Functional and Comparative Genomics

ESCI upf.
School of International Studies

Name of the course: Functional and Comparative genomics
Academic year: 2023/2024
Year: 2nd
Term: 3rd
Code: 52217
Number of credits: 4 credits
Total number of hours committed: 40 hours
Teaching language: English
Lecturer: Toni Gabaldón the subject coordinator. Other professors contributing to the teaching of practical and theoretical classes will be Eduard Ocaña, and some PhD candidates from my group
Timetable: See official calendar


TAs ;

- Saioa Manzano
- Giacomo Mutti
- Moisès Bernabeu



More on our research:

Site: http:/cgenomics.org

Twitter: @gabaldonlab

Scholargoogle, Pubmed


Barcelona
Supercomputing
Center

- **Basic description of contents outlined for the curriculum**

This course covers the basic principles of comparative and functional genomics, in both prokaryotic and eukaryotic organisms. Subjects include genome and chromosomal re-arrangements, synteny analysis; origin of genes; gene duplication and loss, orthology and paralogy, functional evolution of genes, analysis of sequence conservation and reconstruction of evolutionary histories of genes; functional annotation of genes through computational means and gene expression analysis.

Format:

- 9 Theoretical sessions (2h each)
- 9 Practical sessions (2h each)
- Small project on comparative and functional analysis of a gene (pairs)
- Research paper discussion and presentation (groups of 3-4)

Evaluation:

Assessment elements	Time period	Type of assessment		Assessment agent			Type of activity	Grouping		Weight (%)
		Comp	Opt	Lecturer	Self-assess	Co-assess		Indiv	Group (#)	
Theoretical final exam		x		All				x		60%
Final project		x		All					x	15%

- Exam: includes test questions, onpen-ended questions and problems. Covers both theoretical classes and practicals

- Assignment: Study function and evolution of one protein. Present a final report in paper format. (in small groups)

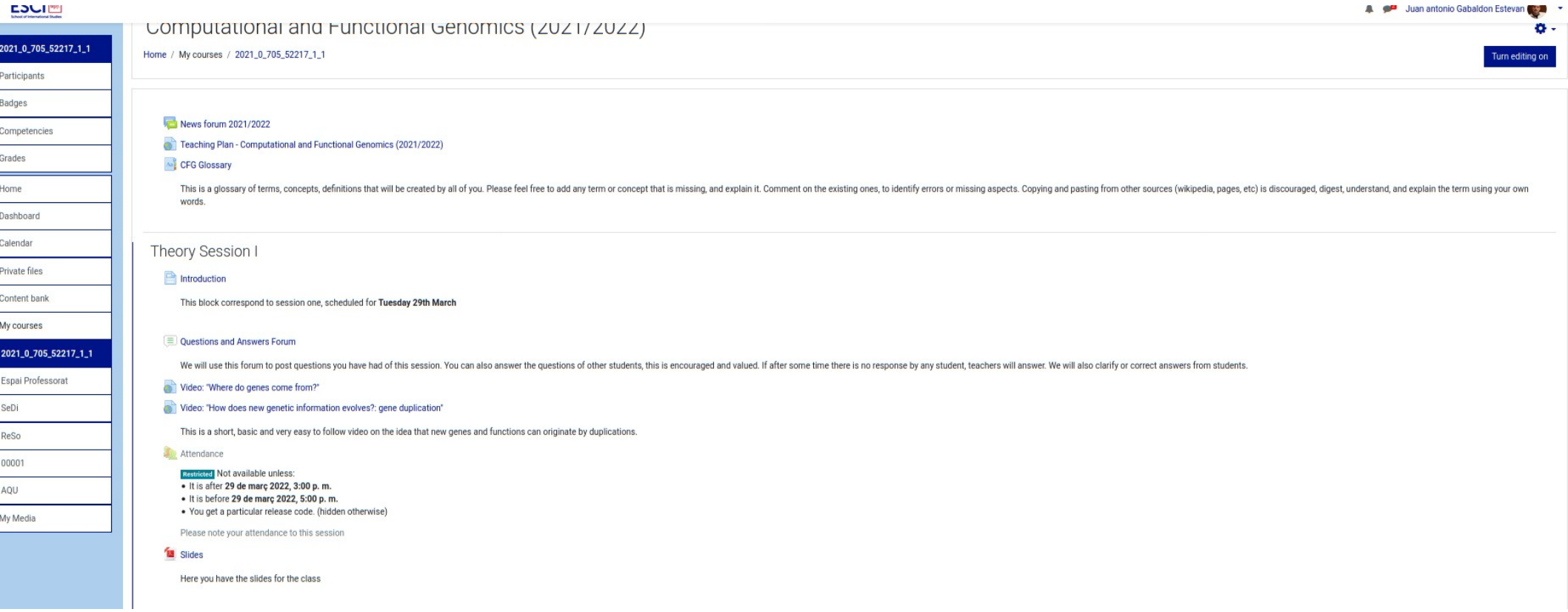
- Journal Club. Read, discuss, synthesize and present a journal article. (in groups of 4-5).

- Participation: attendance, dedication, questions answered during the course

Assessment elements	Grouping		Weight (%)
	Indiv	Group (#)	
Theoretical final exam	x		60%
Final project		x	15%
Seminar		x	10%
Participation in activities	x		15%

Commitment

- Attendance
- Effort
- Interest
- Ask when in doubt
- Dedicate study time during the whole period
- Do not wait till the last minute


Juan antonio Gabaldon Estevan
Computational and Functional Genomics (2021/2022)
Home / My courses / 2021_0_705_52217_1_1
Turn editing on
2021_0_705_52217_1_1
Participants
Badges
Competencies
Grades
Home
Dashboard
Calendar
Private files
Content bank
My courses
2021_0_705_52217_1_1
Espai Professorat
SeDi
ReSo
00001
AQU
My Media
News forum 2021/2022
Teaching Plan - Computational and Functional Genomics (2021/2022)
CFG Glossary
This is a glossary of terms, concepts, definitions that will be created by all of you. Please feel free to add any term or concept that is missing, and explain it. Comment on the existing ones, to identify errors or missing aspects. Copying and pasting from other sources (wikipedia, pages, etc) is discouraged, digest, understand, and explain the term using your own words.
Theory Session I
Introduction
This block correspond to session one, scheduled for Tuesday 29th March
Questions and Answers Forum
We will use this forum to post questions you have had of this session. You can also answer the questions of other students, this is encouraged and valued. If after some time there is no response by any student, teachers will answer. We will also clarify or correct answers from students.
Video: "Where do genes come from?"
Video: "How does new genetic information evolves?: gene duplication"
This is a short, basic and very easy to follow video on the idea that new genes and functions can originate by duplications.
Attendance
Restricted Not available unless:
It is after 29 de març 2022, 3:00 p. m.
It is before 29 de març 2022, 5:00 p. m.
You get a particular release code. (hidden otherwise)
Please note your attendance to this session
Slides
Here you have the slides for the class

Comparative and Functional Genomics

Session 1
Genes and their functions

ESCI upf.
School of International Studies

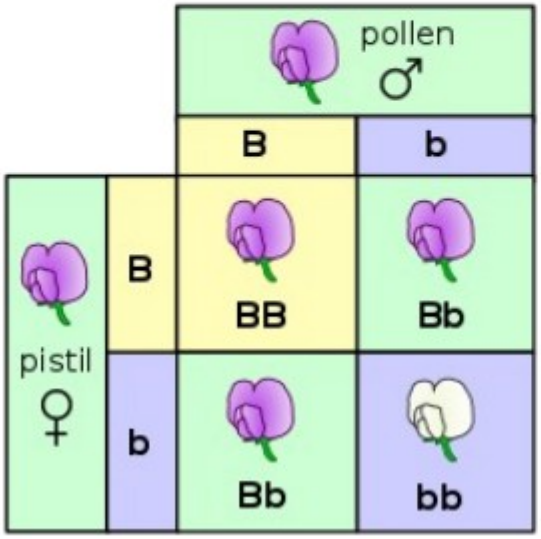
- **Homology, Paralogy and Orthology**

Genes and their Functions

- Protein-coding and non-coding genes.
- Gene structure and expression.
- Functional roles of genes.
- Relationships between sequence, structure, and function and their evolution.
- Homology based functional inference.
- Protein domains and domain shuffling.
- Prediction of protein subcellular localization.
- De novo origin of genes.
- Pathways.
- The Gene Ontology.
- Enrichment analyses.

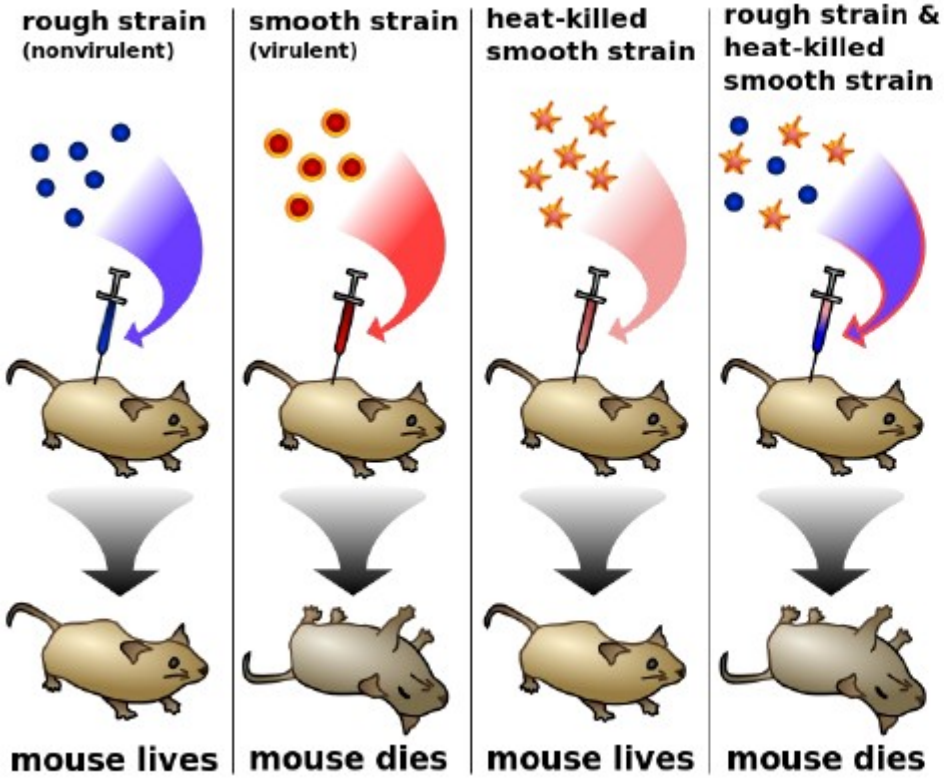
Protein coding and non-coding genes: what is a gene?

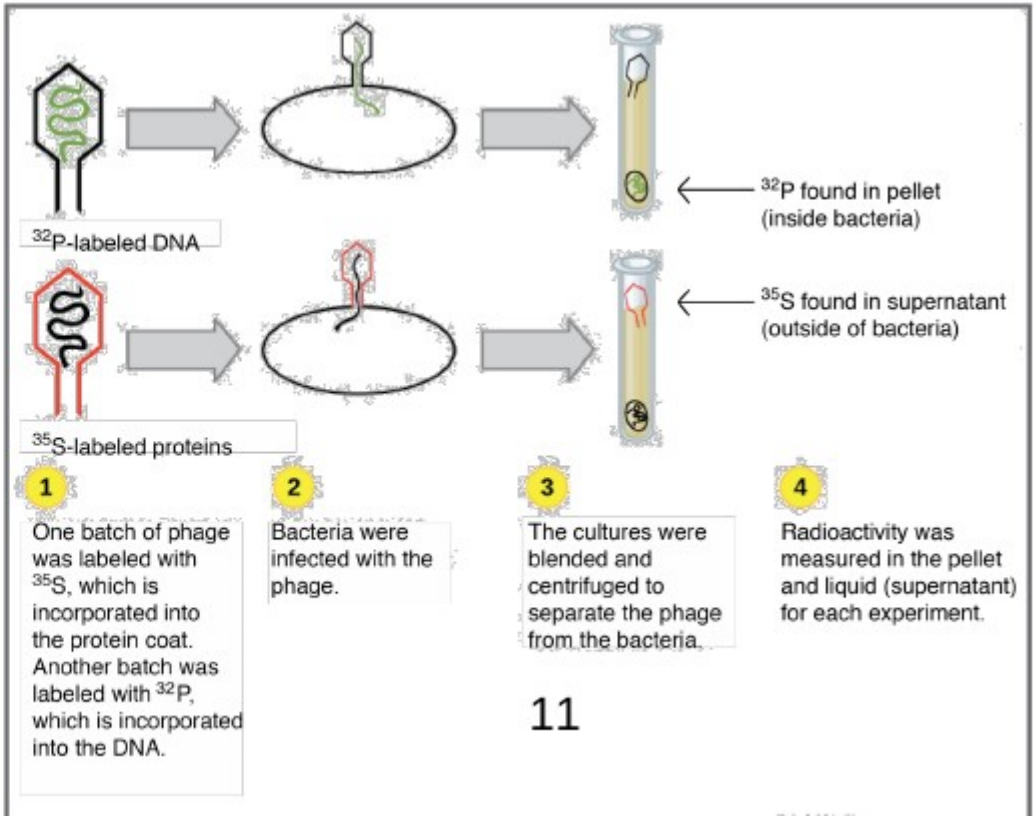
The existence of **discrete inheritable** units was first suggested by Gregor Mendel (1822–1884).



Protein coding and non-coding genes: what is a gene?

Avery–MacLeod–McCarty experiment with *Streptococcus pneumoniae* strains and Hersey-Chase experiments with bacteriophages (1940s) demonstrated that **Nucleic Acids (DNA/RNA) are the substrate of hereditary information**

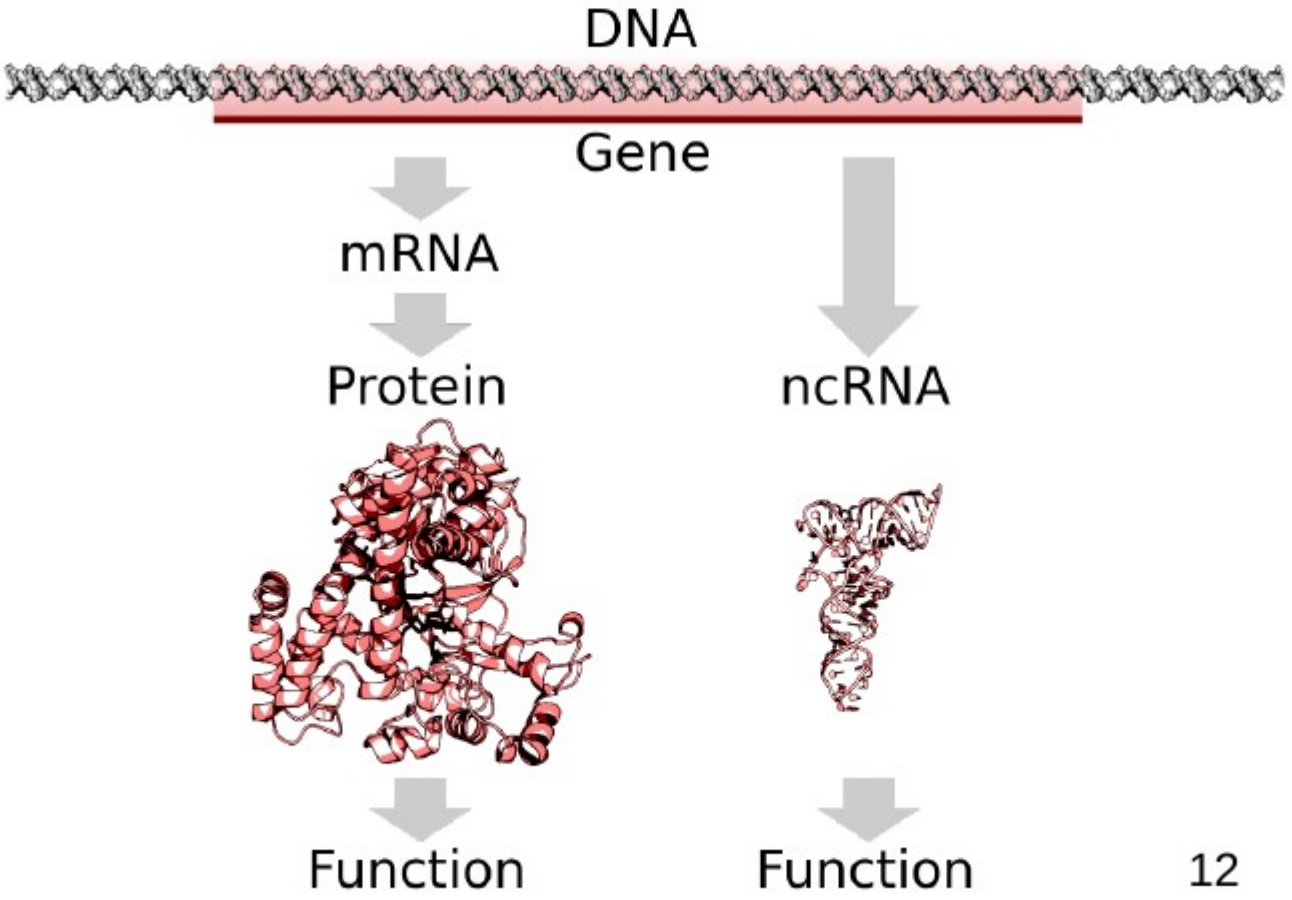




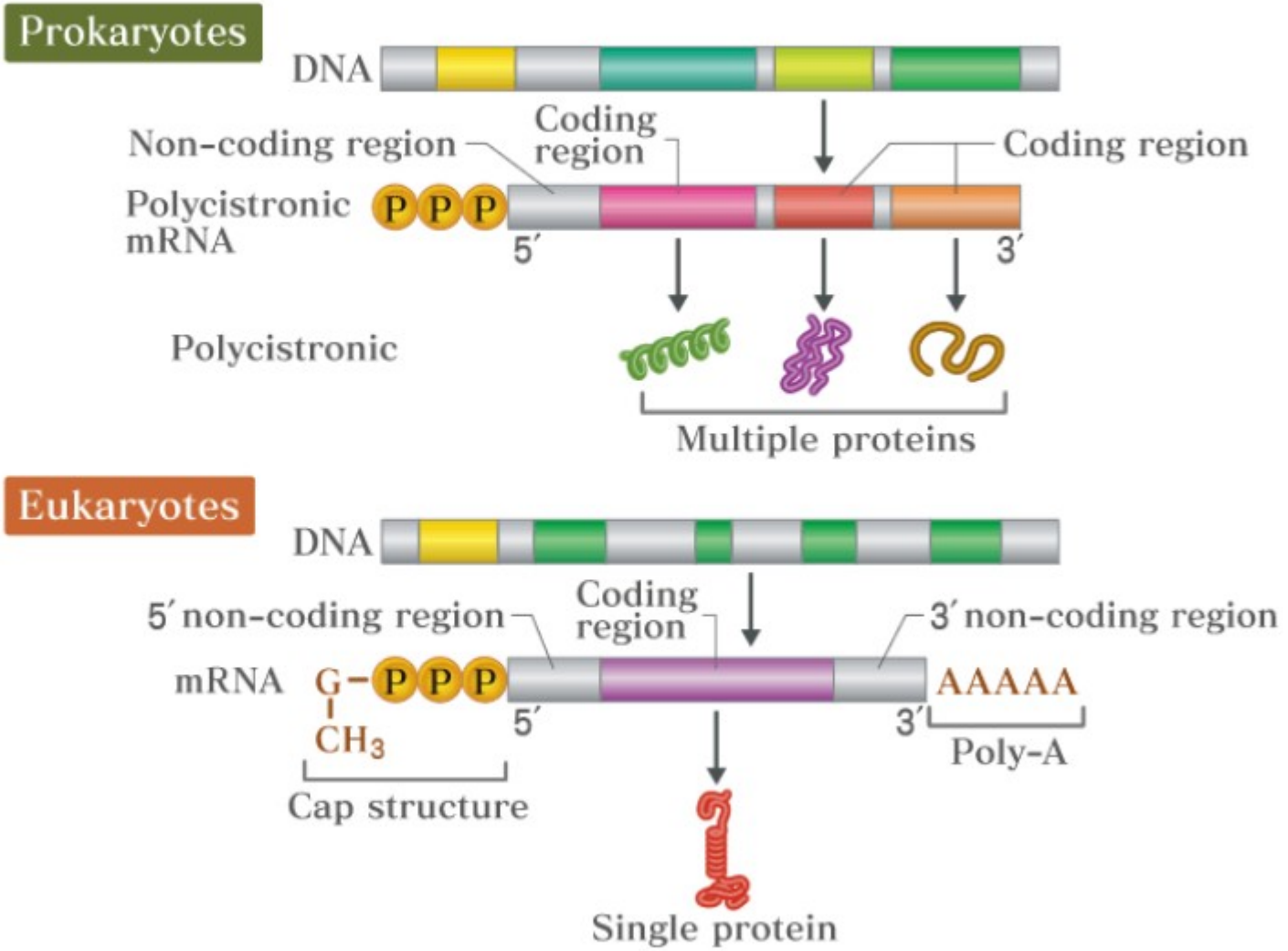
Protein coding and non-coding genes

Modern gene definition:

A sequence of DNA or RNA which **codes** for a molecule that has a function



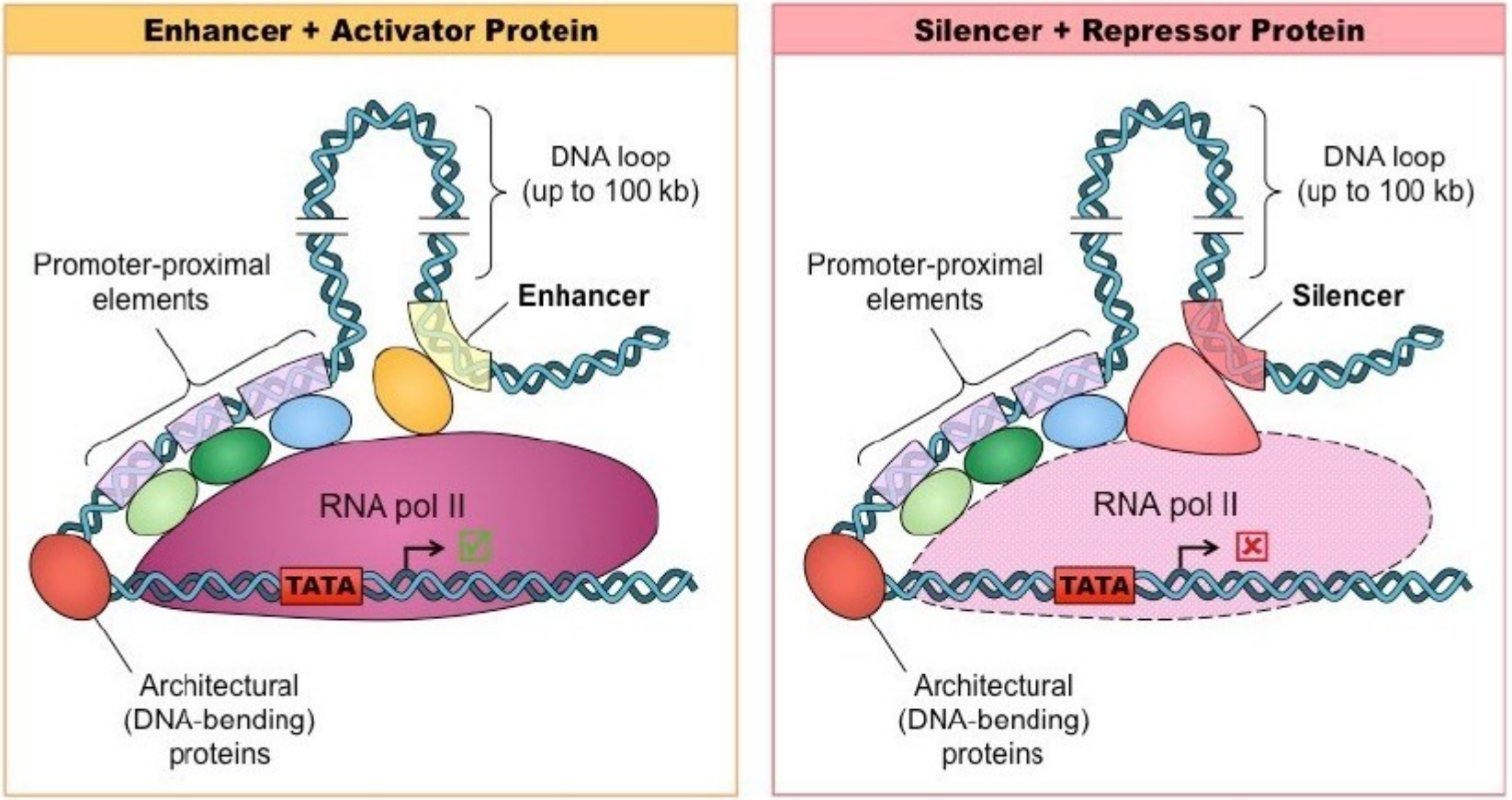
Gene structure and expression.



Introns/Exons
Splicing
5', 3' UTR

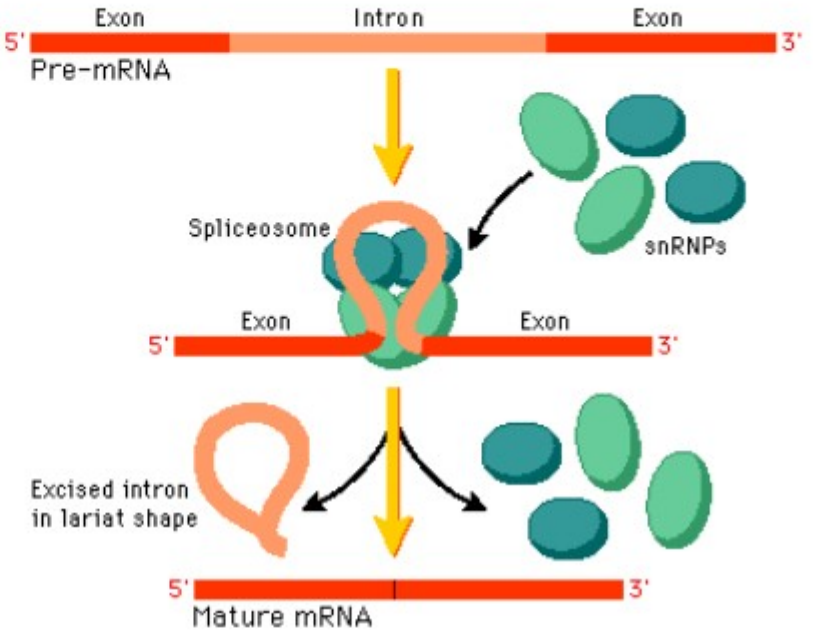
Gene structure and expression.

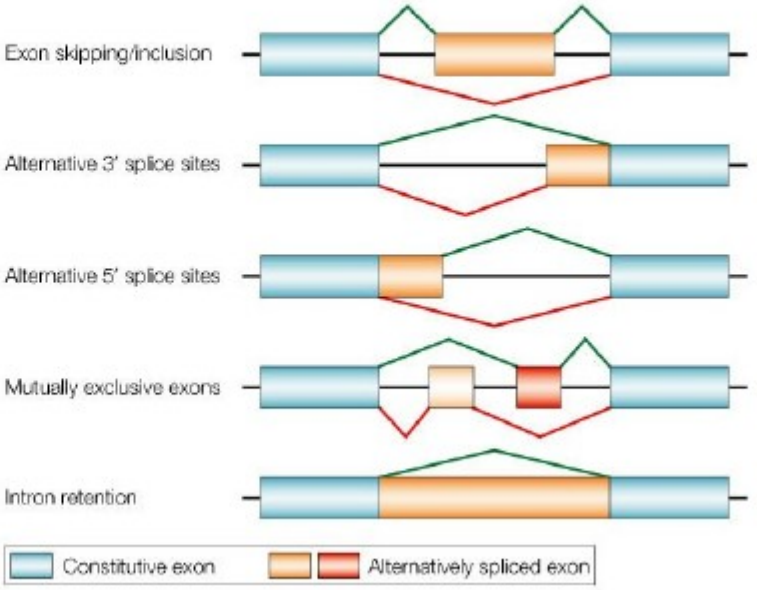
Proteins: Transcription factors (Activators, Repressors)
Sequences: Promoters, Enhancers, TFBinding sites



Gene structure and expression.

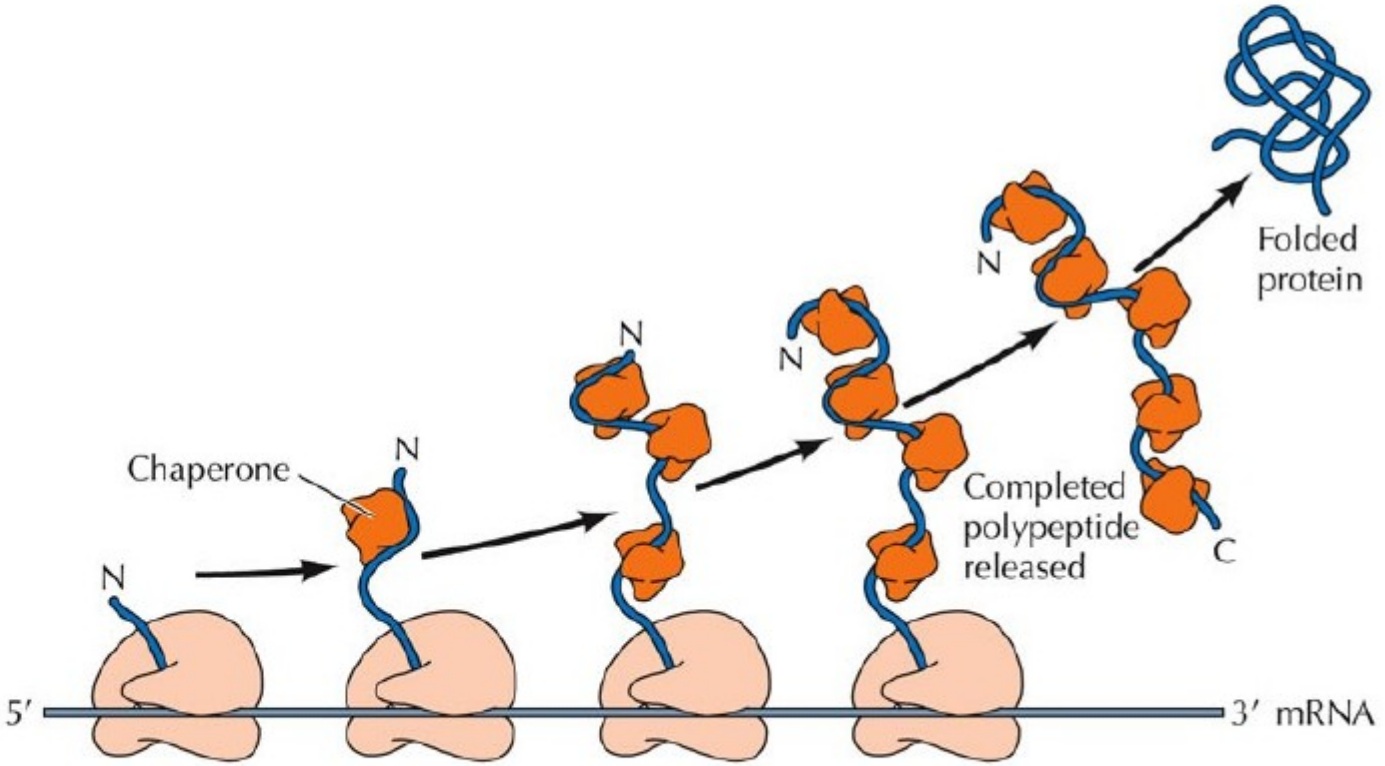
Splicing





Gene structure and expression.

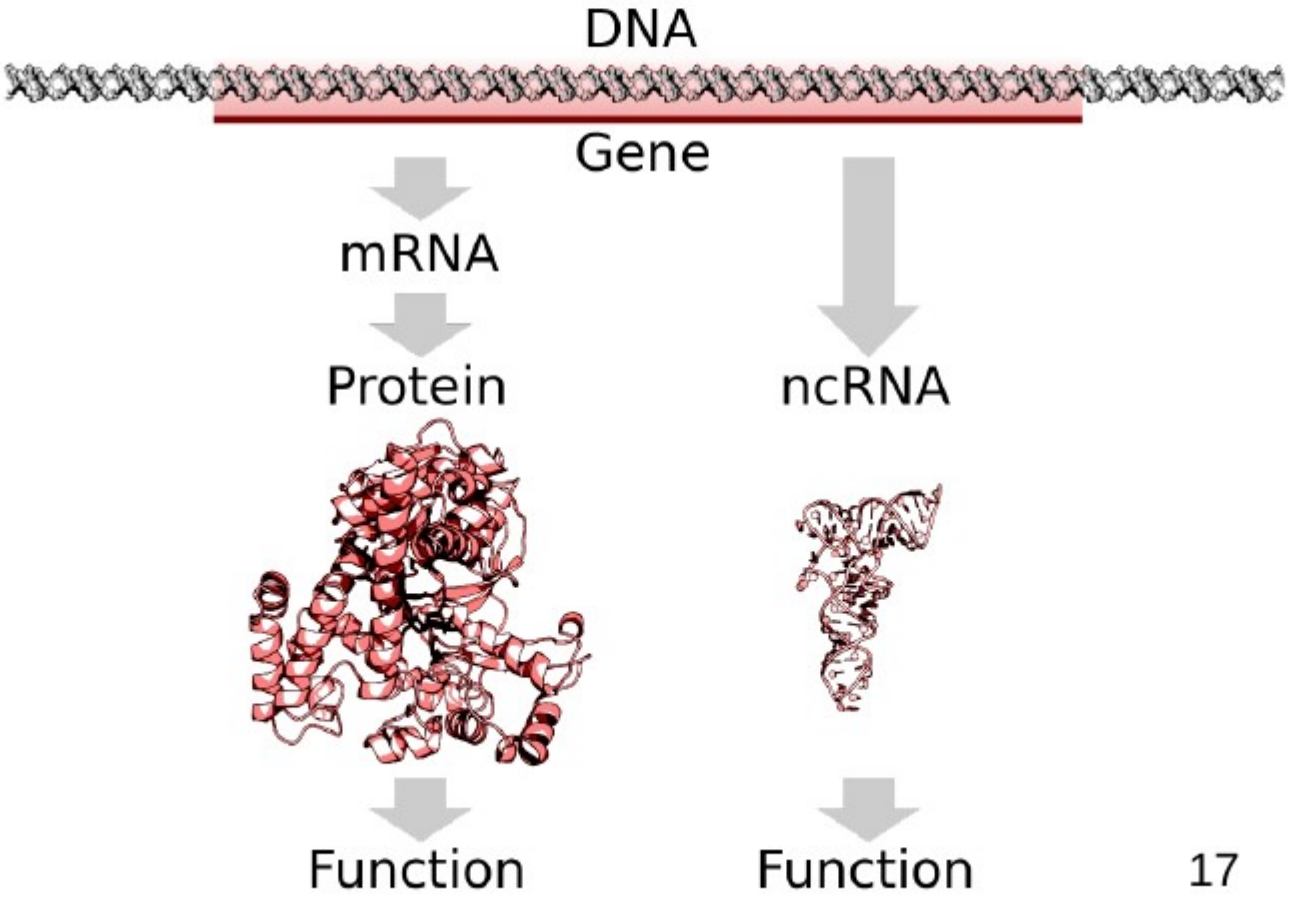
Transcript nuclear export (in eukaryotes)
And protein translation and folding (protein coding genes)



Protein coding and non-coding genes

Modern gene definition:

A sequence of DNA or RNA which **codes** for a molecule that has a function



Functional roles of genes.

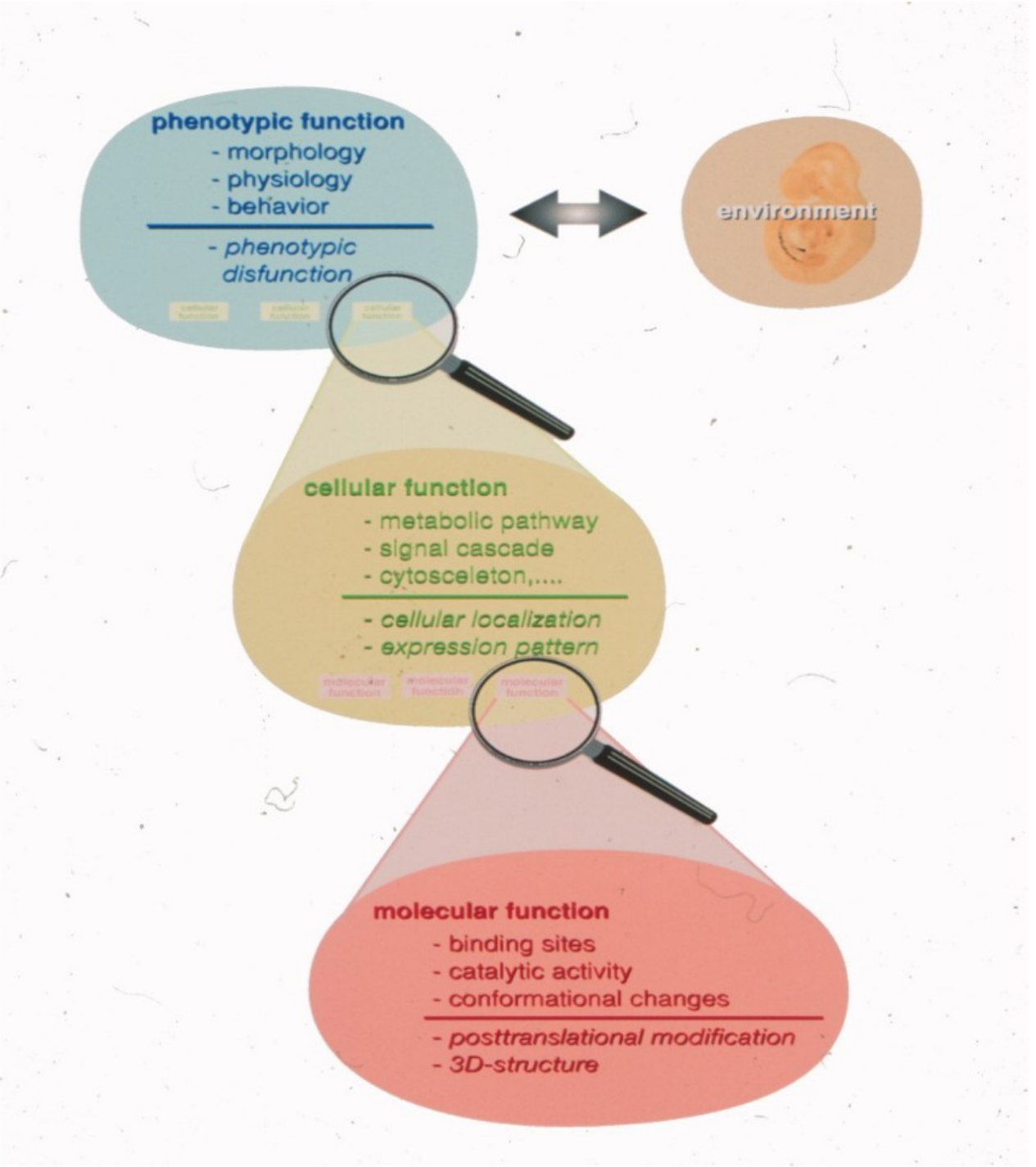
A sequence of DNA or RNA which **codes** for a molecule that has a **function**

Function?

- Structural (i.e. Actin), Catalitic (e.g. Glucogen synthase), Regulatory (i.e. Transcription factor), several functions, etc.

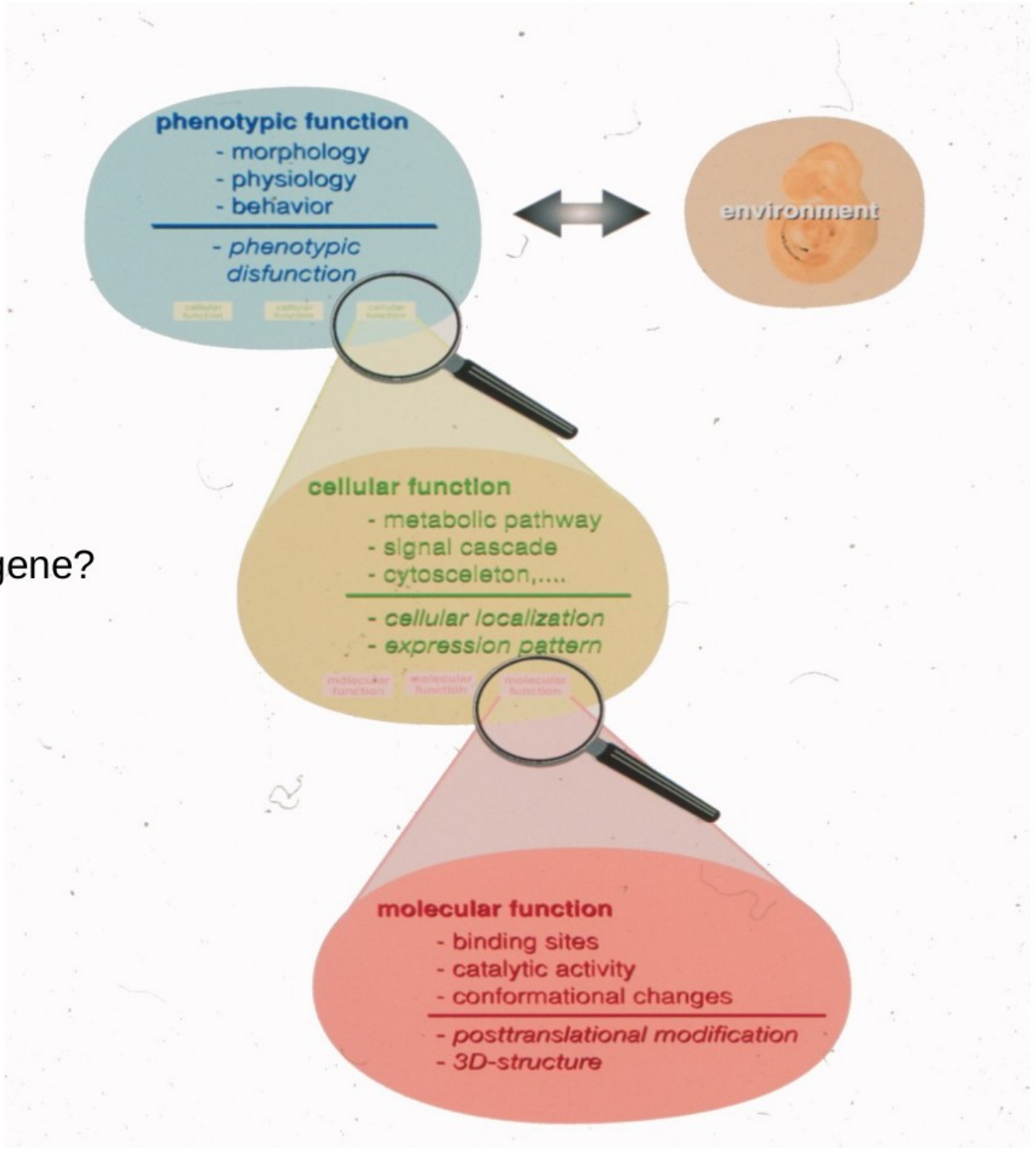
- Essential/non-essential, Constitutive (house-keeping), etc.

Functional roles of genes.



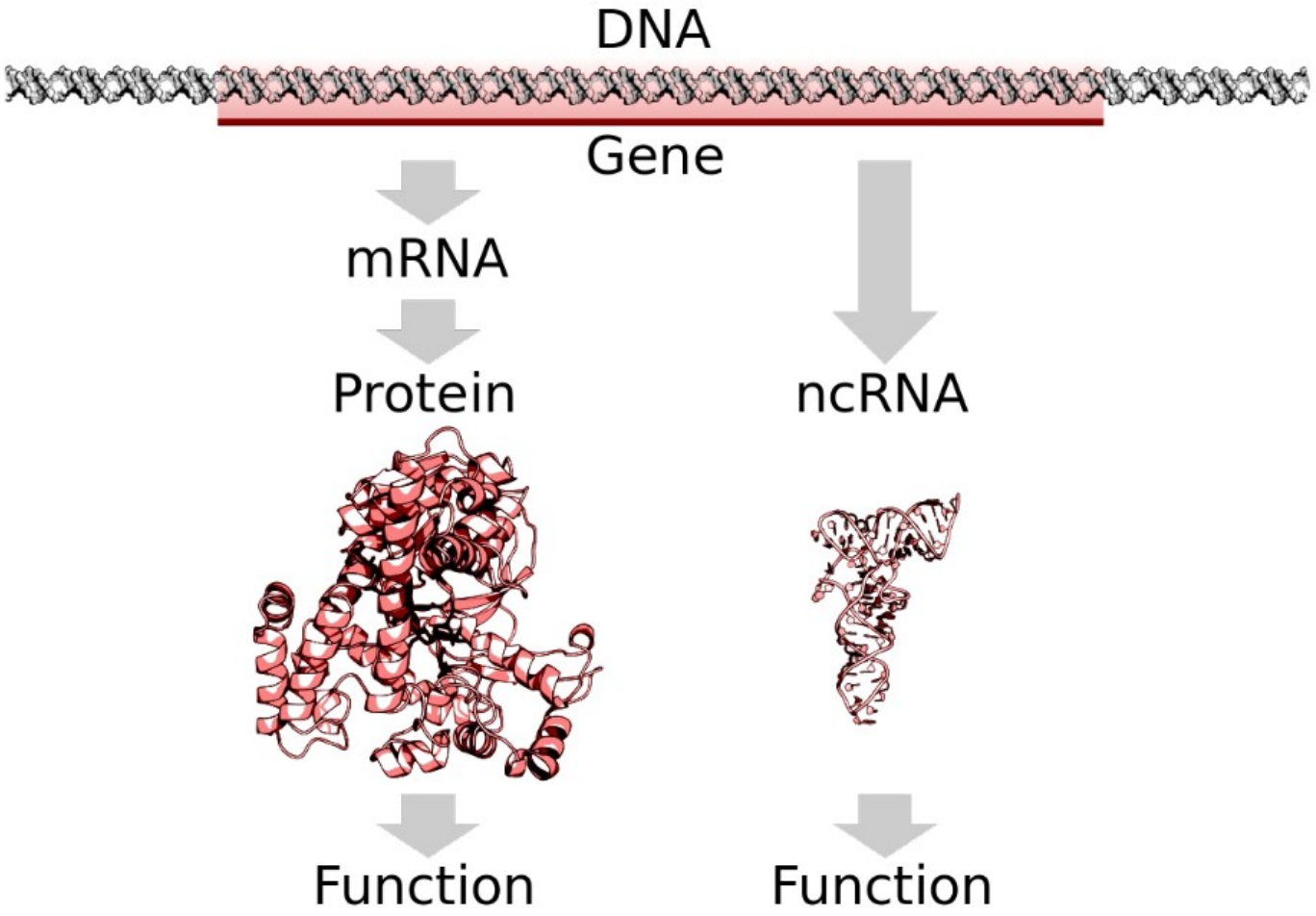
Relationships between sequence, structure, and function and their evolution.

What determines the function of a gene?



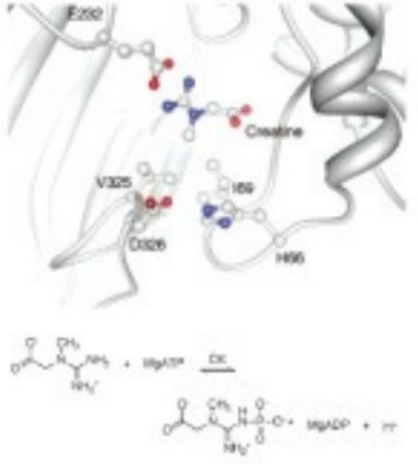
Relationships between sequence, structure, and function and their evolution.

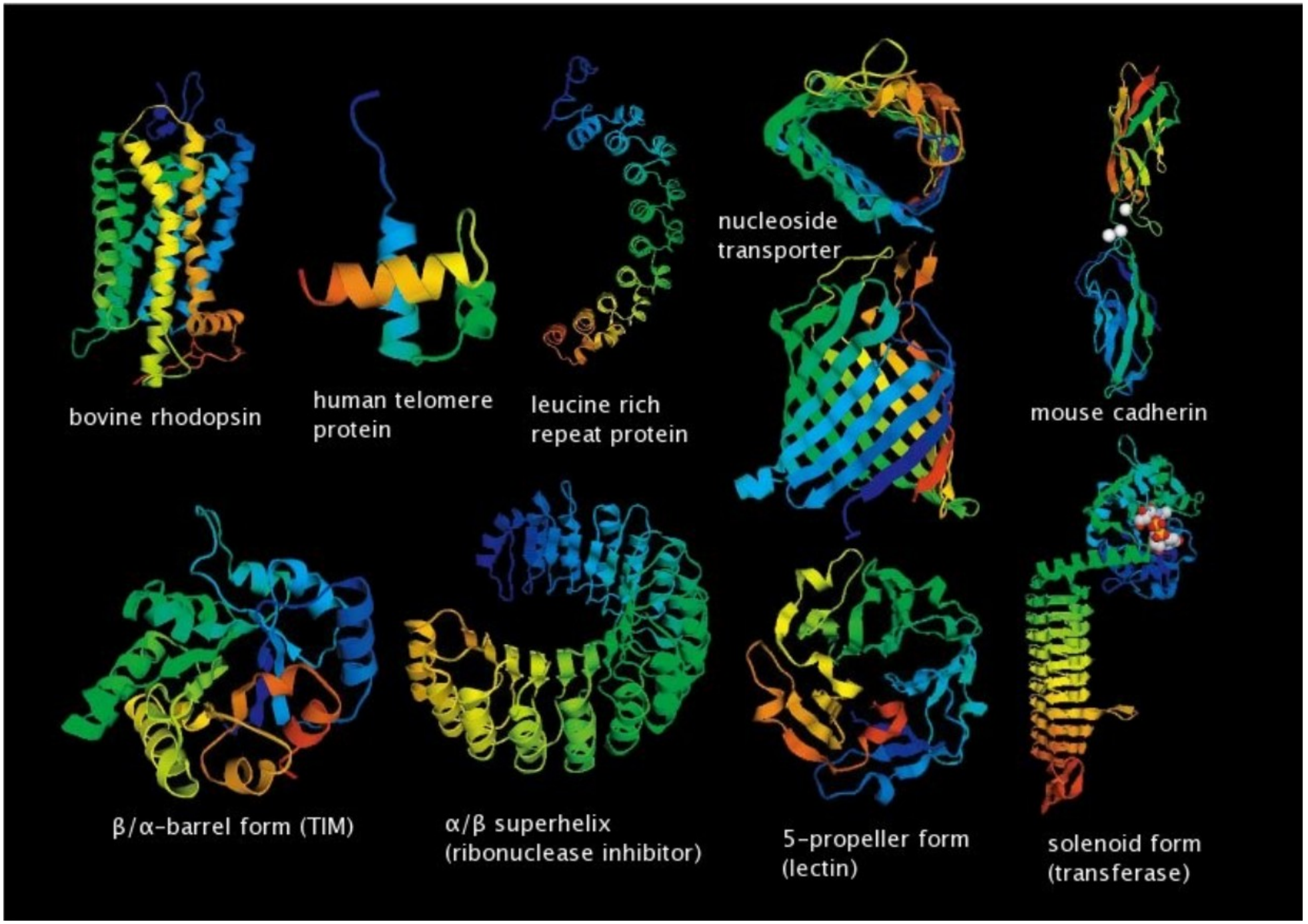
What determines the function of a gene?

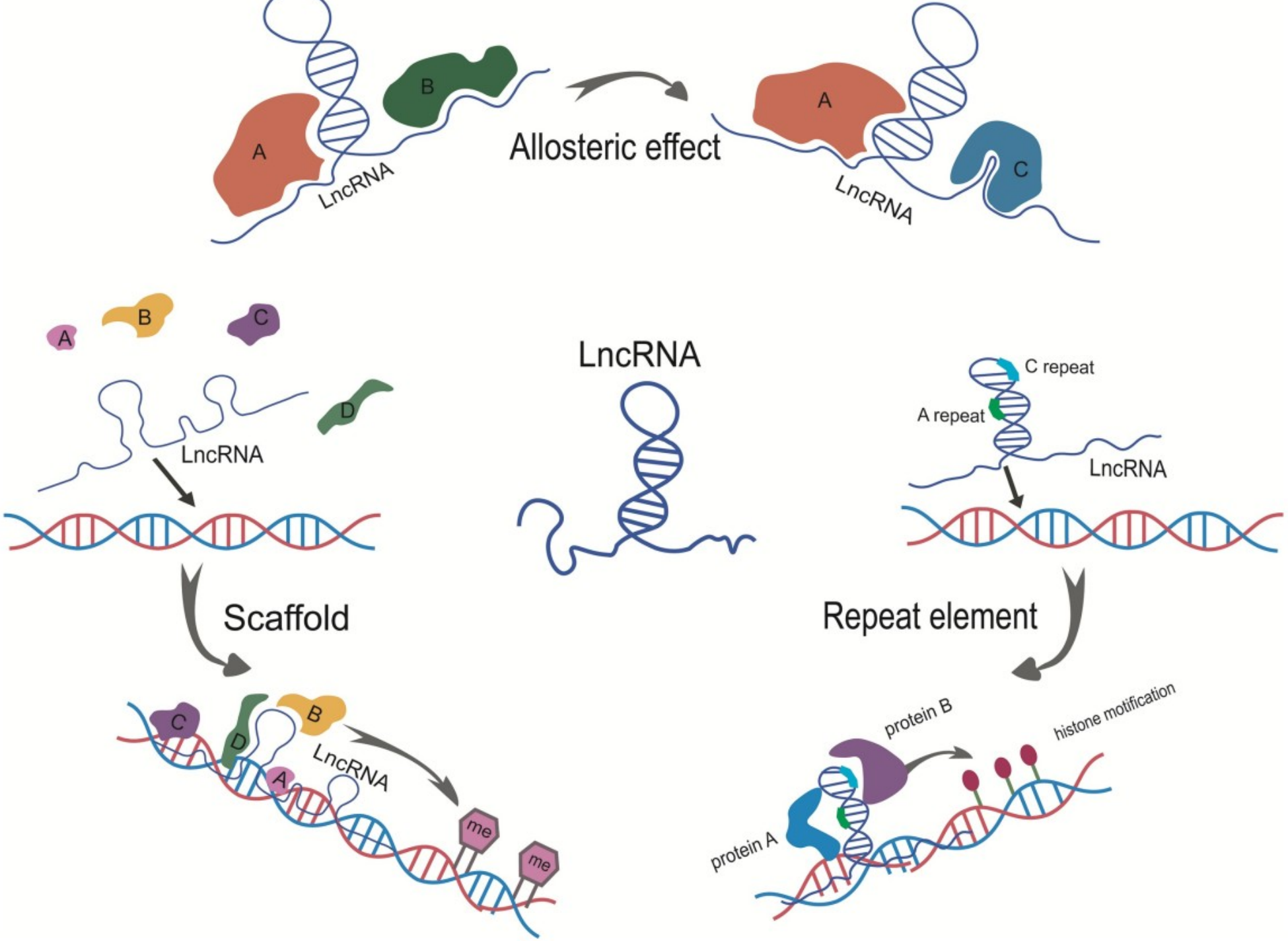


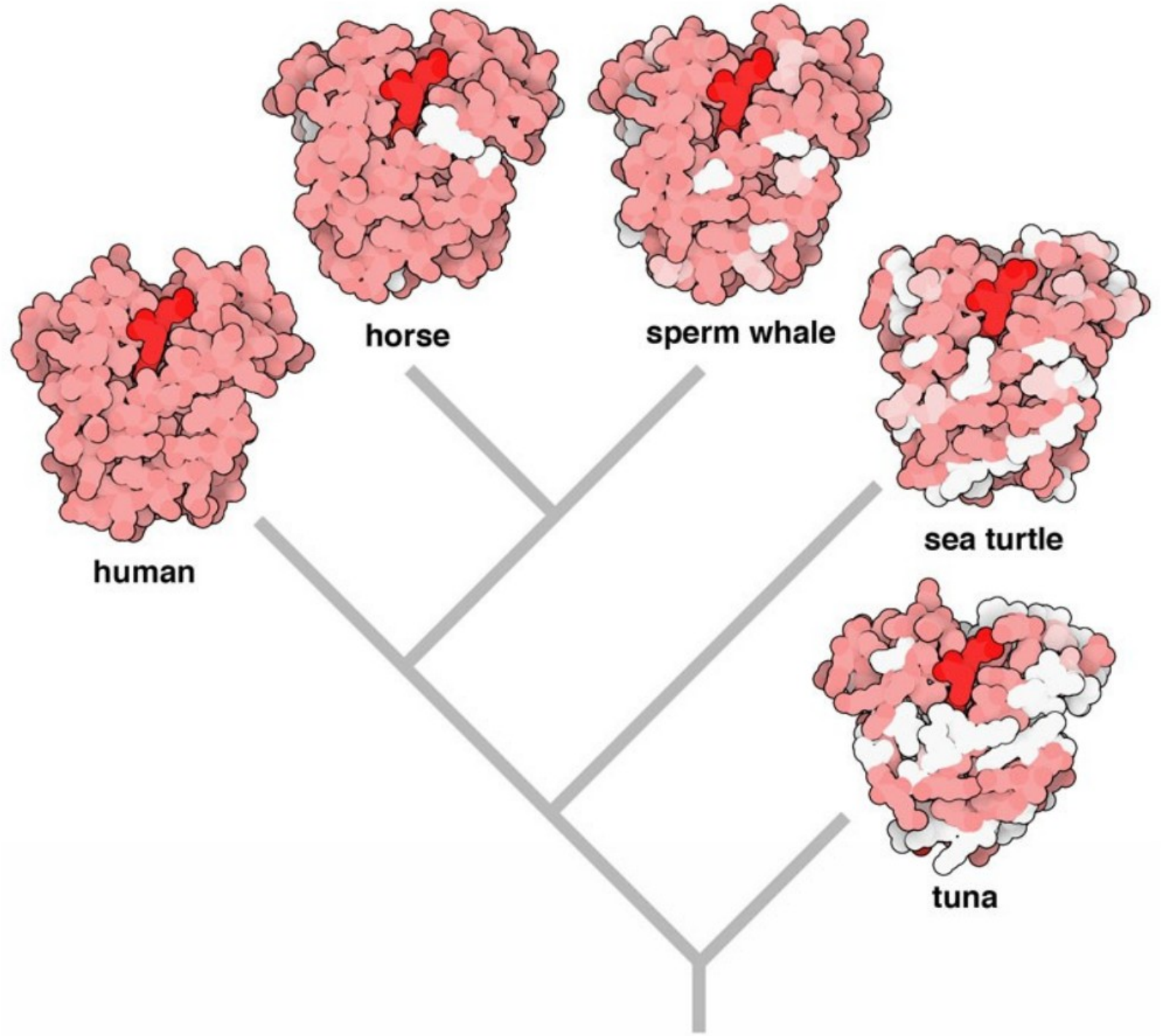
Sequence → Structure → Function

MPFGNTHNKFKL
NYKPEEEYPDLSK
HNNHMAKVLTLE
LYKKLRDKETPSGF
TVDDVIQTGVDNP
GHPFIMTVGCVAG
DEESYEVFKELFDPI
ISDRHGGYKPTD...




bovine rhodopsin
human telomere protein
leucine rich repeat protein
nucleoside transporter
mouse cadherin
β/α-barrel form (TIM)
α/β superhelix (ribonuclease inhibitor)
5-propeller form (lectin)
solenoid form (transferase)


A
B
LncRNA
Allosteric effect
A
C
LncRNA
A
B
C
D
LncRNA
Scaffold
LncRNA
C repeat
A repeat
LncRNA
Repeat element
C
D
B
A
LncRNA
me
me
protein B
histone motification
protein A


horse
sperm whale
human
sea turtle
tuna

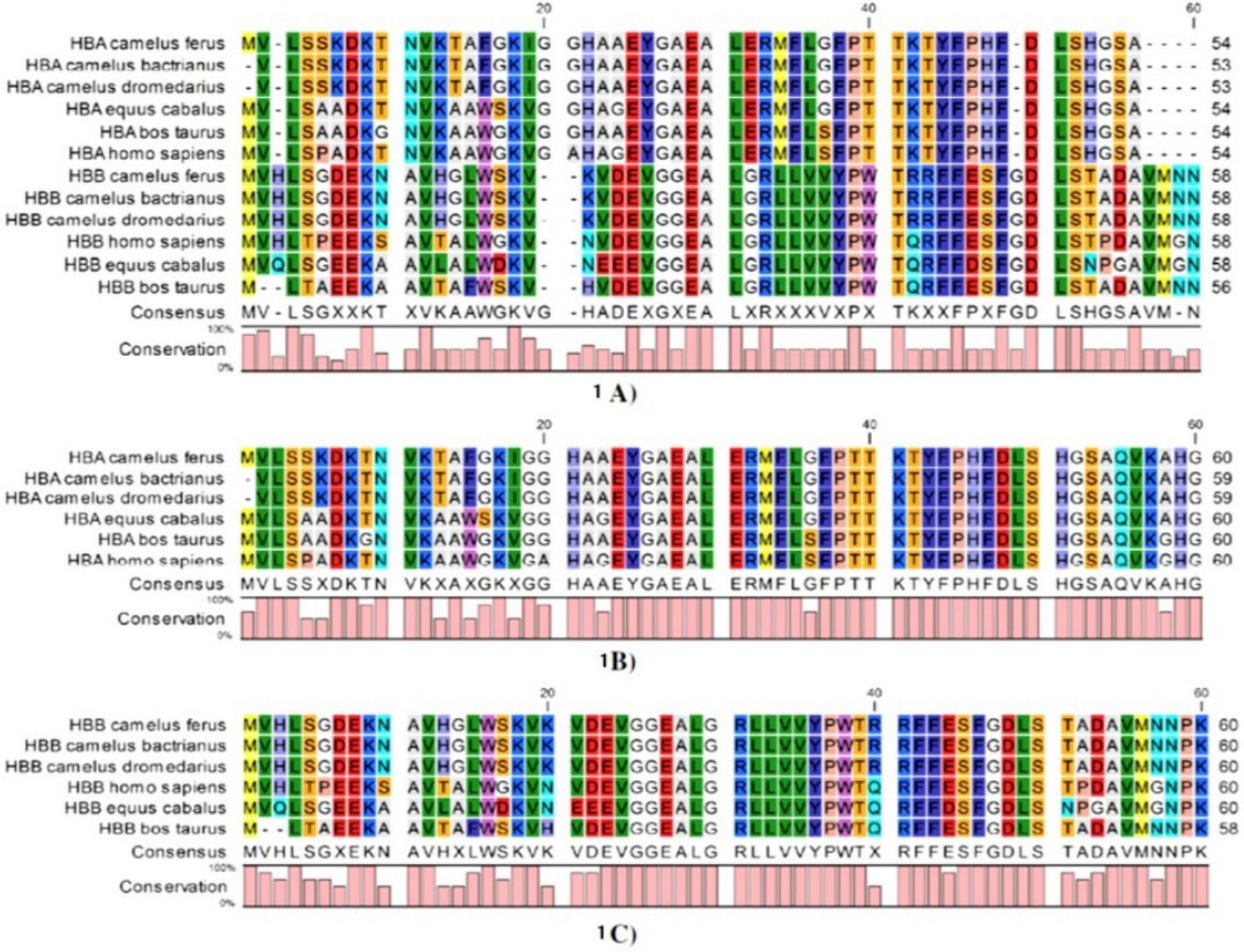
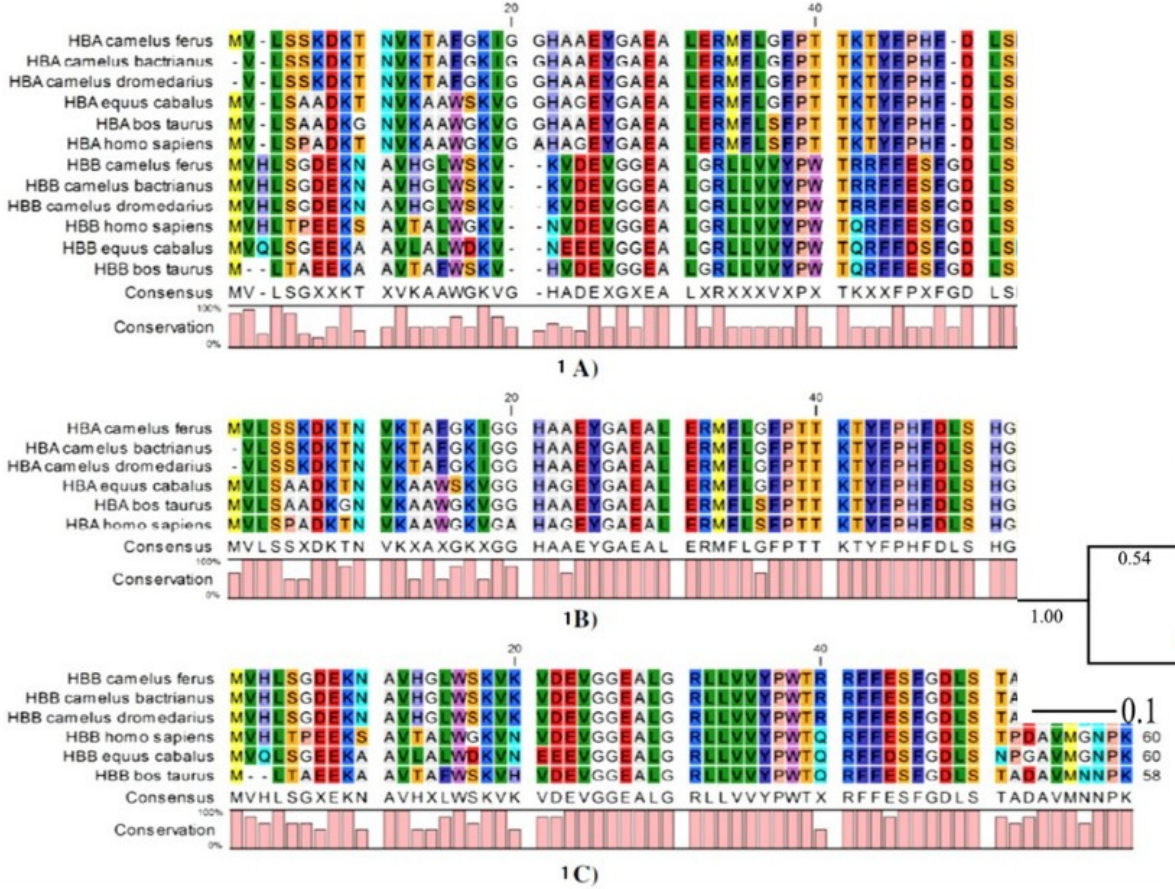
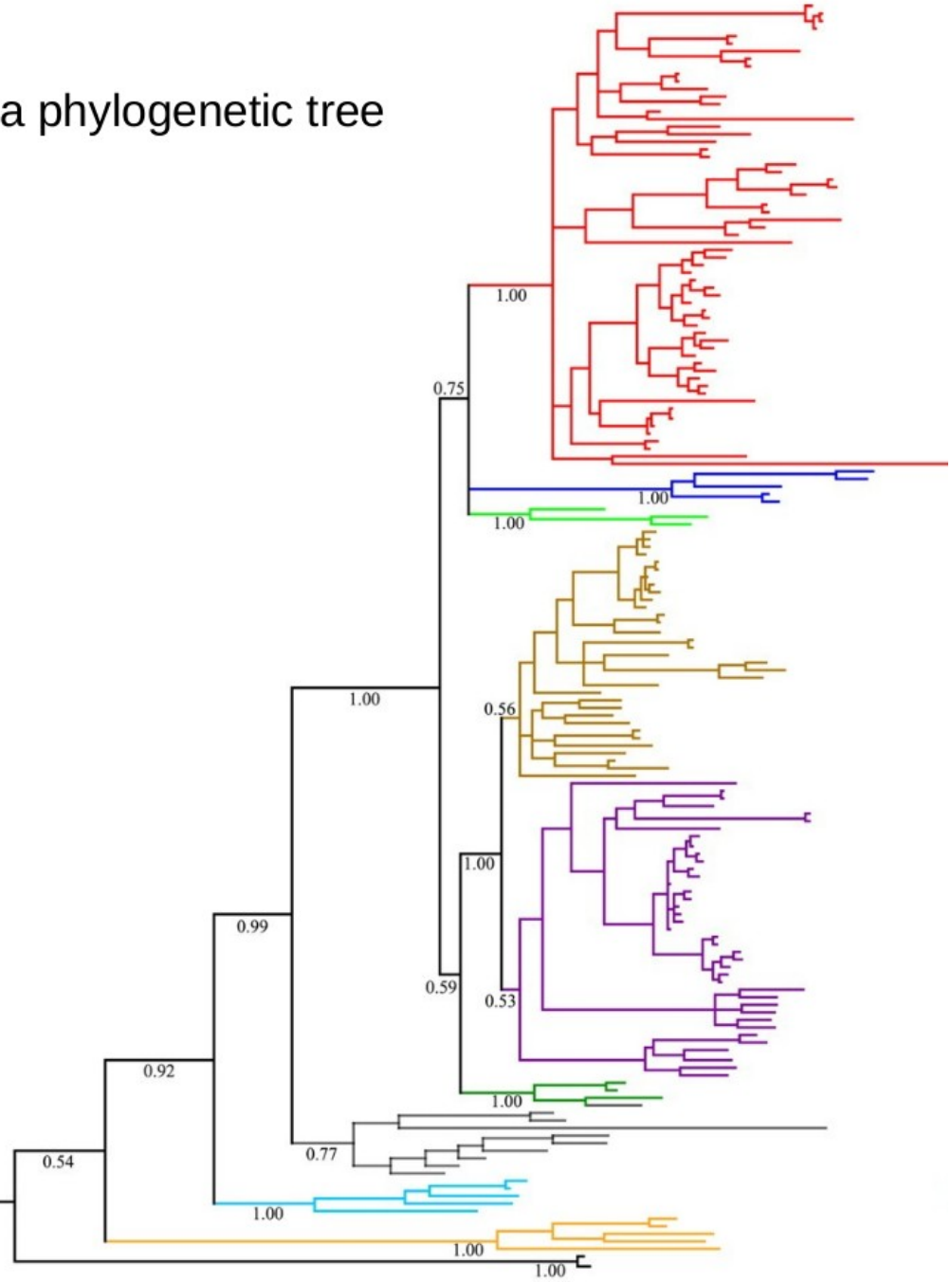


Figure 1 Multiple sequence alignment of HB subunits. 1A) α and ß chains of hemoglobin were aligned between human, domestic one and two-humped camel, wild two humped camel, cow and horse. 1B) A chain of hemoglobin was aligned between mentioned species. 1C) B chain of hemoglobin was aligned between mentioned species. Identical gaps and conserved sequences are indicated by dashes (-) and pink columns respectively

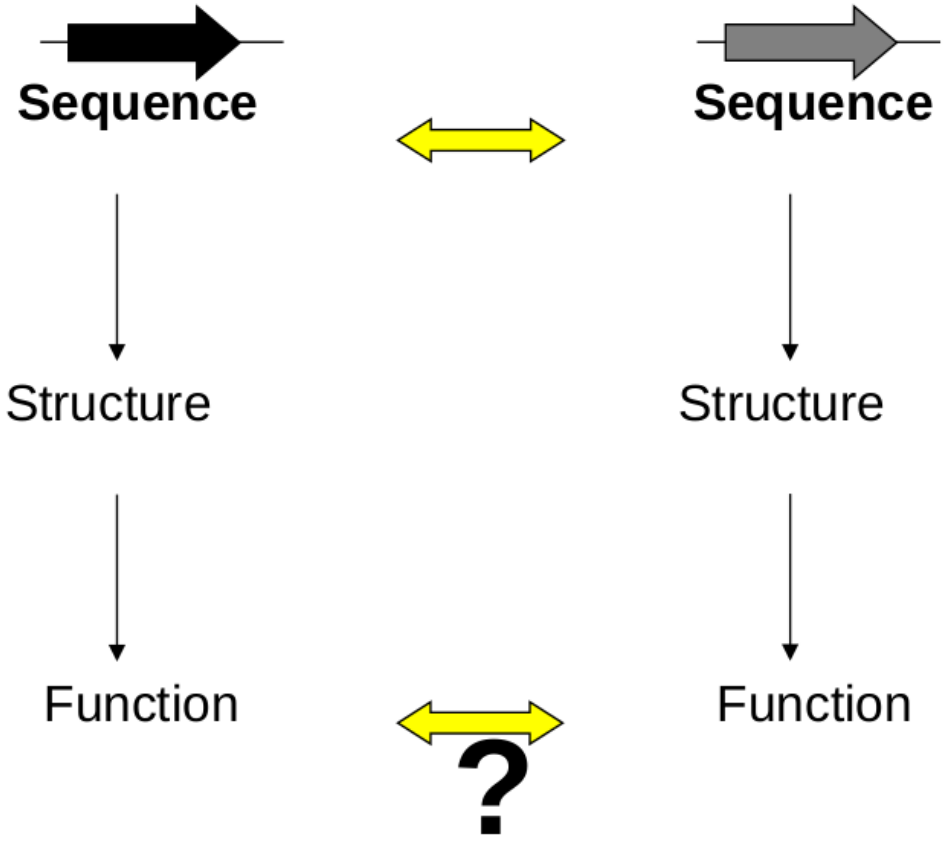
Sequence evolution can be represented with a phylogenetic tree (Gene tree)

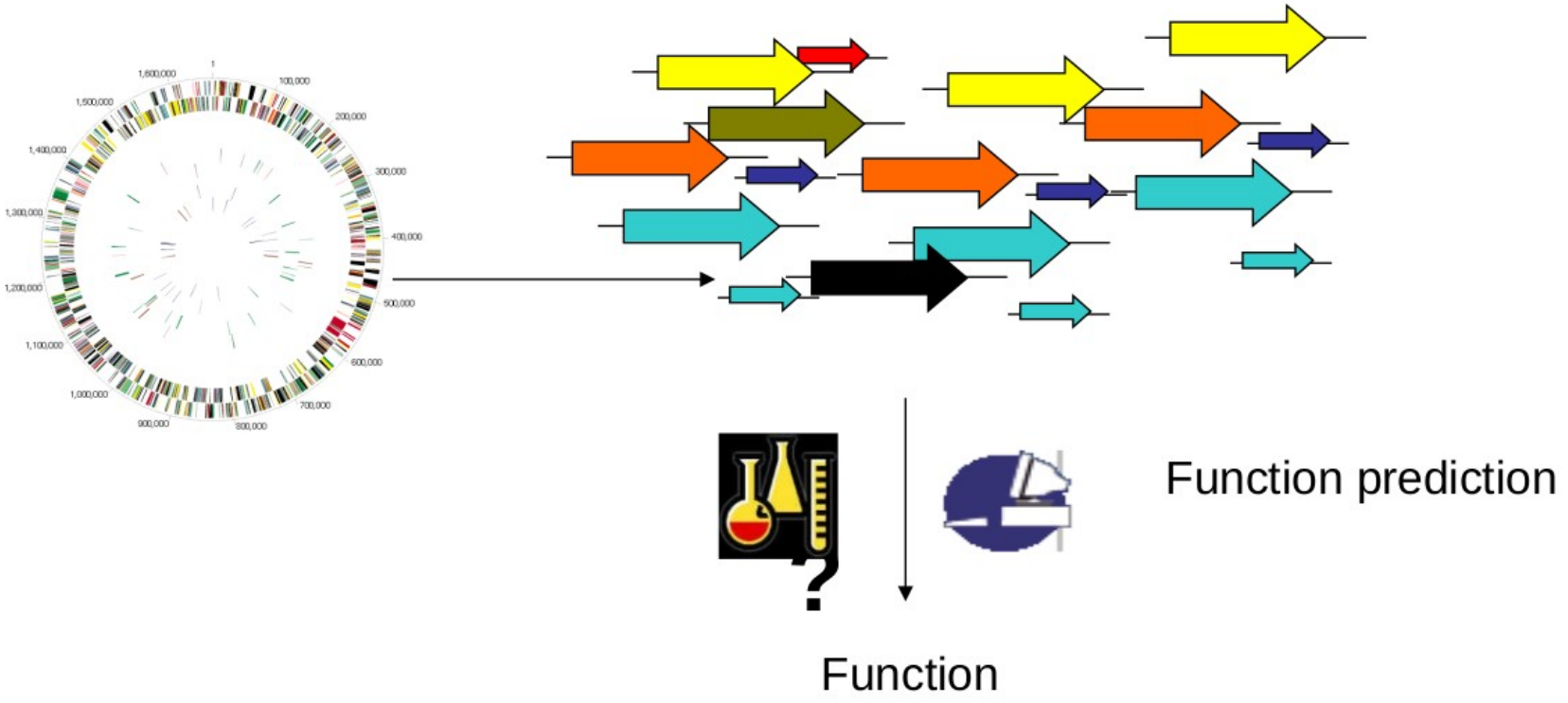




Homology based functional inference.

If sequence determines structure, which determines function, can we predict function from Sequence?





- E. coli, the most intensively studied organism: only 1924 genes (~43%) have been (partially) experimentaly characterized.

>Myprotein
MSEFATSRVESGSQQTSIHSTPIVQKLETDESPIQTKSEYTNAELPAKPIAAYWTVICLC
LMIAFGGFVFGWDTGTISGFVNQTDFKRRFGQMKSDGTYYLSDVRTGLIVGIFNIGCAFG
GLTLGRLGDMYGRRIGLMCVVLVYIVGIVIQIASSDKWYQYFIGRIISGMGVGGIAVLSP
TLISETAPKHIRGTCVSFYQLMITLGIFLGYCTNYGTKDYSNSVQWRVPLGLNFAFAIFM
IAGMLMVPESPRFLVEKGRYEDAKRSLAKSNKVTIEDPSIVAEMDTIMANVETERLAGNA
SWGELFSNKGAILPRVIMGIMIQSLQQLTGNNYFFYYGTTIFNAVGMKDSFQTSIVLGIV
NFASTFVALYTVDKFGRRKCLLGGSASMAICFVIFSTVGVTSLYPNGKDQPSSKAAGNVM
IVFTCLFIFFFAISWAPIAYVIVAESYPLRVKNRAMAIAVGANWIWGFLIGFFTPFITSA
IGFSYGYVFMGCLVFSFFYVFFFVCETKGLTLEEVNEMYVEGVKPWKSGSWISKEKRVSE
F*

Sequences producing significant alignments:

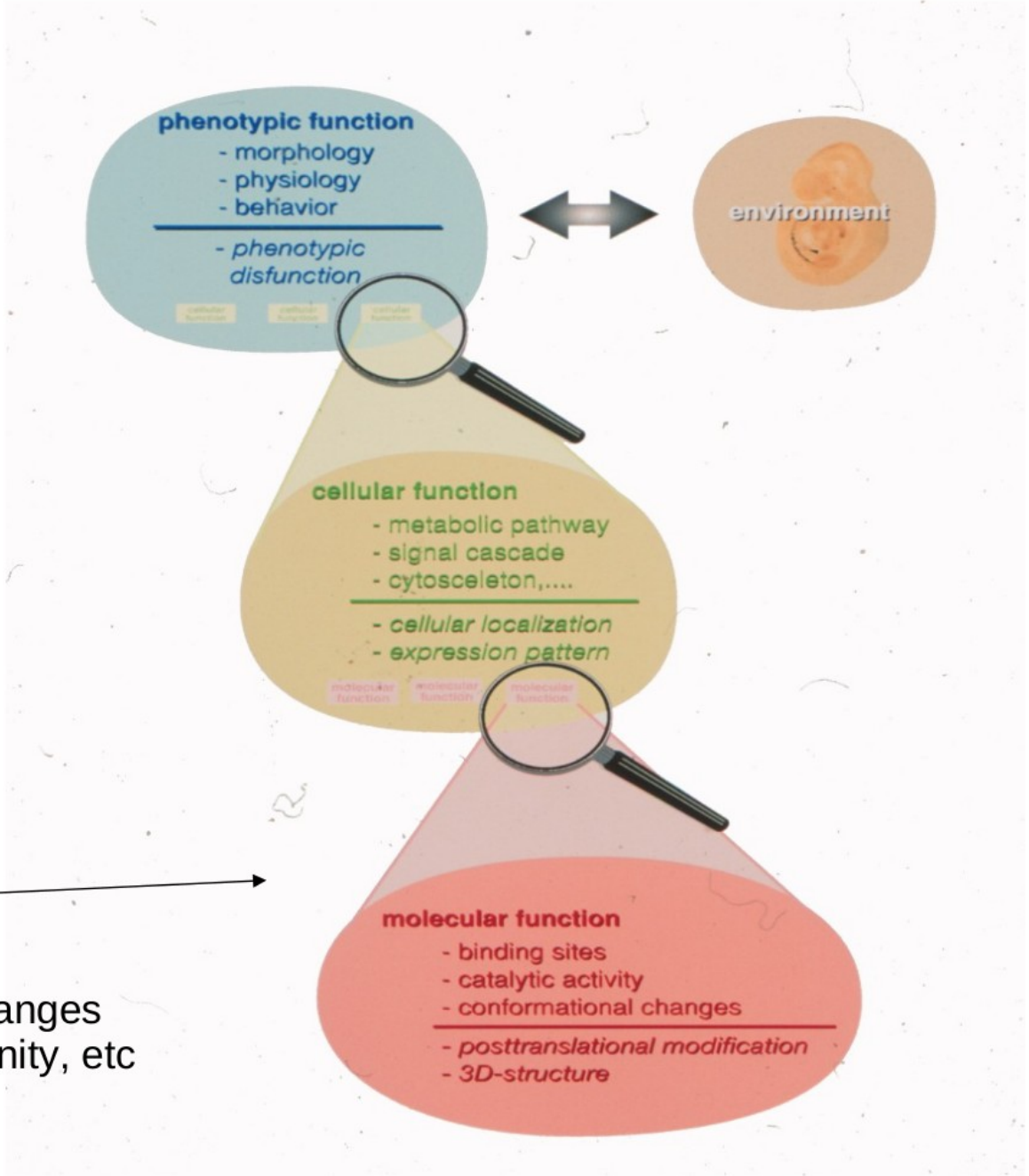
Select: All None Selected:0

Alignments | Download | GenPept | Graphics | Distance tree of results | Multiple alignment

Description	Max score	Total score	Query cover	E value	Ident	Accession
PREDICTED: hexose transporter HXT13-like [Drosophila biarmipes]	534	534	86%	0.0	53%	XP_016968485.1
MFS monosaccharide transporter [Rasamsonia emersonii CBS 393.64]	474	474	96%	1e-159	44%	XP_013332487.1
hypothetical protein PENARI_c007G04528 [Penicillium arizonense]	464	464	97%	4e-155	45%	XP_022489238.1
monosaccharide transporter [Aspergillus taichungensis]	463	463	94%	6e-155	45%	PLN75983.1
high-affinity fructose transporter ght6 [Schizosaccharomyces japonicus yFS275]	462	462	87%	9e-155	47%	XP_002175189.1
hypothetical protein PENSTE_c001G04145 [Penicillium steckii]	462	462	94%	1e-154	45%	OQE32242.1
hypothetical protein PENANT_c040G07082 [Penicillium antarcticum]	461	461	97%	3e-154	44%	OQD79997.1
general substrate transporter [Aspergillus campestris IBT 28561]	461	461	95%	5e-154	45%	PKY01318.1
MFS monosaccharide transporter, putative [Aspergillus flavus NRRL3357]	460	460	88%	1e-153	47%	XP_002384025.1
putative transporter [Aspergillus oryzae 3.042]	459	459	88%	2e-153	47%	EIT72354.1
hypothetical protein PENCOP_c008G06807 [Penicillium coprophilum]	459	459	97%	2e-153	44%	OQE38260.1
Sugar and other transporter [Aspergillus parasiticus SU-1]	458	458	88%	5e-153	47%	KJK67832.1

PREDICTED: hexose transporter HXT13-like [Drosophila biarmipes]
MFS monosaccharide transporter [Rasamsonia emersonii CBS 393.64]
hypothetical protein PENARI_c007G04528 [Penicillium arizonense]
monosaccharide transporter [Aspergillus taichungensis]
ctose transporter ght6 [Schizosaccharomyces japonicus yFS275]
hypothetical protein PENSTE_c001G04145 [Penicillium steckii]
hypothetical protein PENANT_c040G07082 [Penicillium antarcticum]
general substrate transporter [Aspergillus campestris IBT 28561]
MFS monosaccharide transporter, putative [Aspergillus flavus NRRL3357]
putative transporter [Aspergillus oryzae 3.042]
hypothetical protein PENCOP_c008G06807 [Penicillium coprophilum]
Sugar and other transporter [Aspergillus parasiticus SU-1]

eOffice Impress



Homology-based prediction
Good to predict molecular function
but not higher levels

Also, be aware that few residue changes
Can drive changes in substrate affinity, etc

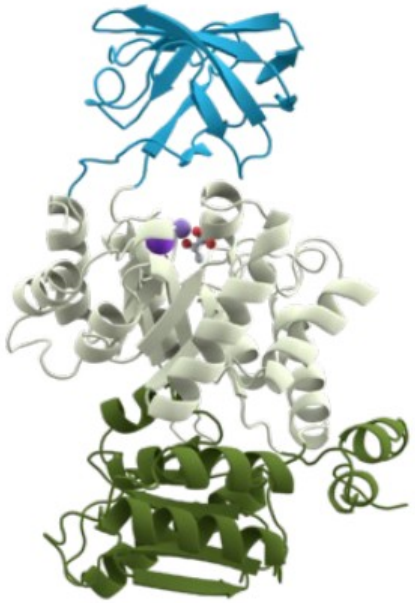
Protein domains and domain shuffling.

Protein domain:
A conserved part of a given protein sequence and (tertiary) structure that can evolve, function, and exist independently of the rest of the protein.

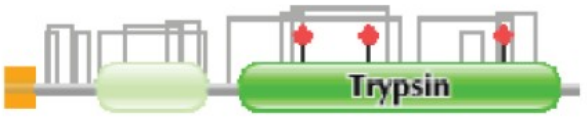
Each domain forms a compact three-dimensional structure and often can be independently stable and folded. Many proteins consist of several structural domains. And a domain may appear in a variety of different proteins.

Domains often form functional units, such as the calcium-binding EF hand domain of calmodulin.

Some domains are **promiscuous** meaning they can appear in diverse families in combination with other domains



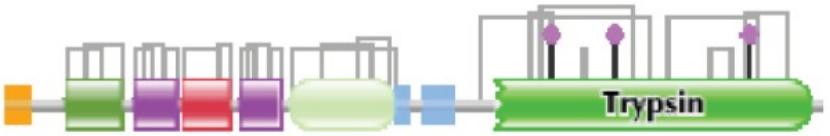
TPA_HUMAN, **DA = FN1, EGF, kringle, kringle, trypsin**



UROK_HUMAN **DA = EGF, kringle, trypsin**

HABP2_HUMAN **DA = EGF, EGF, EGF, kringle, trypsin**

FA12_HUMAN **DA = FN2, EGF, FN1, EGF, kringle, trypsin**



HGFA_HUMAN **DA = FN2, EGF, FN1, EGF, kringle, trypsin**

Note that this can confuse homology-based protein prediction (blast hits)

Protein domains and domain shuffling.

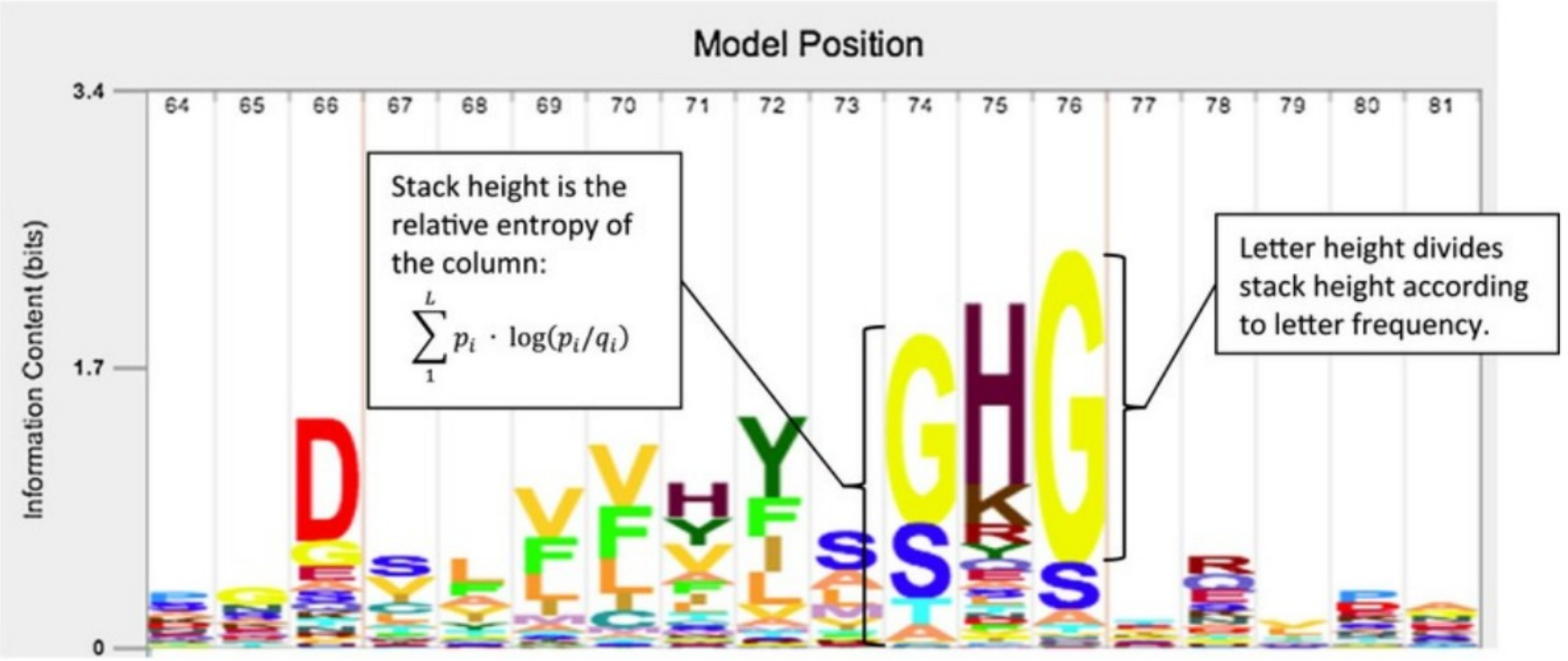
Resources:

Protein domains and domain shuffling.

Domains can be described by Hidden Markov Models (HMM), which specify the likelihood of finding a given residue in a given (relative) position

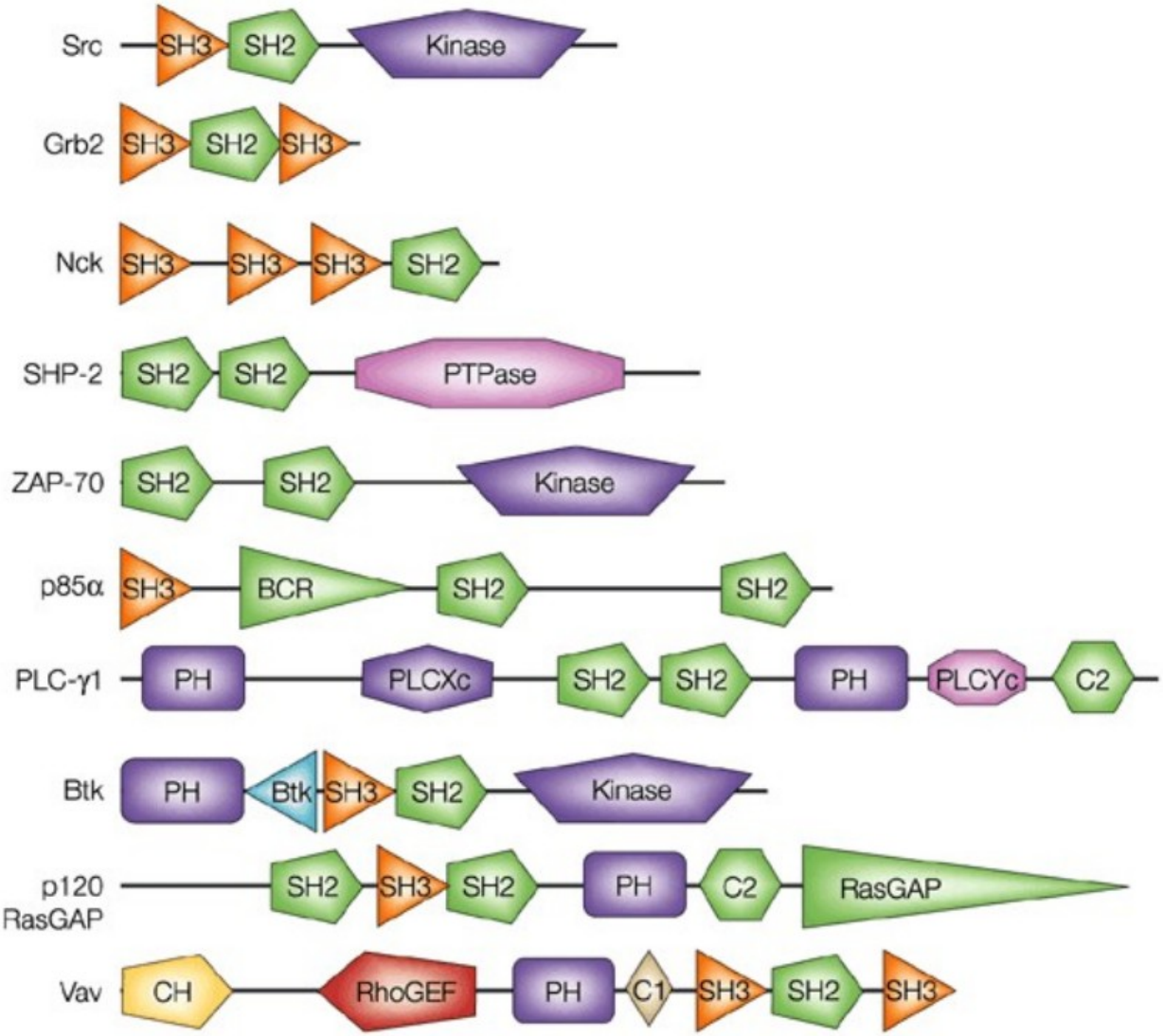
HMMs can be derived from multiple sequence alignments

HMMs can be used to detect the presence of a given domain in a sequence.
(i.e. interproScan)



Protein domains and domain shuffling.

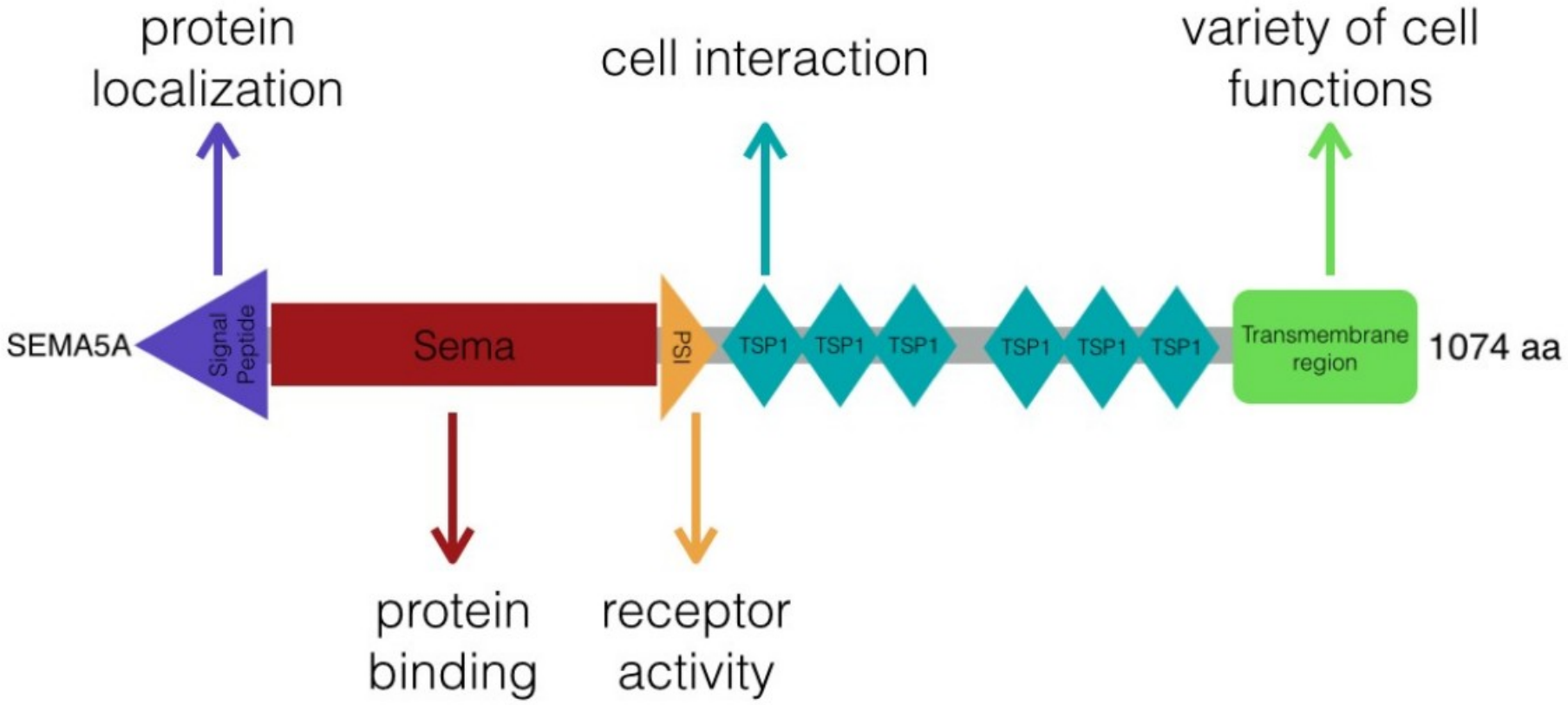
As domains tend to display a conserved function, function can be inferred from the domains present in a given protein

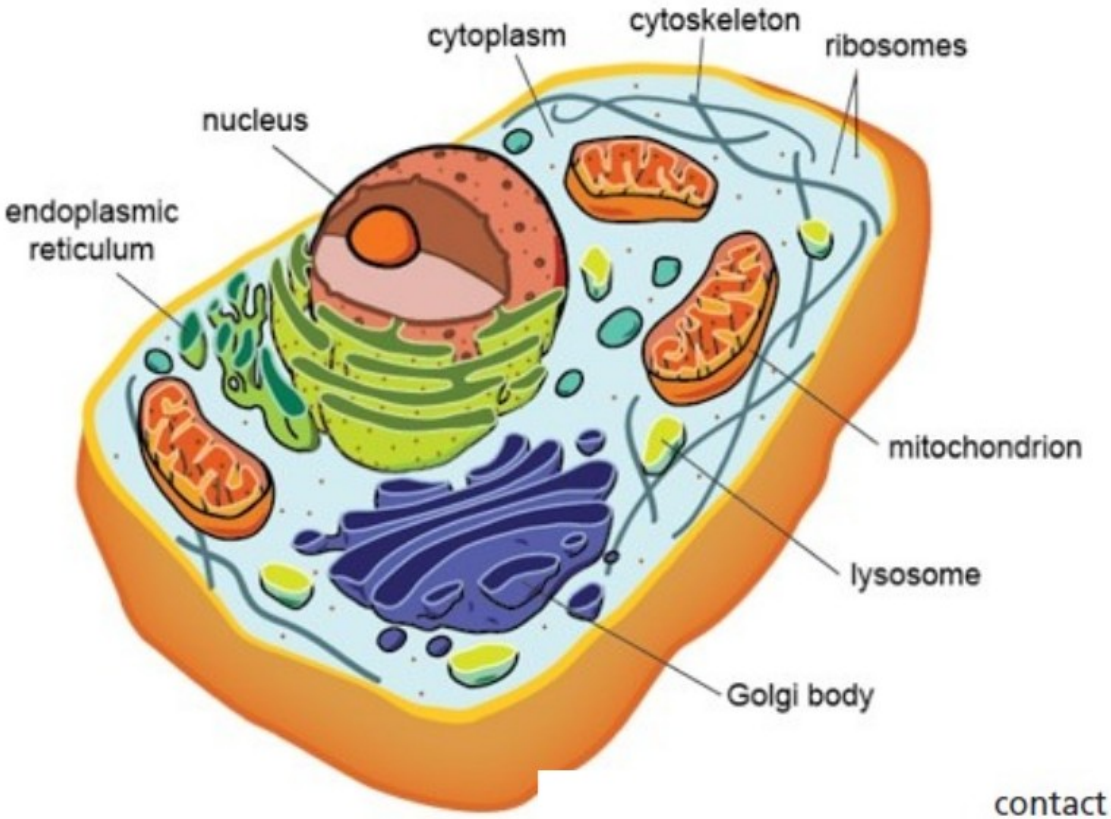


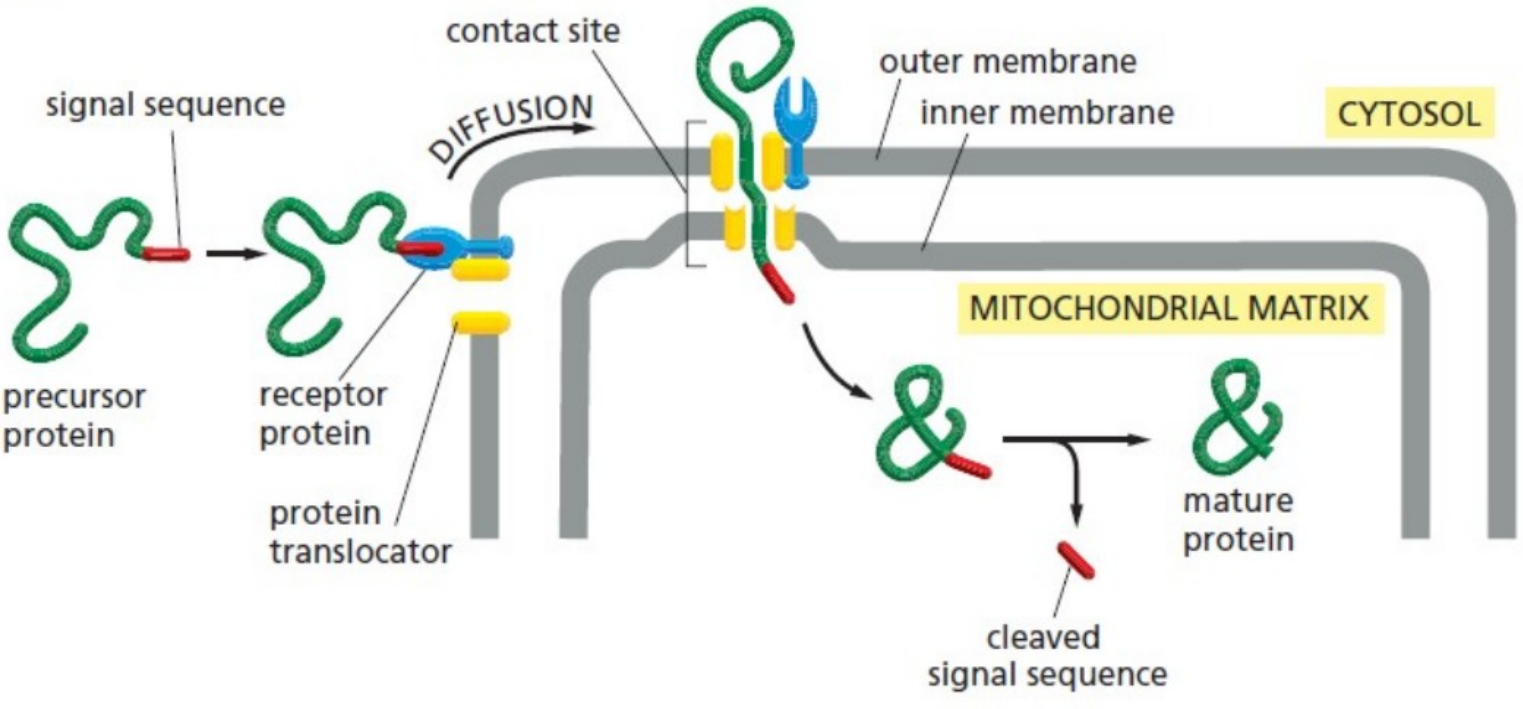
Nature Reviews | Molecular Cell Biology

Prediction of protein subcellular localization

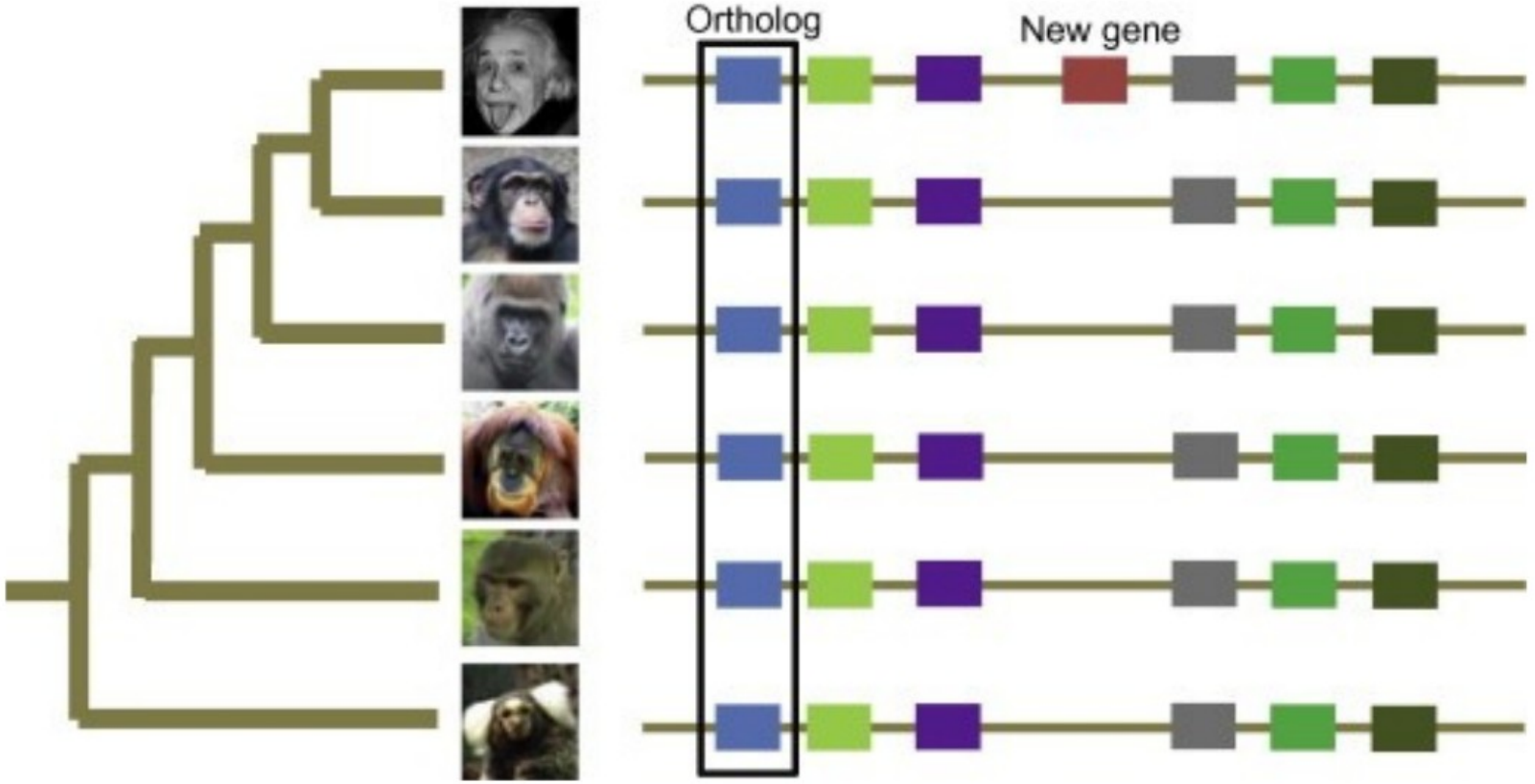
A particular type of protein domain (motif)



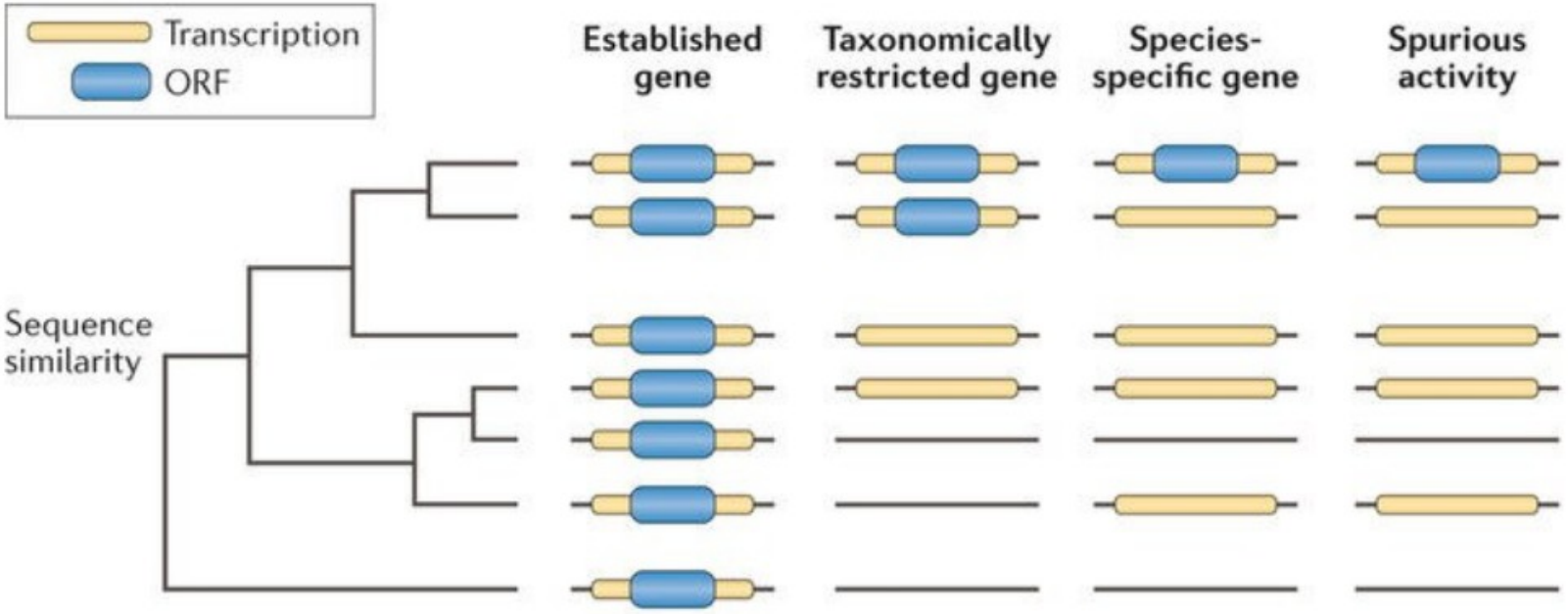

cytoplasm
cytoskeleton
ribosomes
nucleus
endoplasmic reticulum
mitochondrion
lysosome
Golgi body


contact site
outer membrane
inner membrane
CYTOSOL
signal sequence
DIFFUSION
MITOCHONDRIAL MATRIX
precursor protein
receptor protein
protein translocator
mature protein
cleaved signal sequence

De novo origin of genes.



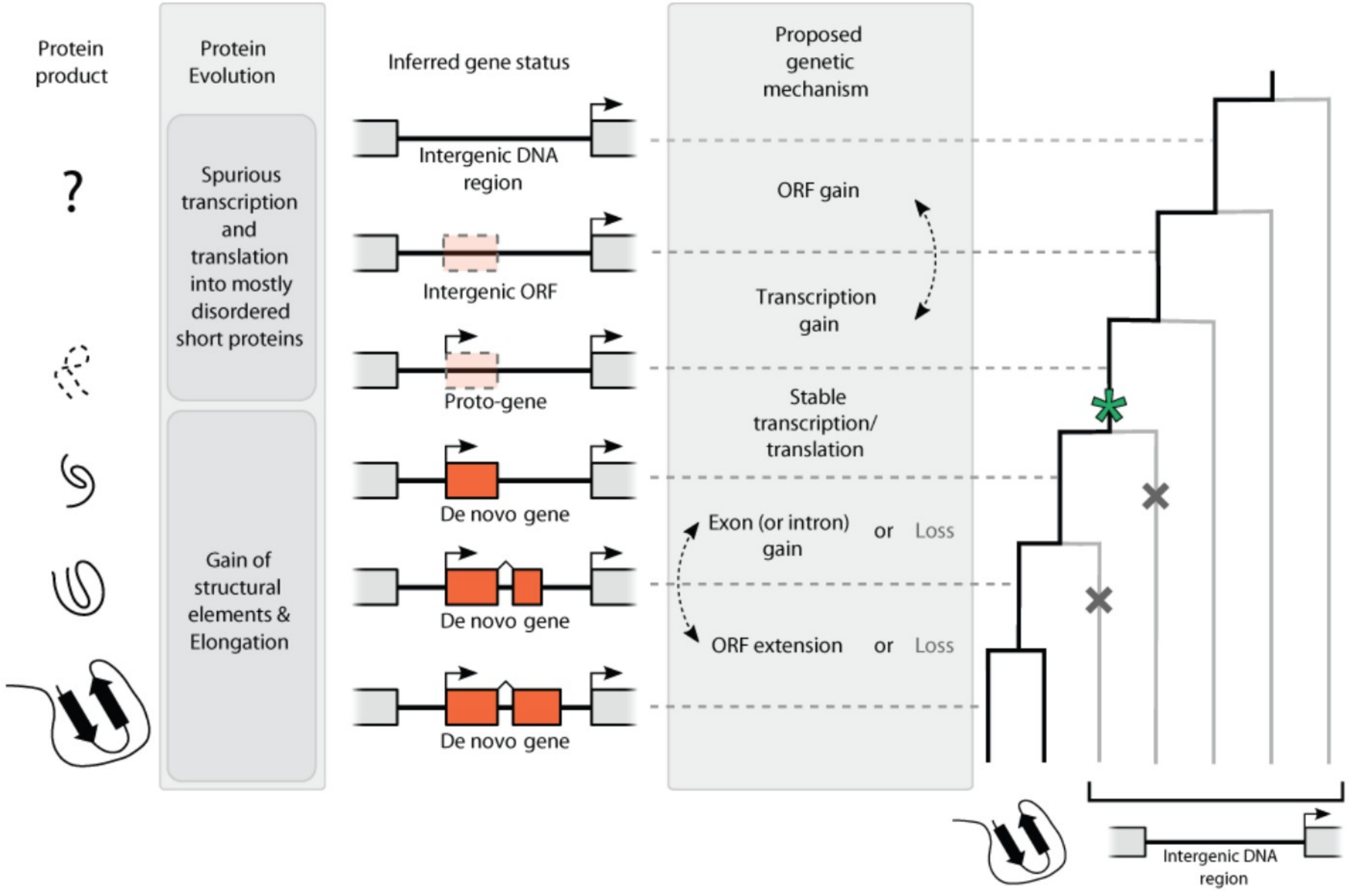
De novo origin of genes.



Evolutionary conservation	✓	✓or✗	✗	✗
Purifying selection ($dN/dS < 1$)	✓	✓	NA	NA
Positive selection ($dN/dS > 1$)	✓or✗	✓	NA	NA
Transcription	✓	✓	✓	✓
Translation	✓	✓	✓	✓
Knockdown or knockout phenotype	✓or✗	✓or✗	✓or✗	✗

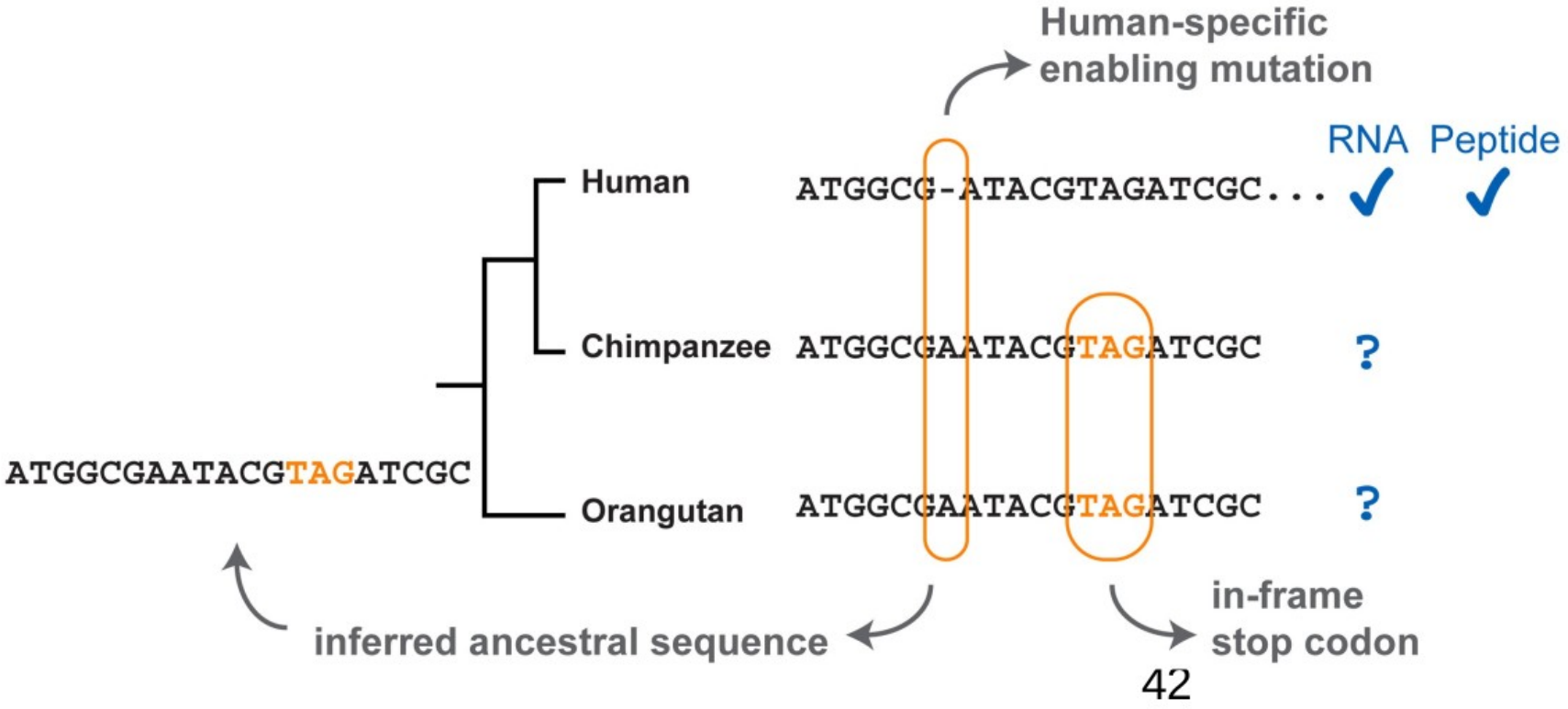
Nature Reviews | Genetics

De novo origin of genes.

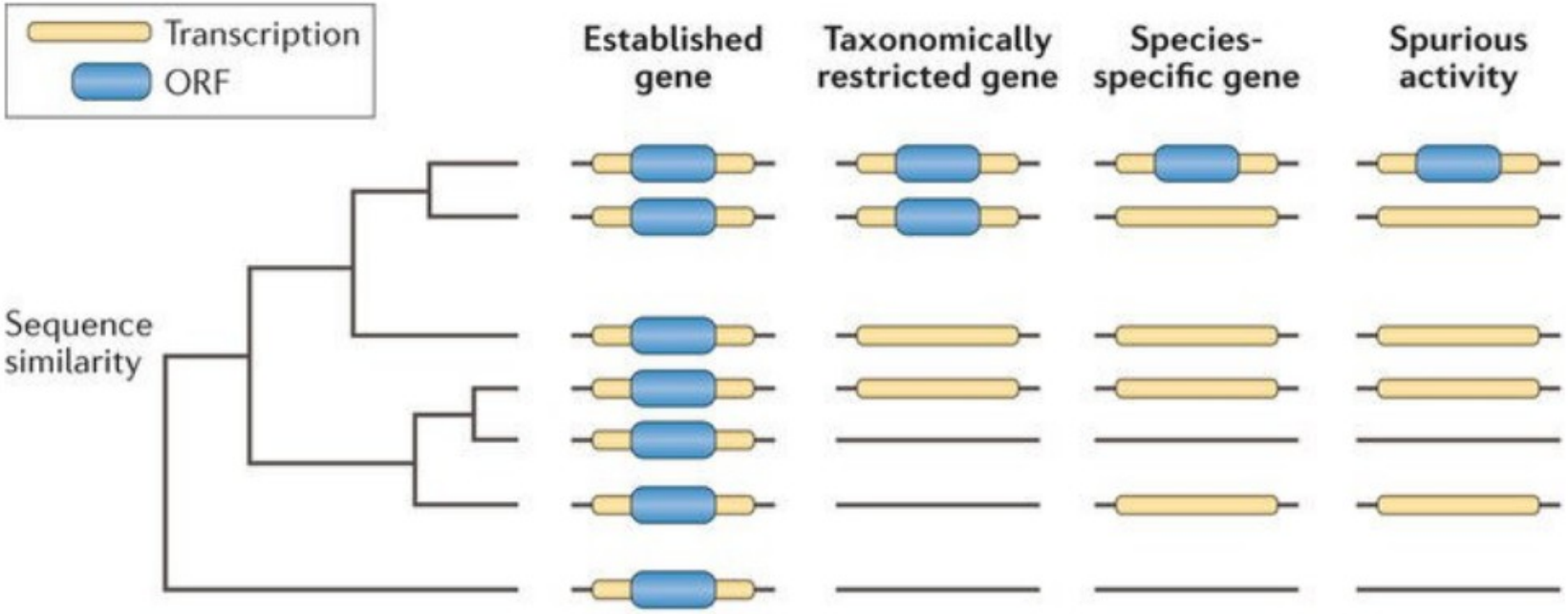


De novo origin of genes.

Pervasive transcription
Transposition of promoters/enhancers increasing expression
Mutations originating promoters, increasing expression, extending ORFs, otimizing codon usage
Perhaps first functions as RNAs, or proto-proteins.



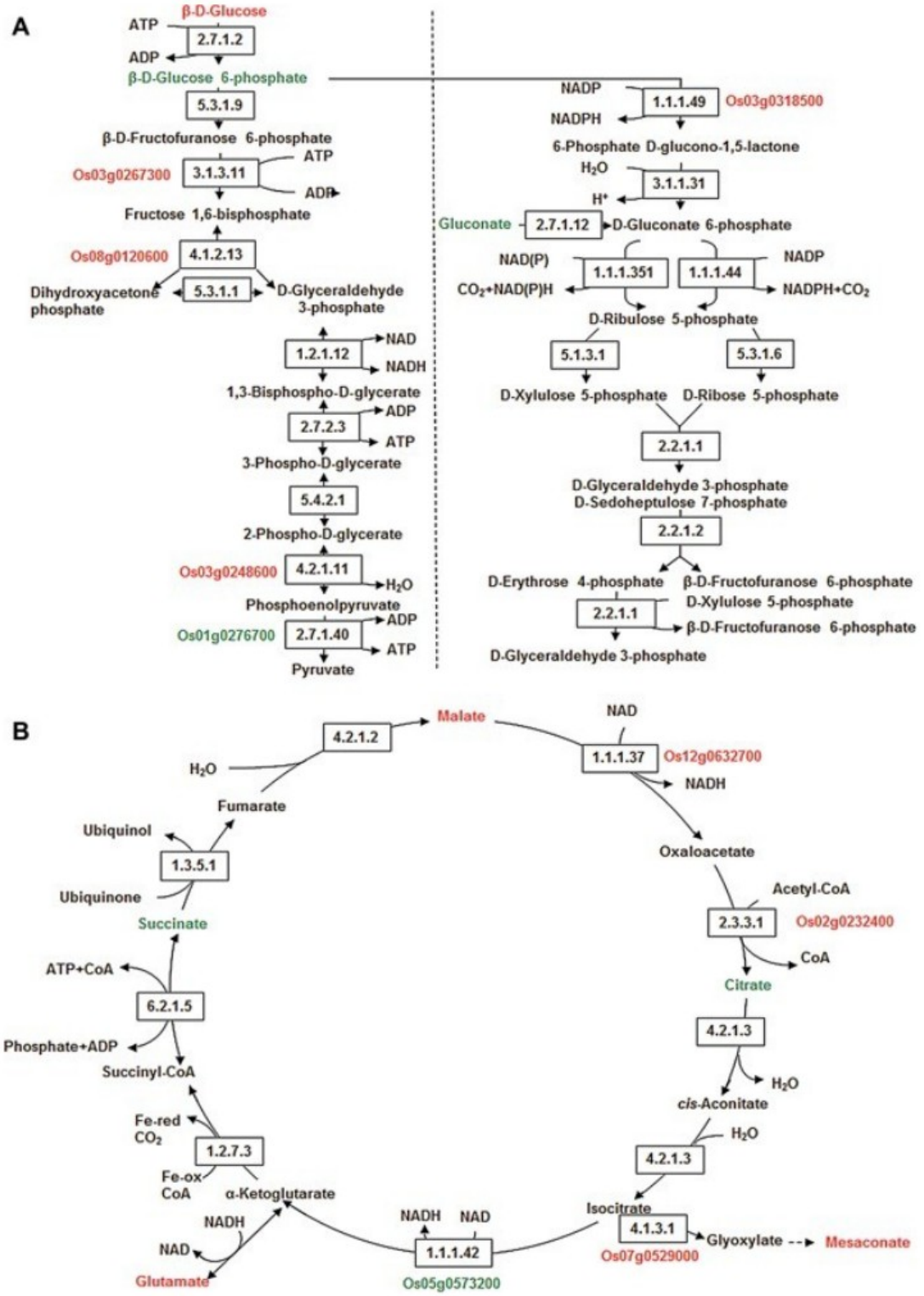
De novo origin of genes.



	Established gene	Taxonomically restricted gene	Species-specific gene	Spurious activity
Evolutionary conservation	✓	✓or✗	✗	✗
Purifying selection ($dN/dS < 1$)	✓	✓	NA	NA
Positive selection ($dN/dS > 1$)	✓or✗	✓	NA	NA
Transcription	✓	✓	✓	✓
Translation	✓	✓	✓	✓
Knockdown or knockout phenotype	✓or✗	✓or✗	✓or✗	✗

Nature Reviews | Genetics

Pathways

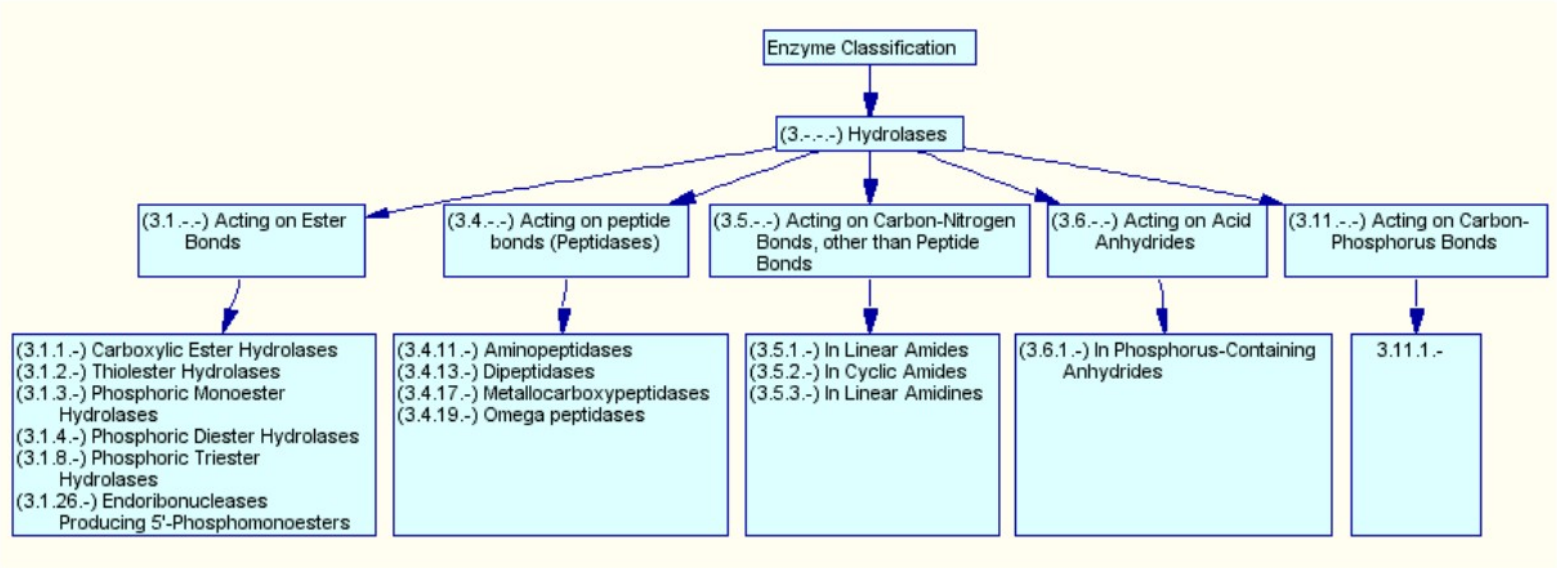

A
β-D-Glucose
ATP
2.7.1.2
ADP
β-D-Glucose 6-phosphate
5.3.1.9
β-D-Fructofuranose 6-phosphate
Os03g0267300
3.1.3.11
ATP
ADP
Fructose 1,6-bisphosphate
Os08g0120600
4.1.2.13
Dihydroxyacetone phosphate
5.3.1.1
D-Glyceraldehyde 3-phosphate
1.2.1.12
NAD
NADH
1,3-Bisphospho-D-glycerate
2.7.2.3
ADP
ATP
3-Phospho-D-glycerate
5.4.2.1
2-Phospho-D-glycerate
Os03g0248600
4.2.1.11
H2O
Phosphoenolpyruvate
Os01g0276700
2.7.1.40
ADP
ATP
Pyruvate
NADP
1.1.1.49
Os03g0318500
NADPH
6-Phosphate D-glucono-1,5-lactone
H2O
3.1.1.31
H+
Gluconate
2.7.1.12
D-Gluconate 6-phosphate
NAD(P)
1.1.1.351
NADP
1.1.1.44
CO2+NAD(P)H
NADPH+CO2
D-Ribulose 5-phosphate
5.1.3.1
5.3.1.6
D-Xylulose 5-phosphate
D-Ribose 5-phosphate
2.2.1.1
D-Glyceraldehyde 3-phosphate
D-Sedoheptulose 7-phosphate
2.2.1.2
D-Erythrose 4-phosphate
β-D-Fructofuranose 6-phosphate
2.2.1.1
D-Xylulose 5-phosphate
β-D-Fructofuranose 6-phosphate
D-Glyceraldehyde 3-phosphate
B
4.2.1.2
Malate
NAD
1.1.1.37
Os12g0632700
NADH
H2O
Fumarate
Ubiquinol
1.3.5.1
Ubiquinone
Succinate
Oxaloacetate
Acetyl-CoA
2.3.3.1
Os02g0232400
CoA
Citrate
ATP+CoA
6.2.1.5
Phosphate+ADP
Succinyl-CoA
4.2.1.3
H2O
cis-Aconitate
Fe-red
CO2
1.2.7.3
Fe-ox
CoA
α-Ketoglutarate
H2O
4.2.1.3
Isocitrate
NADH
NAD
1.1.1.42
Os05g0573200
4.1.3.1
Glyoxylate
Mesaconate
Os07g0529000
NADH
NAD
Glutamate

Pathways

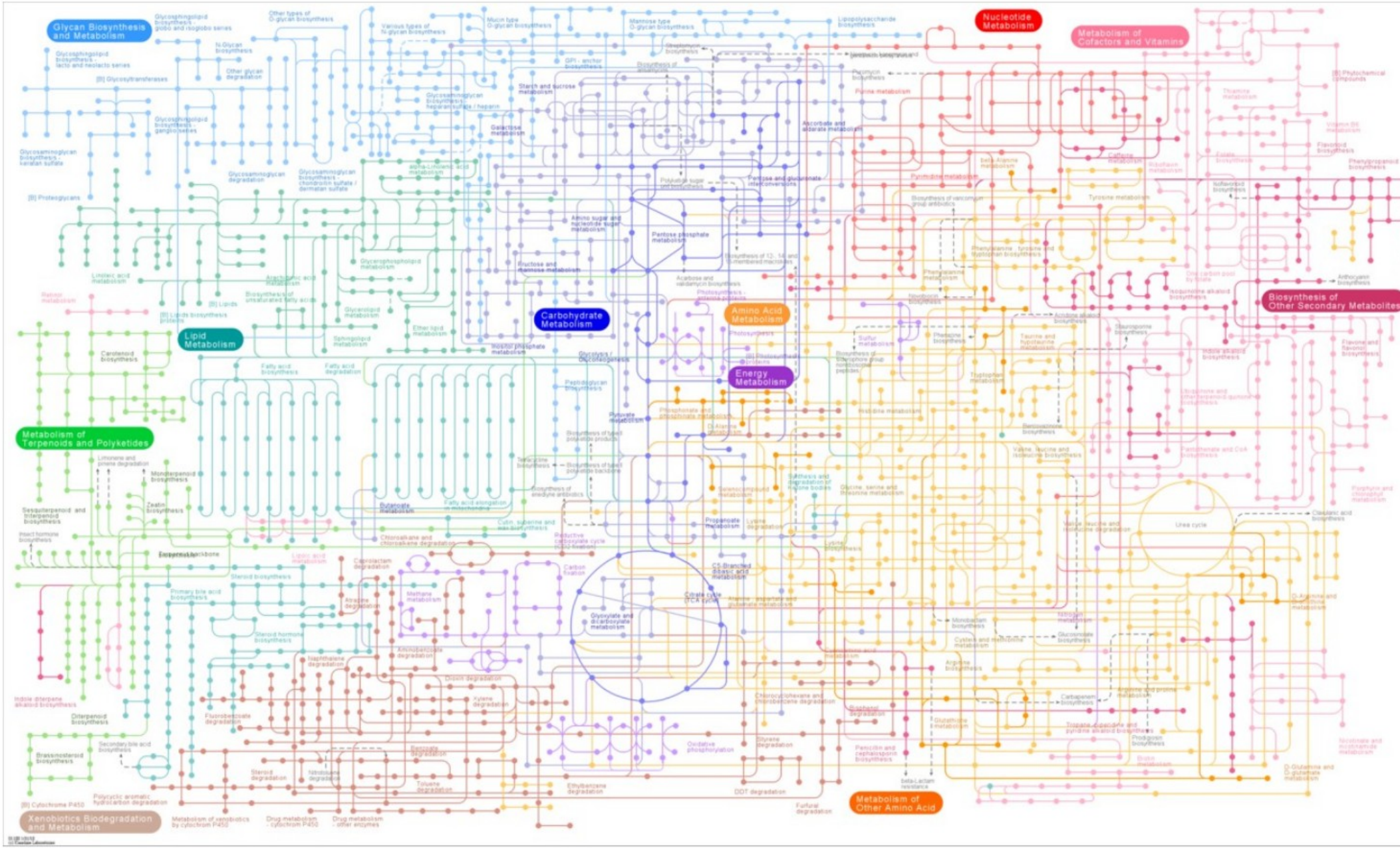
EC 1	Oxidoreductases
EC 1.3	Acting on the CH-CH Group of Donors
EC 1.3.1	With NAD^+ or $NADP^+$ as acceptor
EC 1.3.1.21	7-dehydrocholesterol reductase

IUBMB Enzyme Nomenclature

EC 1.3.1.21

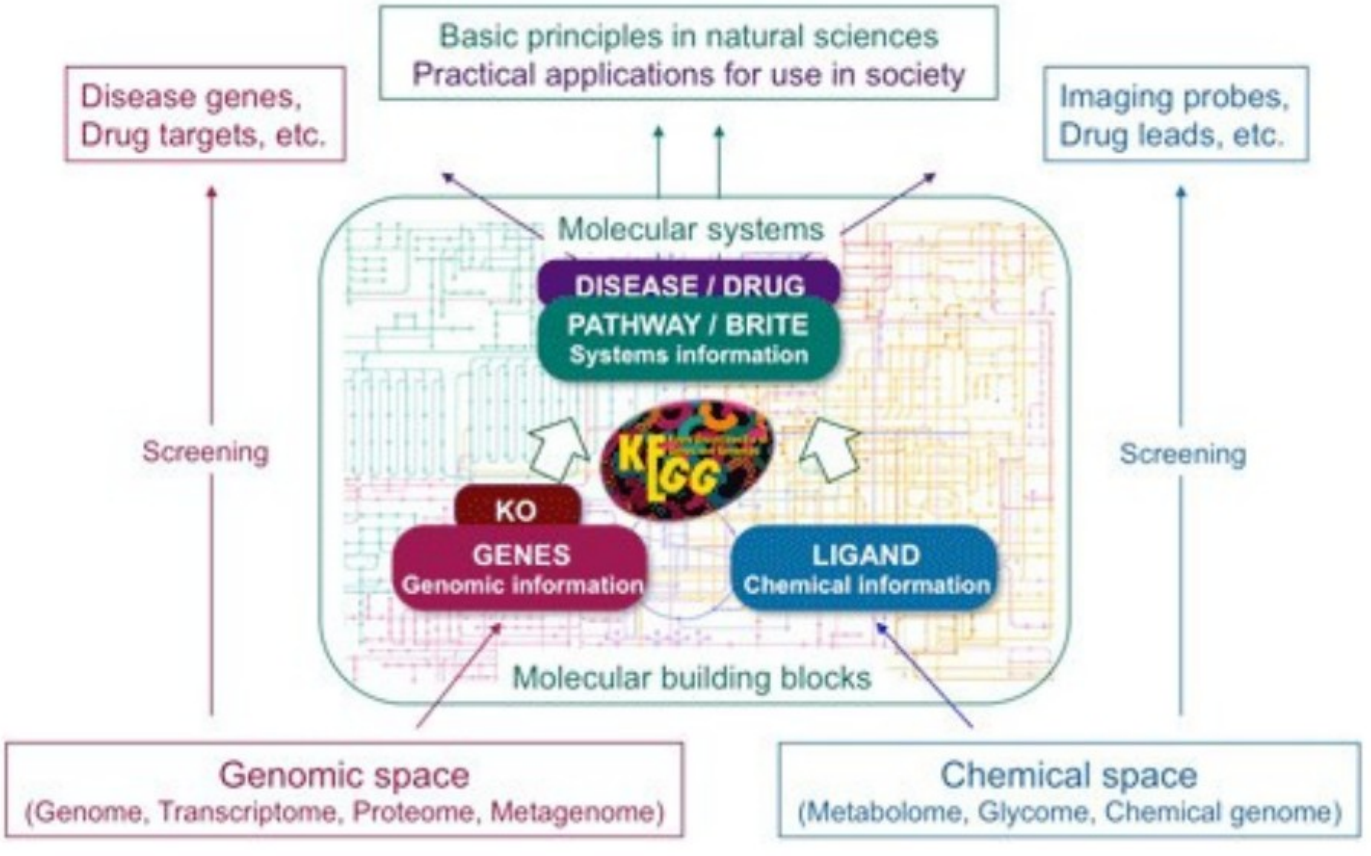


Pathways

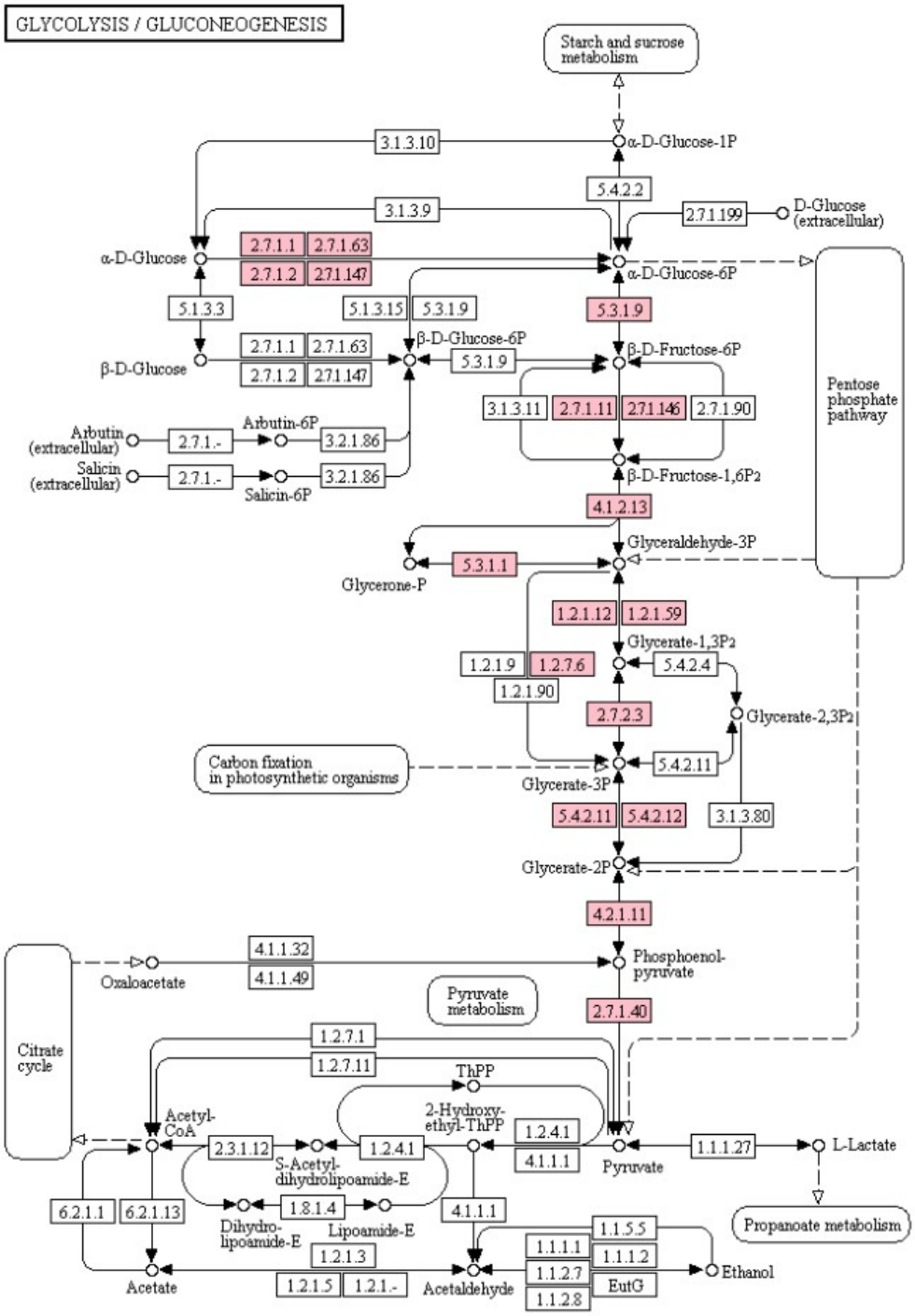

Glycan Biosynthesis and Metabolism
Nucleotide Metabolism
Metabolism of Cofactors and Vitamins
Carbohydrate Metabolism
Amino Acid Metabolism
Lipid Metabolism
Energy Metabolism
Biosynthesis of Other Secondary Metabolites
Metabolism of Terpenoids and Polyketides
Metabolism of Other Amino Acid
Xenobiotics Biodegradation and Metabolism

Pathways

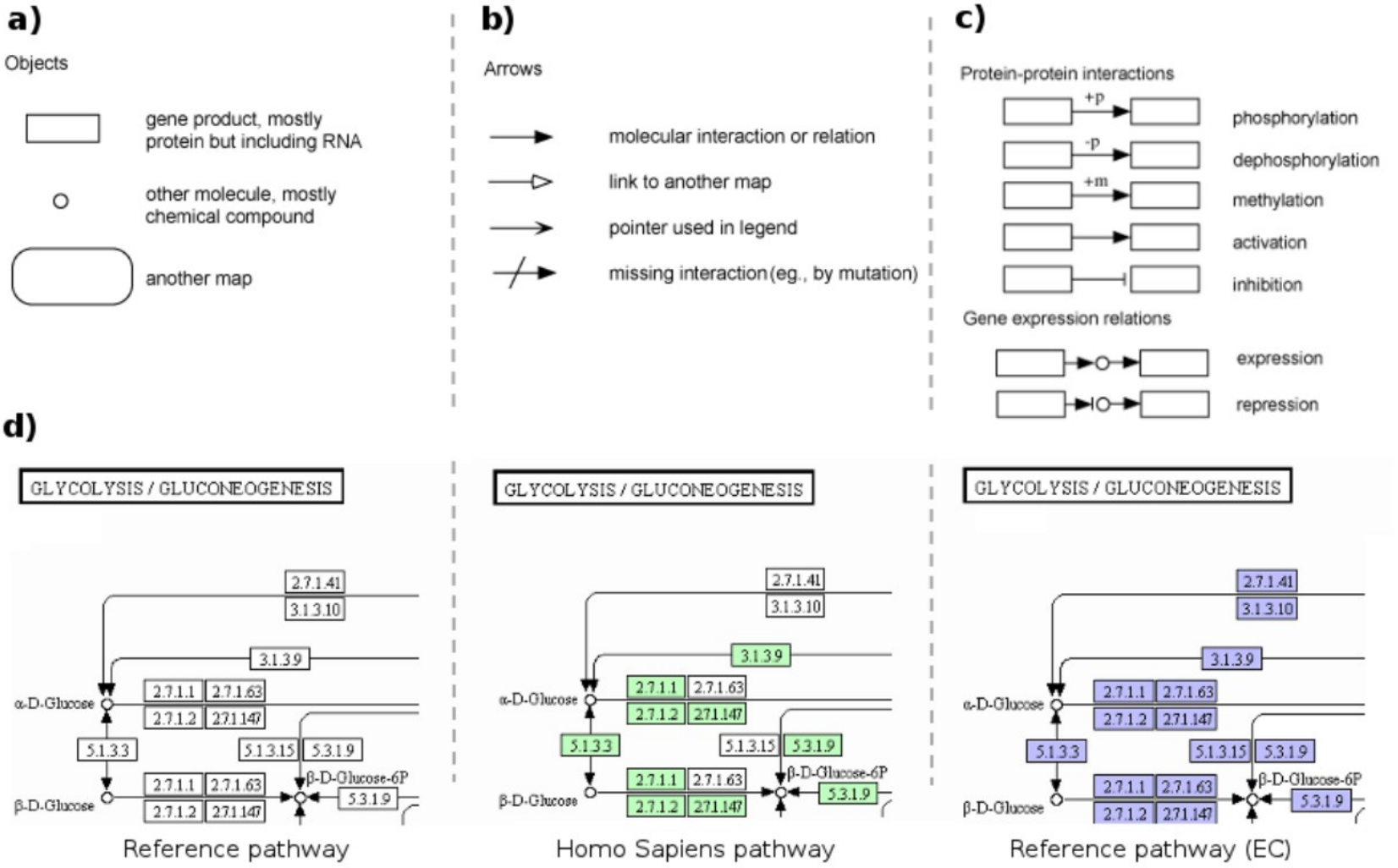

KEGG
Kyoto Encyclopedia of
Genes and Genomes


Basic principles in natural sciences
Practical applications for use in society
Disease genes,
Drug targets, etc.
Imaging probes,
Drug leads, etc.
Molecular systems
DISEASE / DRUG
PATHWAY / BRITE
Systems information
Screening
Screening
KO
GENES
Genomic information
LIGAND
Chemical information
Molecular building blocks
Genomic space
(Genome, Transcriptome, Proteome, Metagenome)
Chemical space
(Metabolome, Glycome, Chemical genome)

Pathways


GLYCOLYSIS / GLUCONEOGENESIS
Starch and sucrose metabolism
α-D-Glucose-1P
D-Glucose (extracellular)
α-D-Glucose
α-D-Glucose-6P
β-D-Glucose-6P
β-D-Glucose
β-D-Fructose-6P
Pentose phosphate pathway
Arbutin (extracellular)
Arbutin-6P
Salicin (extracellular)
Salicin-6P
β-D-Fructose-1,6P2
Glyceraldehyde-3P
Glycerone-P
Glycerate-1,3P2
Glycerate-2,3P2
Carbon fixation in photosynthetic organisms
Glycerate-3P
Glycerate-2P
Phosphoenol-pyruvate
Oxaloacetate
Pyruvate metabolism
Citrate cycle
ThPP
Acetyl-CoA
2-Hydroxy-ethyl-ThPP
S-Acetyl-dihydrolipoamide-E
Pyruvate
L-Lactate
Dihydro-lipoamide-E
Lipoamide-E
Propanoate metabolism
Acetate
Acetaldehyde
Ethanol
EutG
00010 2/19/18
(c) Kanehisa Laboratories

Pathways


a)
Objects
gene product, mostly protein but including RNA
other molecule, mostly chemical compound
another map
b)
Arrows
molecular interaction or relation
link to another map
pointer used in legend
missing interaction (eg., by mutation)
c)
Protein-protein interactions
+p
phosphorylation
-p
dephosphorylation
+m
methylation
activation
inhibition
Gene expression relations
expression
repression
d)
GLYCOLYSIS / GLUCONEOGENESIS
2.7.1.41
3.1.3.10
3.1.3.9
α-D-Glucose
2.7.1.1
2.7.1.63
2.7.1.2
2.7.1.147
5.1.3.3
5.1.3.15
5.3.1.9
β-D-Glucose-6P
β-D-Glucose
Reference pathway
Homo Sapiens pathway
Reference pathway (EC)

Pathways

Category	Database	Content
Systems information	KEGG PATHWAY	KEGG pathway maps
	KEGG BRITE	BRITE functional hierarchies
	KEGG MODULE	KEGG modules of functional units
	KEGG ORTHOLOGY	KEGG Orthology (KO) groups
Genomic information	KEGG GENOME	KEGG organisms with complete genomes
	KEGG GENES	Gene catalogs of complete genomes
	KEGG SSDB	Sequence similarity database for GENES
Chemical information	KEGG COMPOUND	Metabolites and other small molecules
	KEGG GLYCAN	Glycans
	KEGG REACTION	Biochemical reactions
	KEGG RPAIR	Reactant pair chemical transformations
	KEGG RCLASS	Reaction class defined by RPAIR
	KEGG ENZYME	Enzyme nomenclature
Health information	KEGG DISEASE	Human diseases
	KEGG DRUG	Drugs
	KEGG DGROUP	Drug groups
	KEGG ENVIRON	Crude drugs and health-related substances


GENEONTOLOGY
Unifying Biology

The Gene Ontology

Ontology: The study of “being”. Ontology often deals with questions concerning **what entities exist and how such entities may be grouped, related within a hierarchy, and subdivided according to similarities and differences.**

The **Gene Ontology** project aims to:

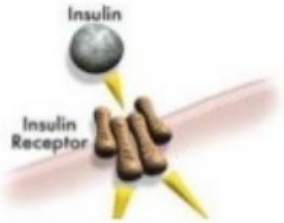
1) maintain and develop a controlled vocabulary of gene and gene product attributes;

2) annotate genes and gene products, and assimilate and disseminate annotation data;

3) provide tools for easy access to all aspects of the data provided by the project, and to enable functional interpretation of experimental data using the GO, for example via enrichment analysis.

1. Molecular Function

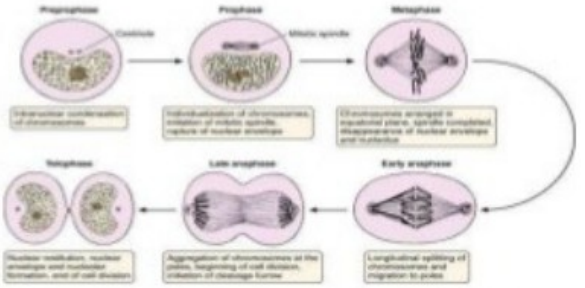
An elemental activity or task or job



- protein kinase activity
- insulin receptor activity

2. Biological Process

A commonly recognized series of events



- cell division

3. Cellular Component

Where a gene product is located

- mitochondrion
- mitochondrial matrix
- mitochondrial inner membrane

EMBL-EBI

id: **GO:0048074**

name: **negative regulation of eye pigmentation**

definition: "Any process that stops, prevents, or reduces the frequency, rate or extent of establishment of a pattern of pigment in the eye of an organism.

"synonym: "down regulation of eye pigmentation" EXACT []

synonym: "down-regulation of eye pigmentation" EXACT []

synonym: "downregulation of eye pigmentation" EXACT []

synonym: "inhibition of eye pigmentation" NARROW []

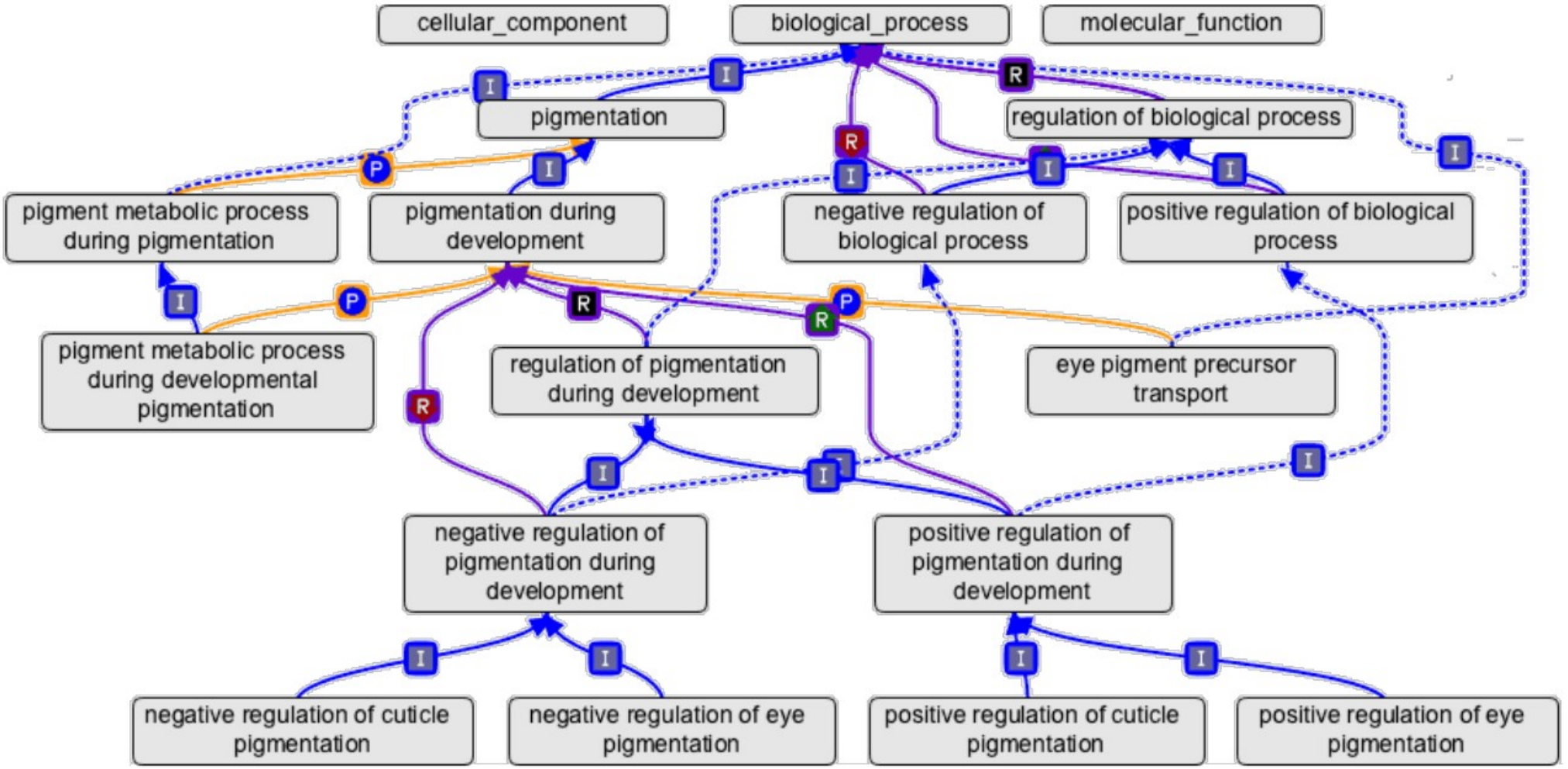
is_a: GO:0048073 ! regulation of eye pigmentation

is_a: GO:0048086 ! negative regulation of developmental pigmentation

relationship: negatively_regulates: GO:0048069 ! eye pigmentation

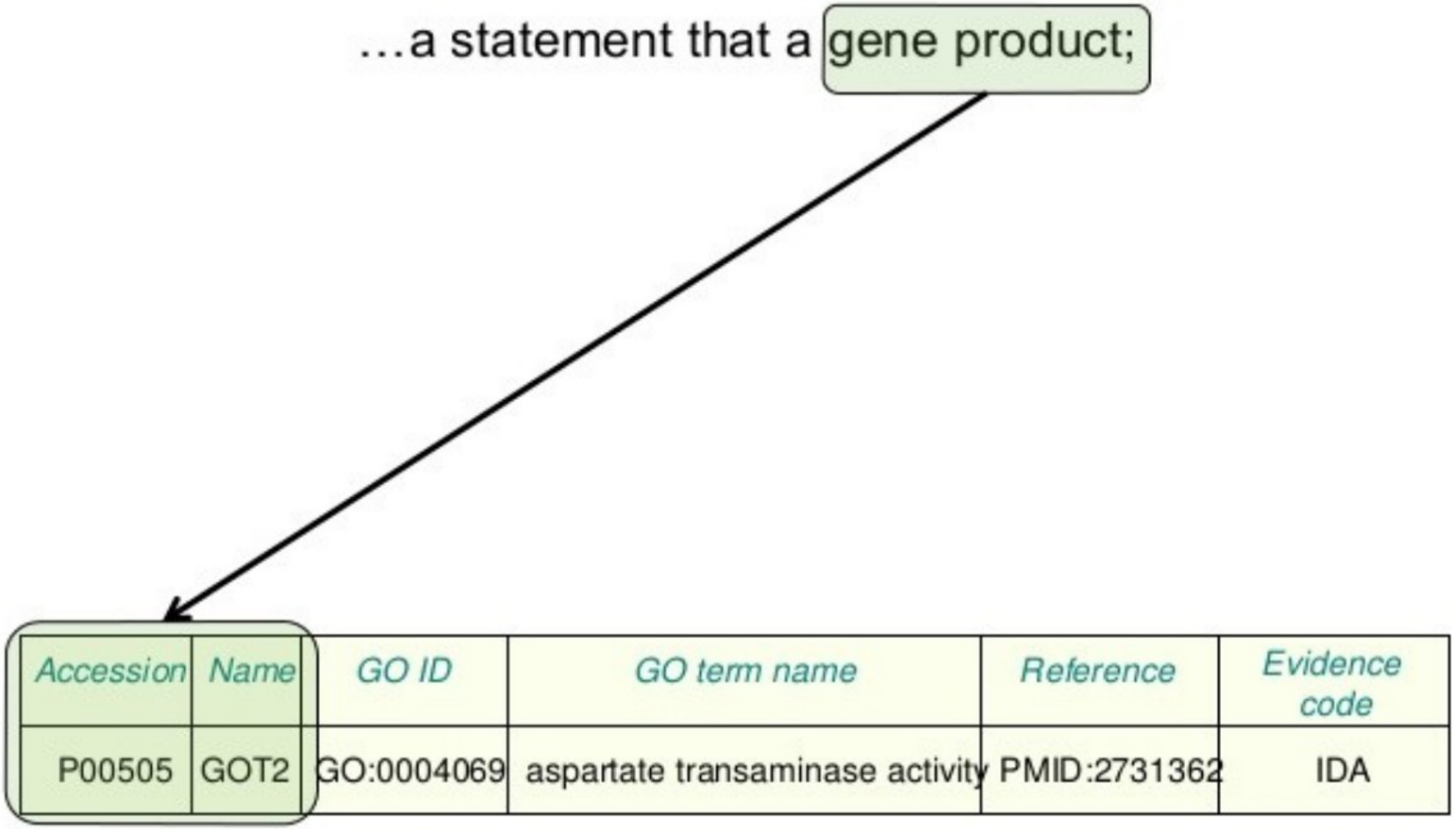
intersection_of: GO:0008150 ! biological_process

intersection_of: negatively_regulates GO:0048069 ! eye pigmentation


cellular_component
biological_process
molecular_function
pigmentation
regulation of biological process
pigment metabolic process during pigmentation
pigmentation during development
negative regulation of biological process
positive regulation of biological process
pigment metabolic process during developmental pigmentation
regulation of pigmentation during development
eye pigment precursor transport
negative regulation of pigmentation during development
positive regulation of pigmentation during development
negative regulation of cuticle pigmentation
negative regulation of eye pigmentation
positive regulation of cuticle pigmentation
positive regulation of eye pigmentation

A GO annotation is …

…a statement that a gene product;



Accession	Name	GO ID	GO term name	Reference	Evidence code
P00505	GOT2	GO:0004069	aspartate transaminase activity	PMID:2731362	IDA

A GO annotation is …

…a statement that a gene product;

1. has a particular **molecular function**
   or is involved in a particular **biological process**
   or is located within a certain **cellular component**

Accession	*Name*	*GO ID*	*GO term name*	*Reference*	*Evidence code*
P00505	GOT2	GO:0004069	aspartate transaminase activity	PMID:2731362	IDA

A GO annotation is …

…a statement that a gene product;

1. has a particular **molecular function**
   or is involved in a particular **biological process**
   or is located within a certain **cellular component**

2. as described in a particular reference

Accession	*Name*	*GO ID*	*GO term name*	*Reference*	*Evidence code*
P00505	GOT2	GO:0004069	aspartate transaminase activity	PMID:2731362	IDA

A GO annotation is …

…a statement that a gene product;

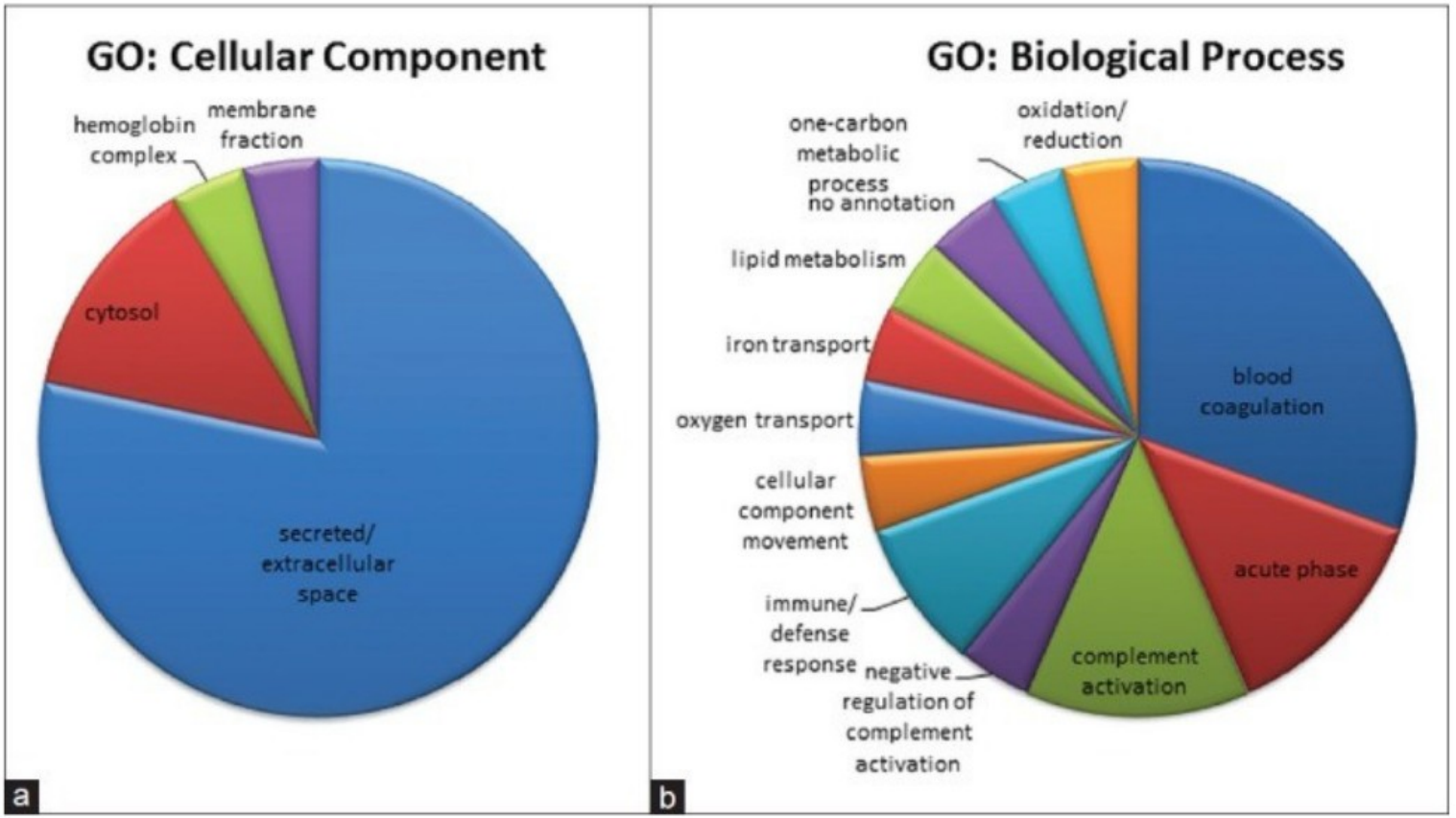
1. has a particular molecular function
   or is involved in a particular biological process
   or is located within a certain cellular component

2. as described in a particular reference

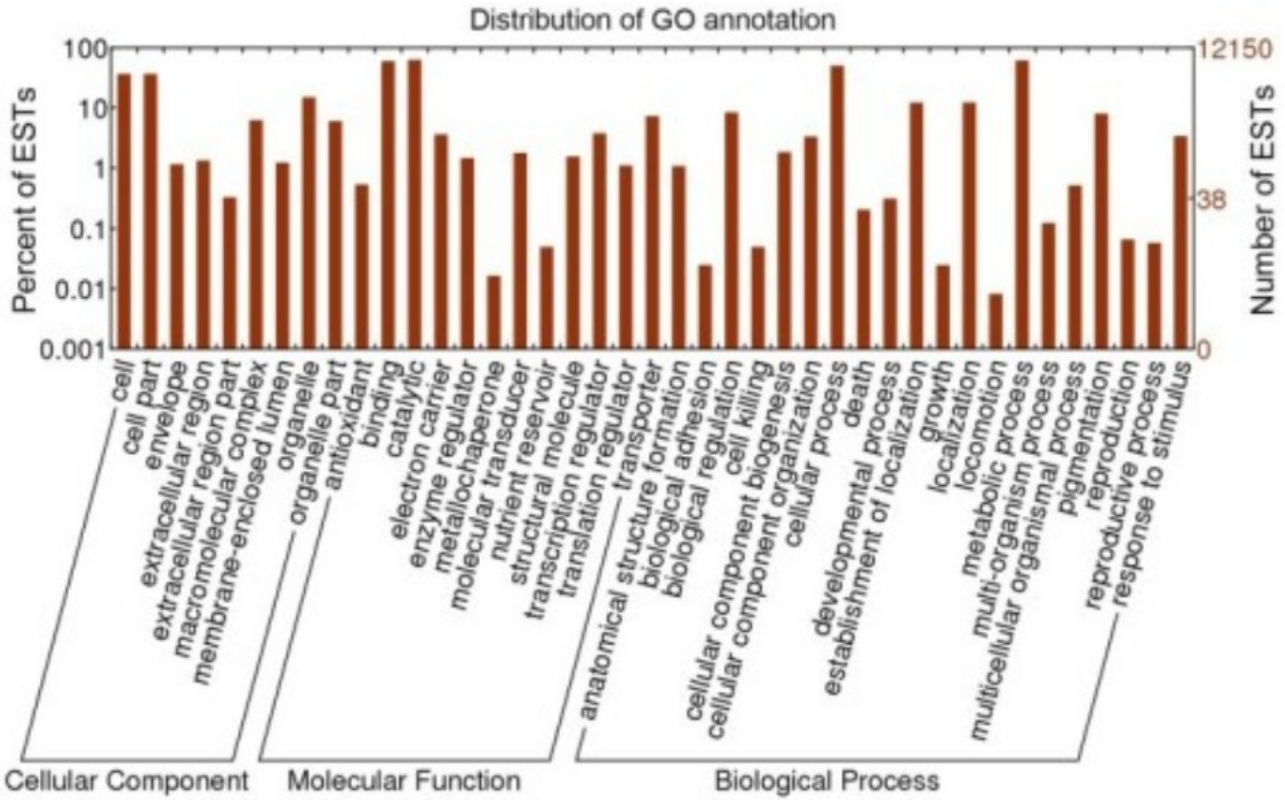
3. as determined by a particular method

Accession	*Name*	*GO ID*	*GO term name*	*Reference*	*Evidence code*
P00505	GOT2	GO:0004069	aspartate transaminase activity	PMID:2731362	IDA

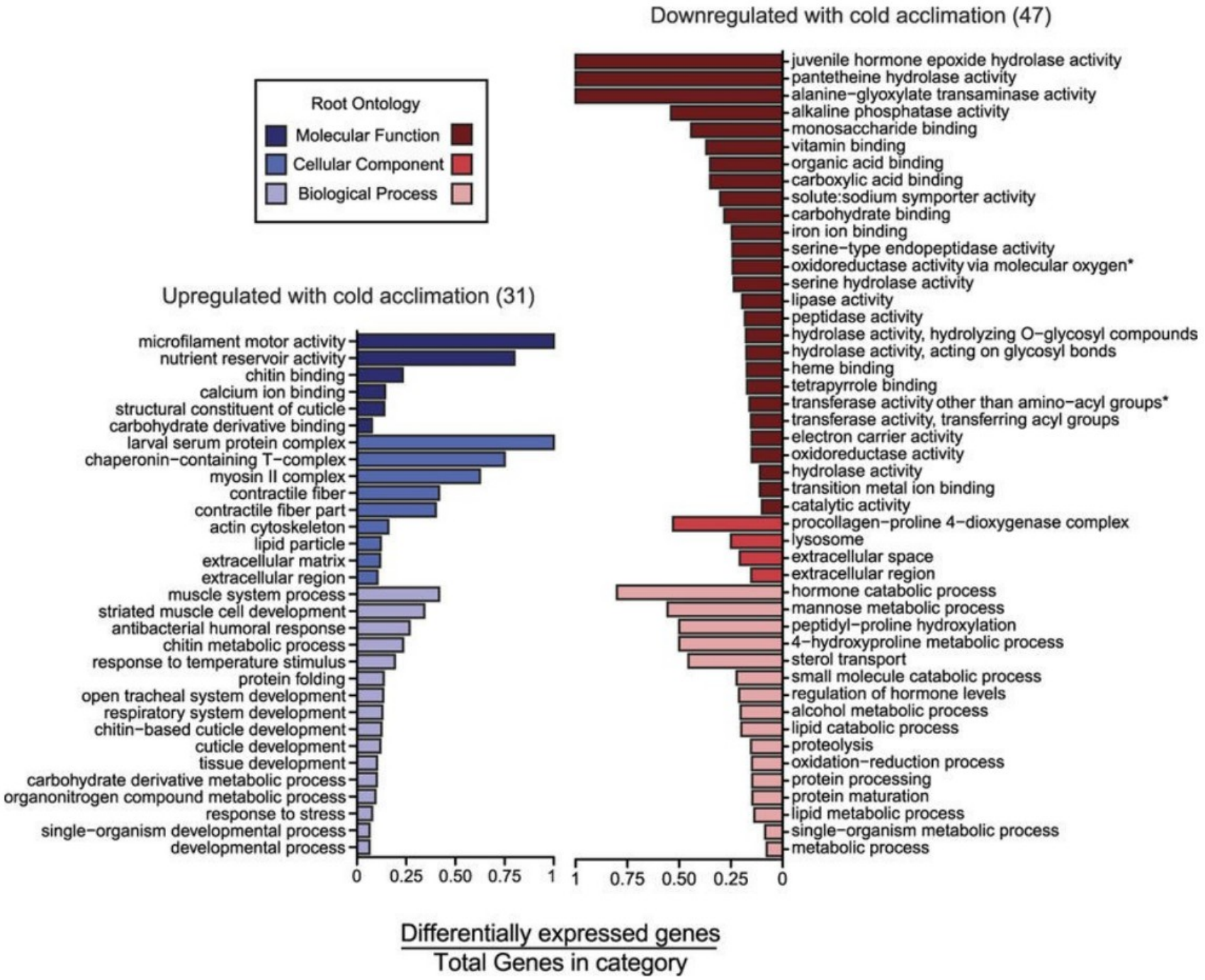
The Gene Ontology



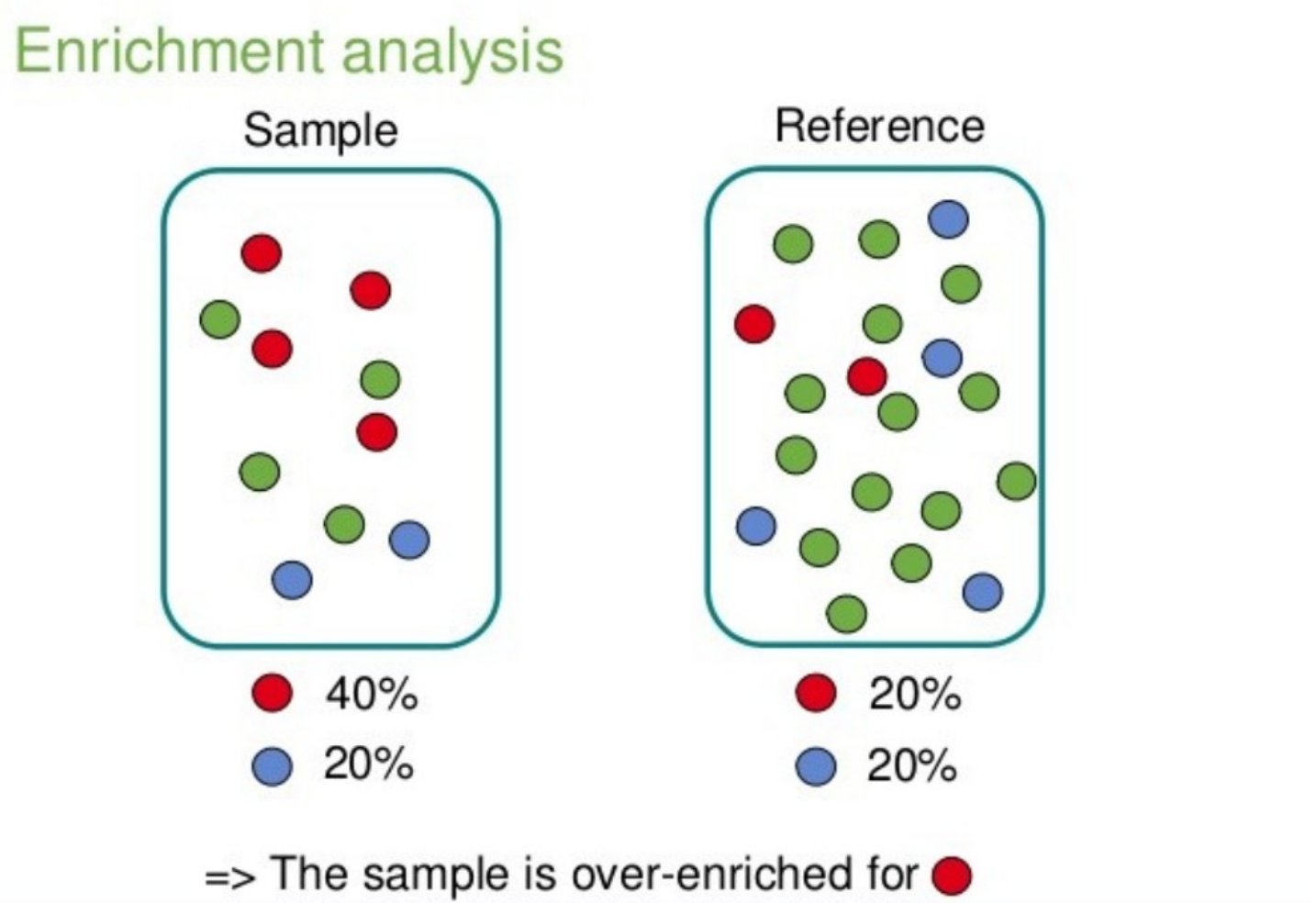
The Gene Ontology



The Gene Ontology



Enrichment analysis



Enrichment analysis

Fisher's exact test

Test for the exact probability of observing a deviation from a background population in a sample (comparing observed and expected values, or values in a contingency table). It is equivalent to the chi-square test but can be used for cases when the number of values is small (<6).

Correction for multiple testing

You can test for differences in the same sample of many categories (i.e. enrichment for different "colors" or functions) and you can perform a Fisher's test for each of these categories. But p-values should be corrected for multiple testing (Bonferroni correction, False Discovery Rate)

Examples

Are duplicated genes in a genome enriched for certain functions?
Are overexpressed genes in a given condition enriched for certain function?
When comparing a set of genomes, are the shared genes enriched in certain functions?