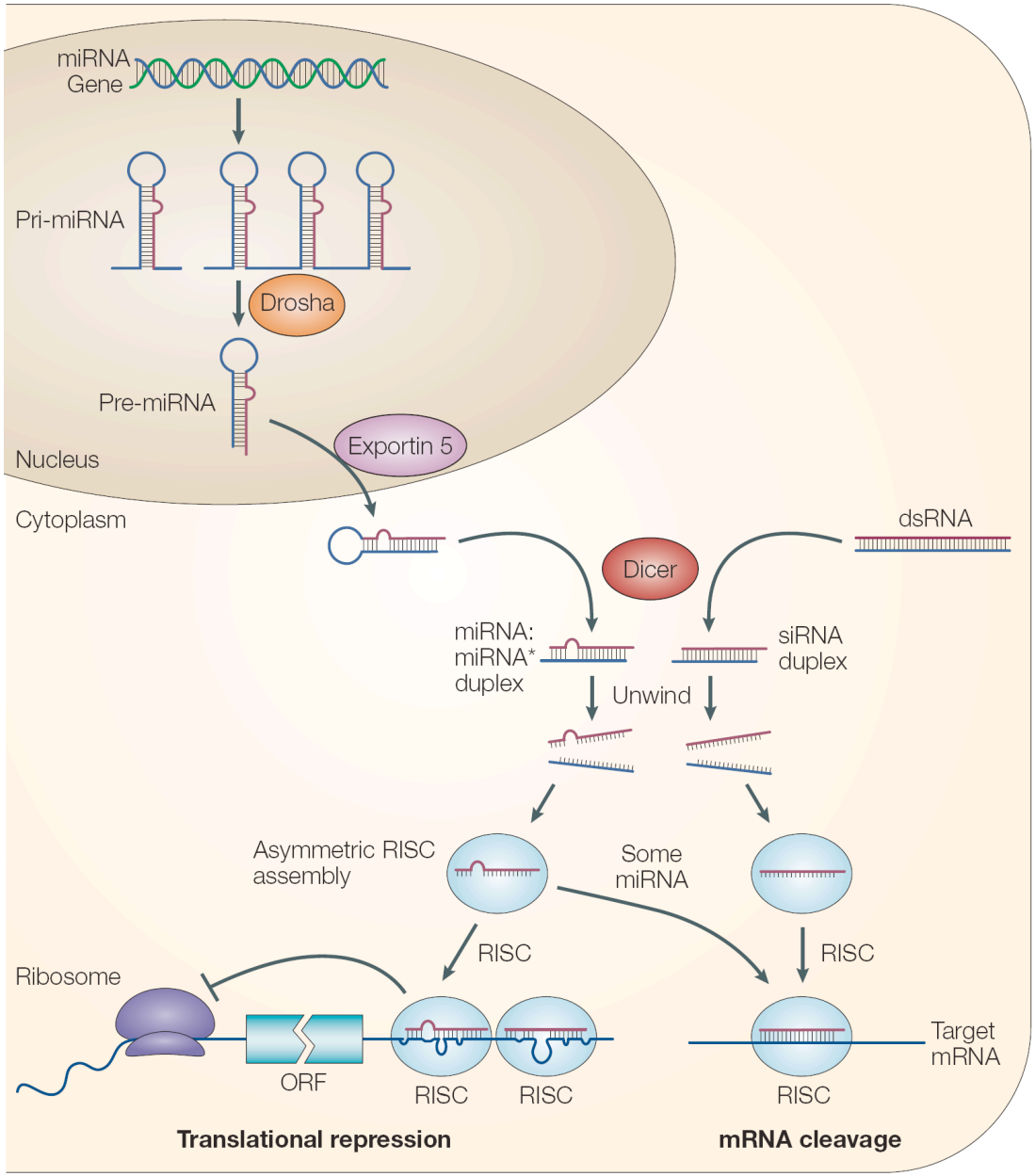
- Some lower organisms have been identified to have comparable numbers of protein-coding genes when measured up to humans
- A small numerical difference in coding genes is not sufficient to explain the diversity of cell types and tissues found in complex organisms
- The developmental complexity of organisms is more closely related to the amount of non-coding sequences in genomes, suggesting that they harbor critical regulatory information
- Most of the mammalian trascriptome does not correspond to annotated exons of protein-coding genes
- ……therefore, the fraction of the mammalian genome that is used as "information messenger" is much greater that previously predicted

The ceRNA

- The microRNAs
- The coding genes
- Pseudogenes
- Long non-coding RNAs (lncRNAs)

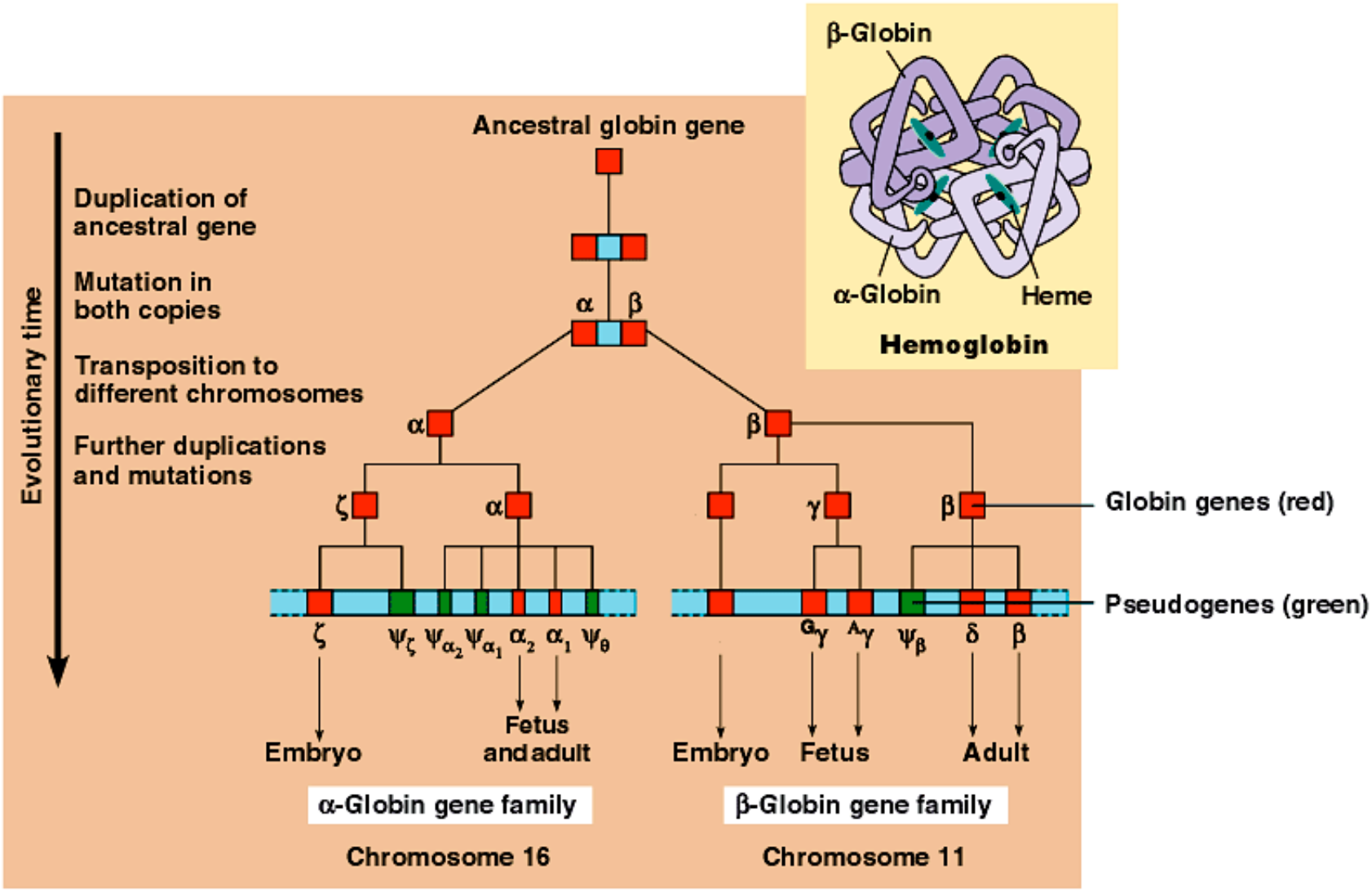
The microRNAs

- the human genome is predicted to encode several thousand microRNAs
- microRNAs bind preferentially, but not exclusively, to the 3’ UTR of their targets through recognition of sequences referred to as microRNA recognition elements (MREs)
- microRNAs can function in a combinatorial manner if mRNA transcrips harbor numerous MREs for more than one microRNA
- since each microRNA may repress up to hundreds of mRNAs the mammalian transcriptome is regulated in this manner


miRNA Gene
Pri-miRNA
Drosha
Pre-miRNA
Exportin 5
Nucleus
Cytoplasm
dsRNA
Dicer
miRNA: miRNA* duplex
siRNA duplex
Unwind
Asymmetric RISC assembly
Some miRNA
RISC
RISC
Ribosome
ORF
RISC
RISC
RISC
Target mRNA
Translational repression
mRNA cleavage

The coding genes

- In the human genome, approximately 20,000 protein-coding genes have been identified
- mRNA sequences of coding genes are often densely covered in MREs, which explains how their expression is regulated through microRNAs



Pseudogenes

- A pseugene resembles a known gene (previously defined as “non-functional”, “junk” or “evolutionary relic”)
- Pseudogenes contain mutations which prevent them from encoding fully functional proteins
- Human sequencing genome revealed the existence of about 19,000 pseudogenes, many of which are expressed as RNA transcripts
- Pseudogene sequences are often well conserve
- Expressed pseudogene RNAs will be targeted by the same microRNAs that target their ancestral related genes

Long non-coding RNAs (lncRNAs)

- lncRNAs are typically 300 to thousands of nucleotides in length
- The precise number of lncRNAs is expanding, approximating to date the 10,000 genetic units
- They are involved in regulatory mechanisms (X chromosome inactivation, transcriptional initiation in eukariotes…….)
- Similar to coding genes and pseudogenes, lncRNAs are densely populated with MREs

The competing endogenous RNA language

- microRNA are negative regulators of gene expression through binding to specific MRE sequences and, consequently, decreasing the stability of target RNAs and/or limiting their translation
- RNAs can regulate each other through their ability to compete for microRNA binding: target RNAs can sequester microRNAs, thereby protecting other target RNAs from down-regulation by the sequestered microRNAs
- protein-coding mRNAs may possess a second independent and genetically determined function through their ability to regulate other RNAs

ceRNA networks will depend on the identity, concentration and subcellular distribution of the RNA and microRNA species that are present in a given cell type at a given moment

we have neglected the full function of a gene, which we now know emanates from both the protein and the transcript, independently

ARTICLE

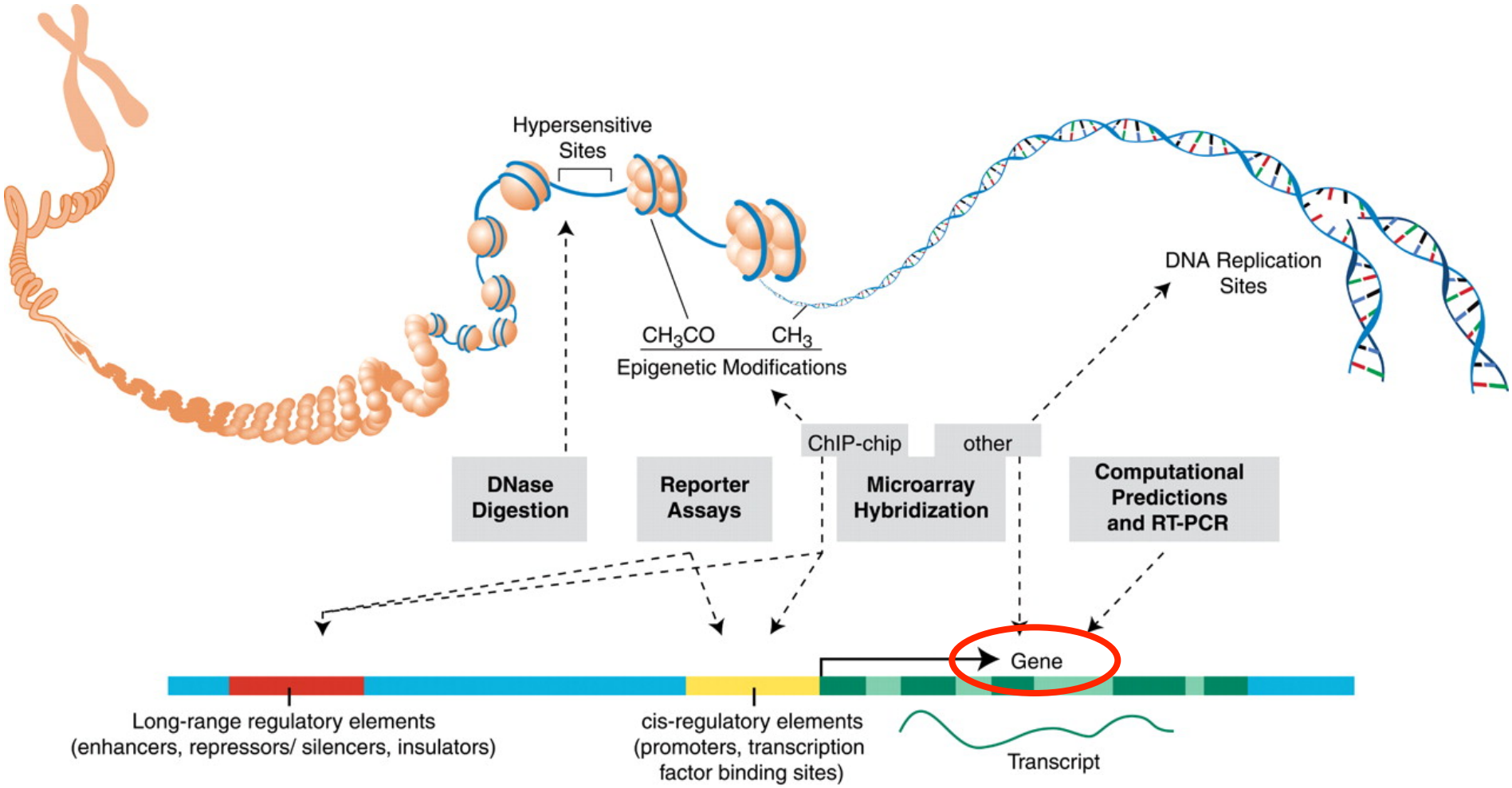
doi:10.1038/nature11247

An integrated encyclopedia of DNA elements in the human genome

The ENCODE Project Consortium*

The human genome encodes the blueprint of life, but the function of the vast majority of its nearly three billion bases is unknown. The Encyclopedia of DNA Elements (ENCODE) project has systematically mapped regions of transcription, transcription factor association, chromatin structure and histone modification. These data enabled us to assign biochemical functions for 80% of the genome, in particular outside of the well-studied protein-coding regions. Many discovered candidate regulatory elements are physically associated with one another and with expressed genes, providing new insights into the mechanisms of gene regulation. The newly identified elements also show a statistical correspondence to sequence variants linked to human disease, and can thereby guide interpretation of this variation. Overall, the project provides new insights into the organization and regulation of our genes and genome, and is an expansive resource of functional annotations for biomedical research.

Functional genomic elements being identified by the ENCODE pilot phase

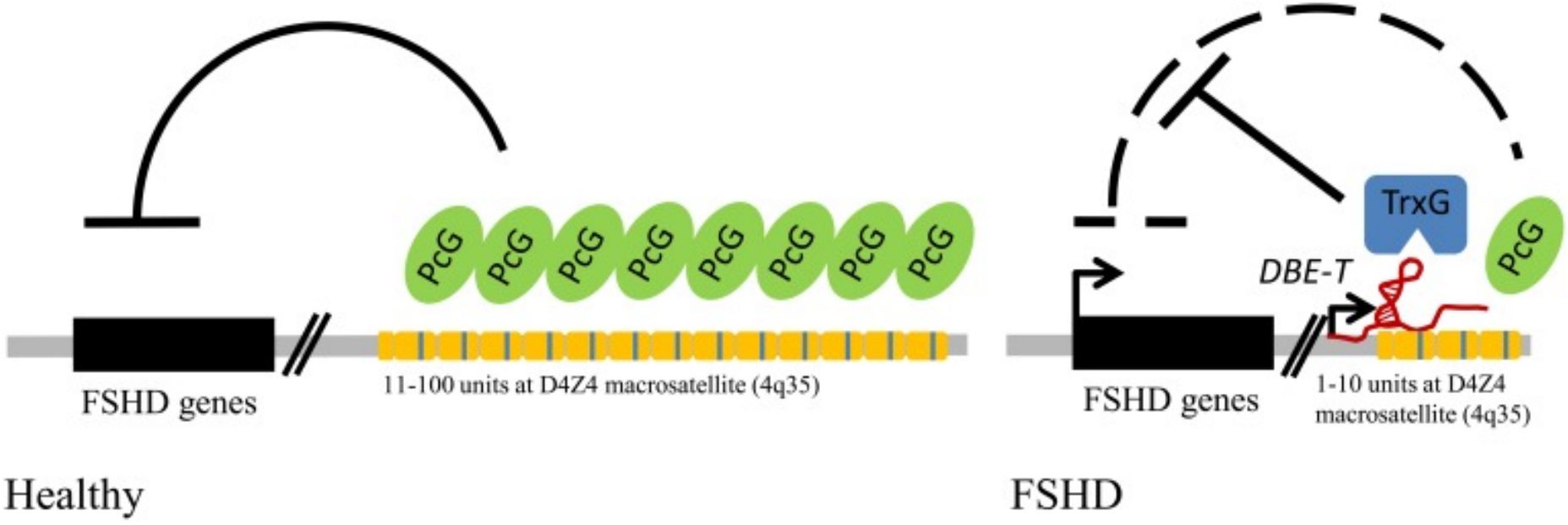


The ENCODE Project Consortium Science 306, 636 -640 (2004)

Published by AAAS

Science
AAAS

Recent advances in facioscapulohumeral muscular dystrophy



FSHD muscular dystrophy links repetitive elements, Polycomb proteins, and ncRNAs in a human genetic disease. Model for FSHD molecular pathogenesis: in healthy individuals the repetitive elements (yellow modules) of the D4Z4 macrosatellite at 4q35 are bound by Polycomb (PcG) proteins, which mediate gene repression; in FSHD patients the shortening below the threshold of 11 copies generates an epigenetic remodeling of the locus, sustained by a long non-coding RNA (DBE-T), and the recruitment of Trithorax (TrxG) proteins, driving histone H3 lysine 36 dimethylation, chromatin remodeling and leading to up-regulation of disease genes.

HGP, ENCODE……

Cancer Genome Project

Genomics & Genetics
Overview
CGP
Faculty
Stratton
Futreal

Projects
Cancer Gene Census
COSMIC
CGP Resequencing Studies
Copy Number Mapping
NCI-60
Planned studies
Genomics of Drug Sensitivity in Cancer
Software

Information
Links
News
Publications
Conditions of use

Hide Navigation

Cancer Gene Census

Overview

The Cancer Gene Census is an ongoing effort to catalogue those genes for which mutations have been causally implicated in cancer. The original census and analysis was published in Nature Reviews Cancer and supplemental analysis information related to the paper is also available.

The census is not static but rather is updated regularly/as needed. In particular we are grateful to Felix Mitelman and his colleagues in providing information on more genes involved in uncommon translocations in leukaemias and lymphomas. Currently, more than 1% of all human genes are implicated via mutation in cancer. Of these, approximately 90% have somatic mutations in cancer, 20% bear germline mutations that predispose to cancer and 10% show both somatic and germline mutations.

The Cancer Genome Atlas

The Cancer Genome Atlas (TCGA) is a comprehensive and coordinated effort to accelerate our understanding of the molecular basis of cancer through the application of genome analysis technologies, including large-scale genome sequencing. TCGA is a joint effort of the National Cancer Institute (NCI) and the National Human Genome Research Institute (NHGRI), which are both part of the National Institutes of Health, U.S. Department of Health and Human Services.

Mission and Goal

The Cancer Genome Atlas will assess the feasibility of a full-scale effort to systematically explore the entire spectrum of genomic changes involved in human cancer.

The overarching goal of The Cancer Genome Atlas is to improve our ability to diagnose, treat and prevent cancer.

The -omics era

- Genomics: is a discipline that applies DNA sequencing methods and bioinformatics to sequence, assemble, and analyze the function and structure of genomes (the complete set of DNA within a single cell of an organism)
- Transcriptomics, or genome-wide expression profiling, aims to catalogue the complete set of RNA transcripts produced by the genome
- Proteomics is the large-scale study of proteins, particularly their structures and functions.
- Methylomics….metabolomics…….pharmacogenomics…….

- -omics and back again!!

Thank you