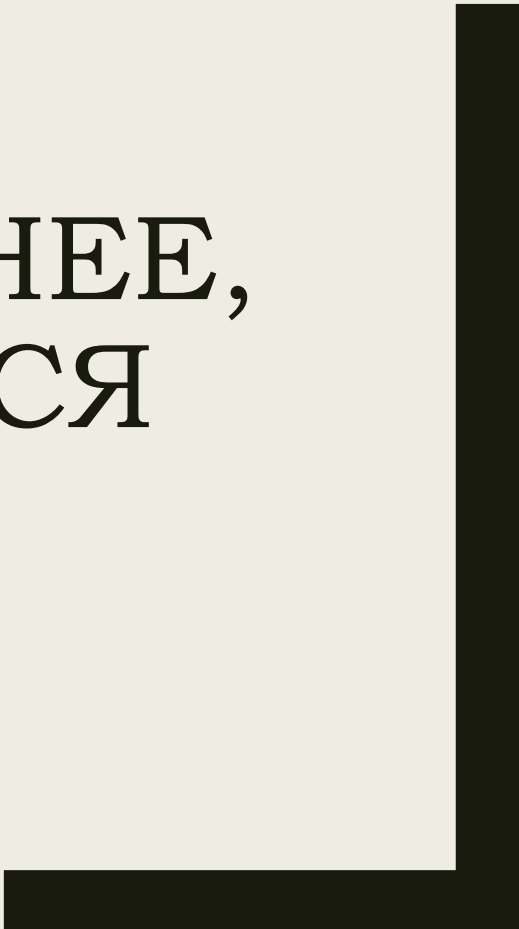




МИР РНК РАЗНООБРАЗНЕЕ, ЧЕМ КАЖЕТСЯ

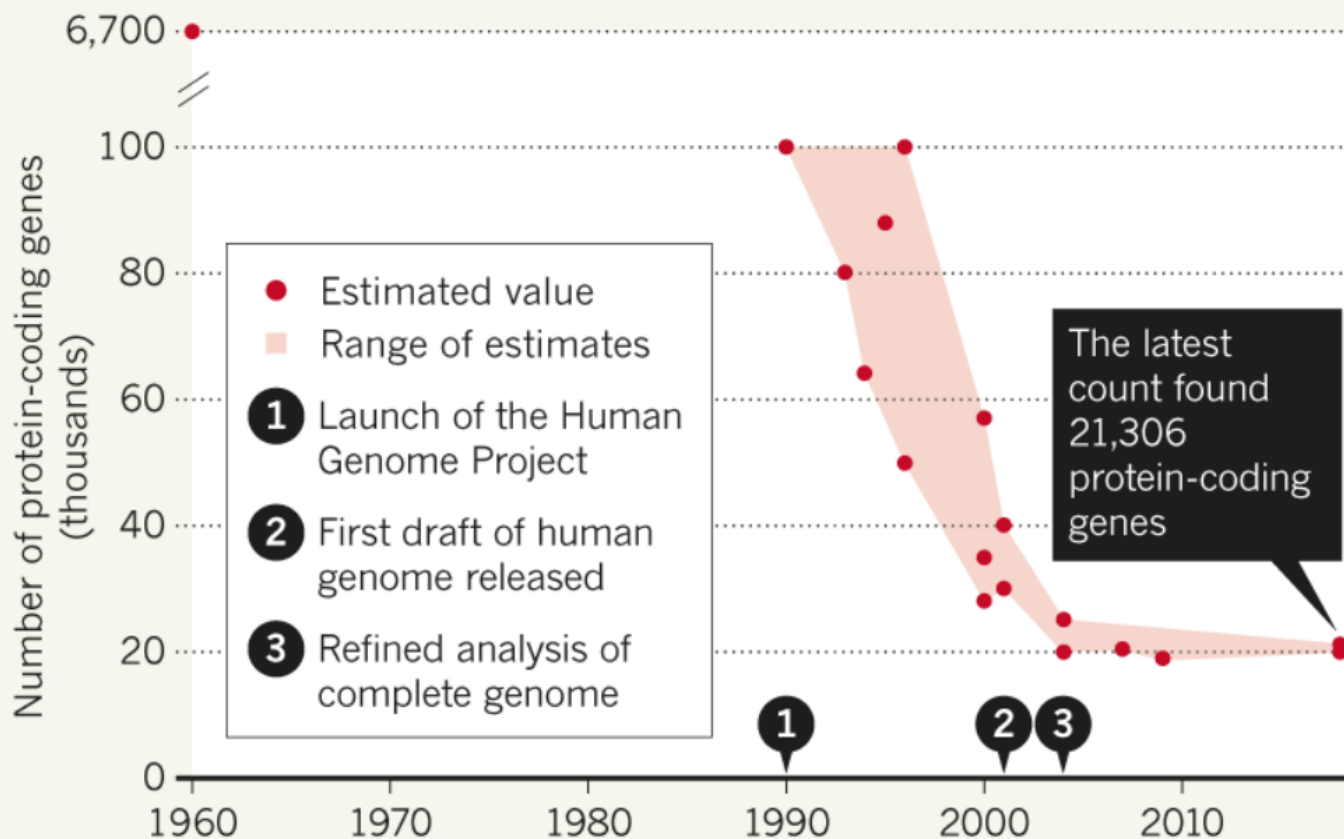
Анастасия Жарикова – 2022
azharikova89@gmail.com



СКОЛЬКО ЖЕ ГЕНОВ?

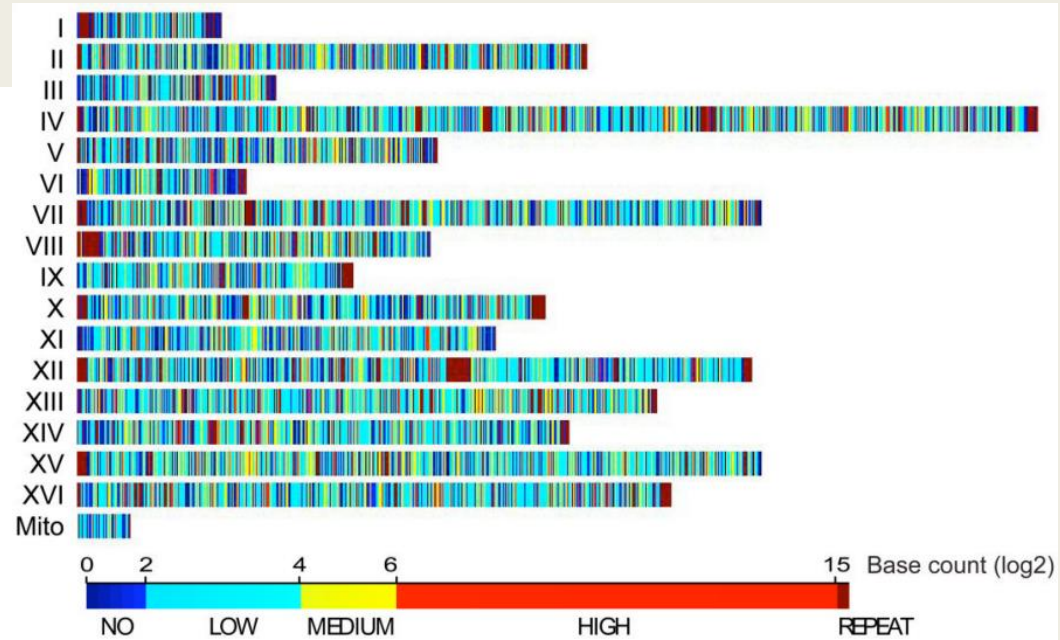
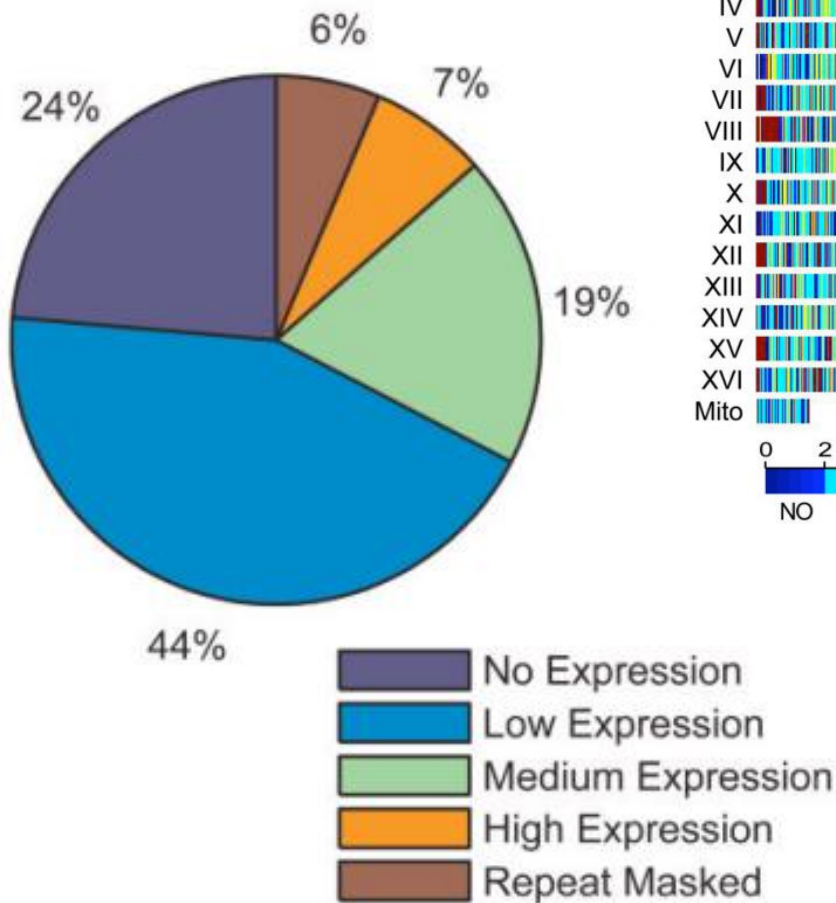
GENE TALLY

Scientists still don't agree on how many protein-making genes the human genome holds, but the range of their estimates has narrowed in recent years.



2018 год

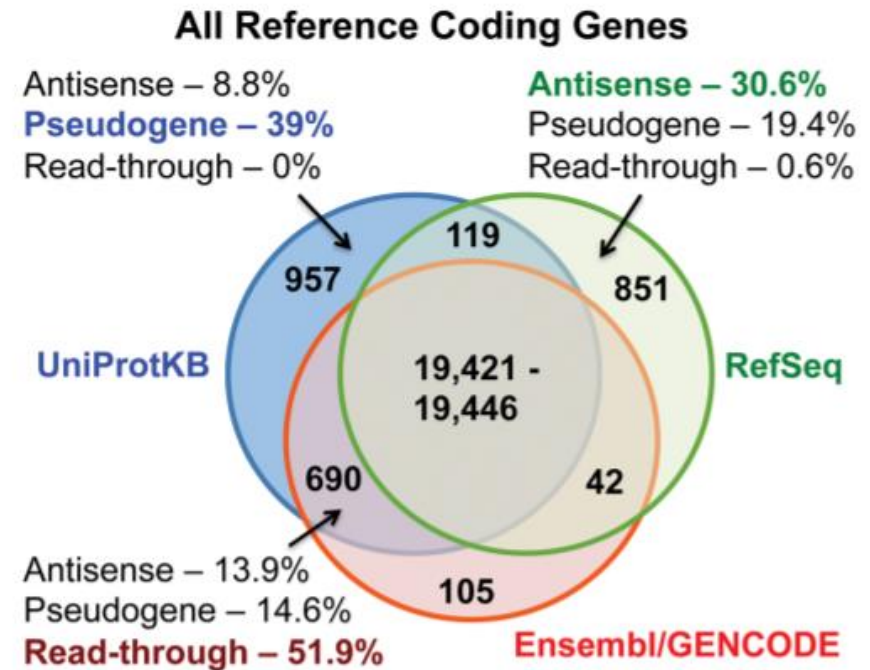
Анализ транскриптомов



GENCODE

Release 39 (GRCh38.p13)

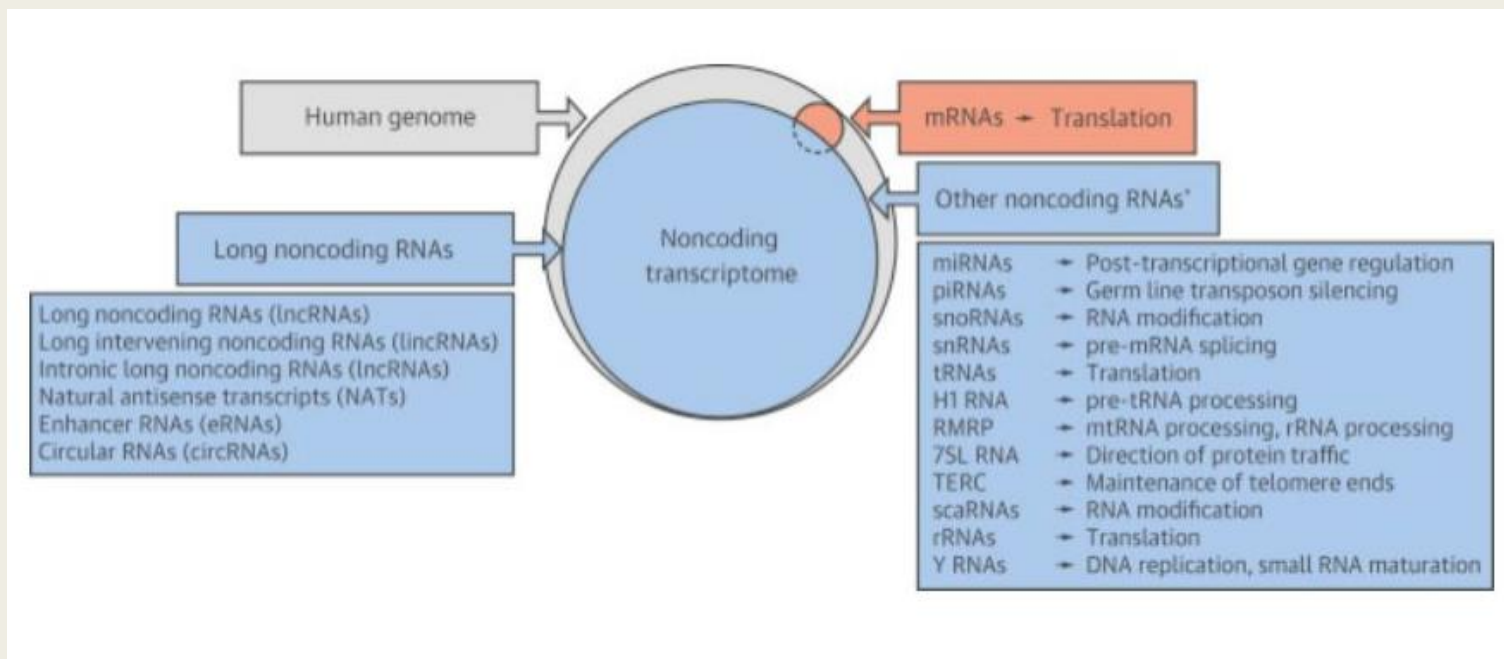
Total No of Genes	61533
Protein-coding genes	19982
Long non-coding RNA genes	18811
Small non-coding RNA genes	7567
Pseudogenes	14763
- processed pseudogenes	10662
- unprocessed pseudogenes	3557
- unitary pseudogenes	243
- polymorphic pseudogenes	50
- pseudogenes	15
Immunoglobulin/T-cell receptor gene segments	
- protein coding segments	409
- pseudogenes	236



<https://pubmed.ncbi.nlm.nih.gov/29982784/>

Следим за версиями всего!

Какие РНК бывают?



Некодирующие РНК

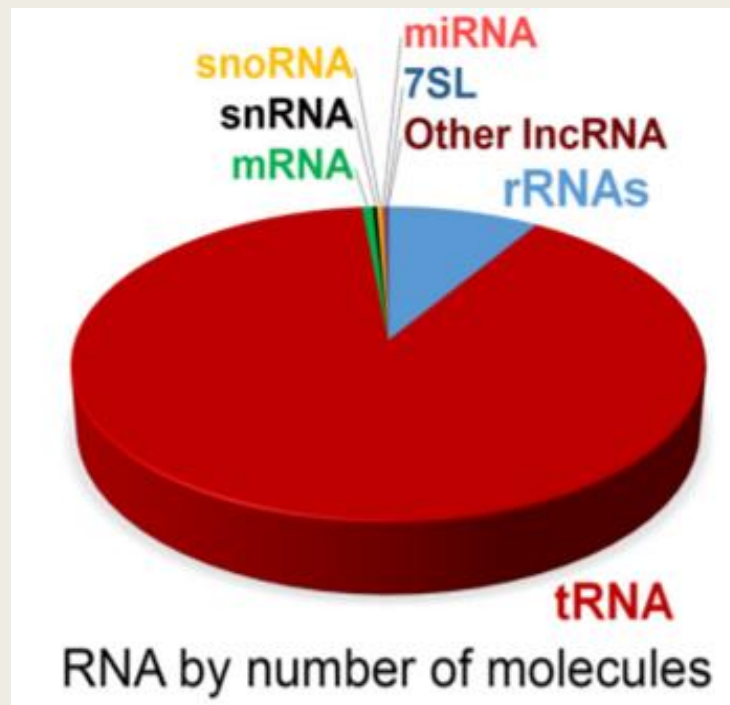
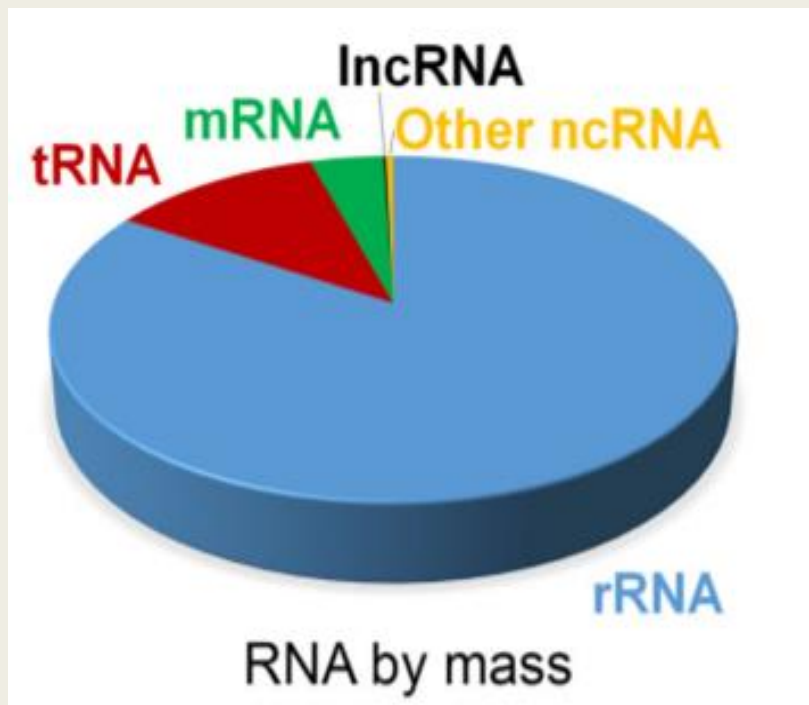
Noncoding RNAs Databases: Current Status and Trends

Vinicius Maracaja-Coutinho, Alexandre Rossi Paschoal, José Carlos Caris-Maldonado, Pedro Vinícius Borges, Almir José Ferreira, and Alan Mitchell Durham

Abstract

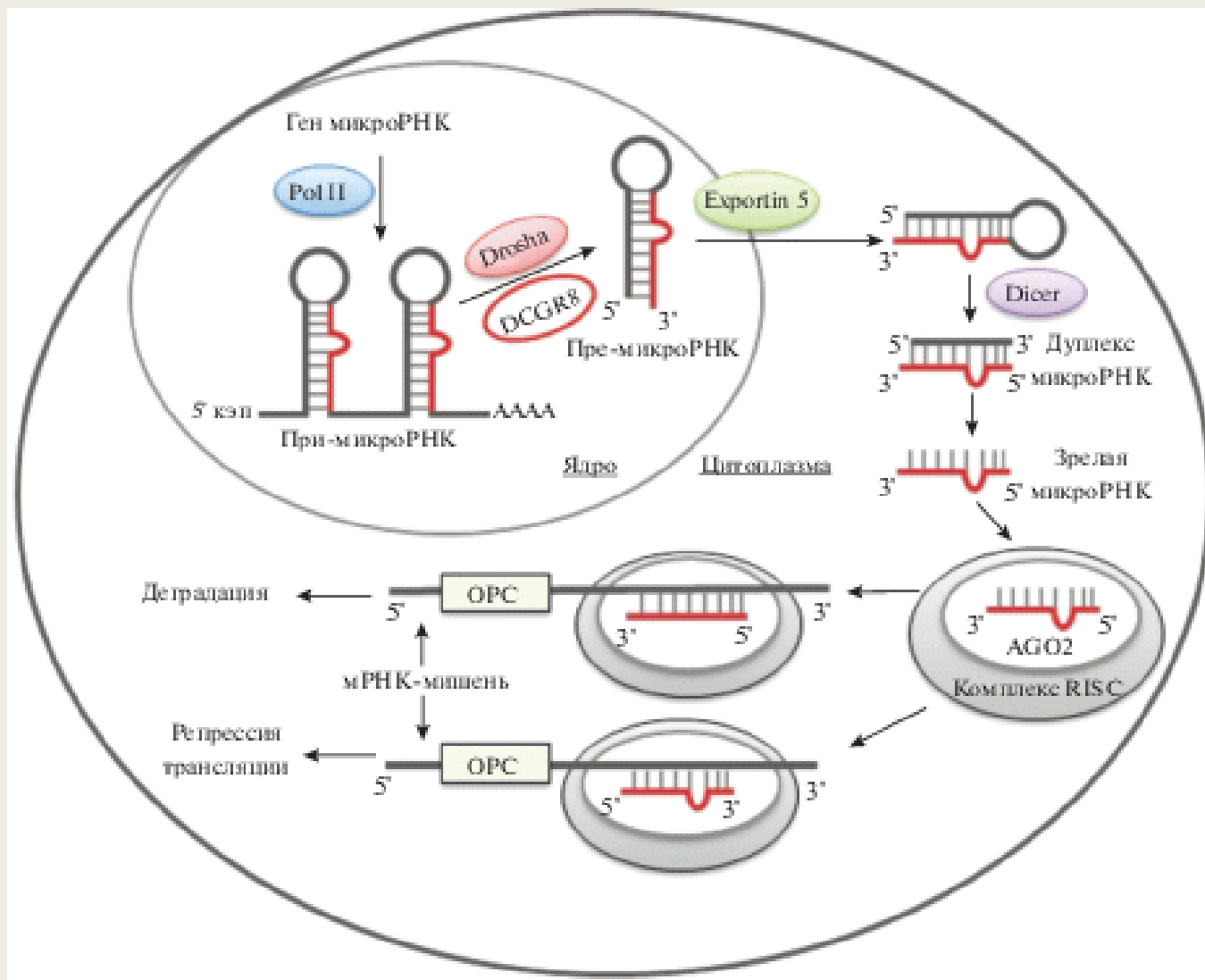
One of the most important resources for researchers of noncoding RNAs is the information available in public databases spread over the internet. However, the effective exploration of this data can represent a daunting task, given the large amount of databases available and the variety of stored data. This chapter describes a classification of databases based on information source, type of RNA, source organisms, data formats, and the mechanisms for information retrieval, detailing the relevance of each of these classifications and its usability by researchers. This classification is used to update a 2012 review, indexing now more than 229 public databases. This review will include an assessment of the new trends for ncRNA research based on the information that is being offered by the databases. Additionally, we will expand the previous analysis focusing on the usability and application of these databases in pathogen and disease research. Finally, this chapter will analyze how currently available database schemas can help the development of new and improved web resources.

Типы РНК

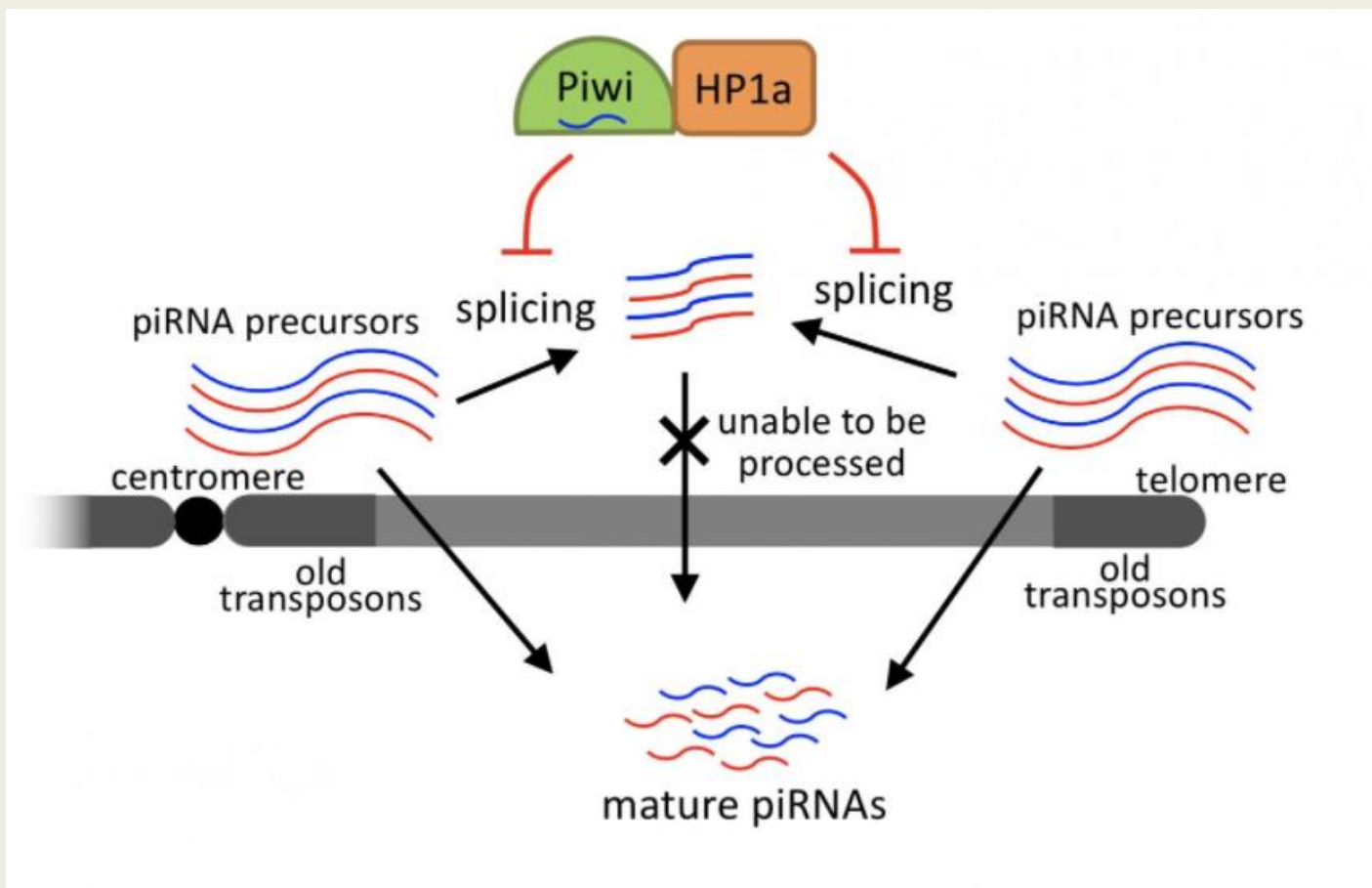


При секвенировании тотальной РНК удаляют рРНК

микроРНК

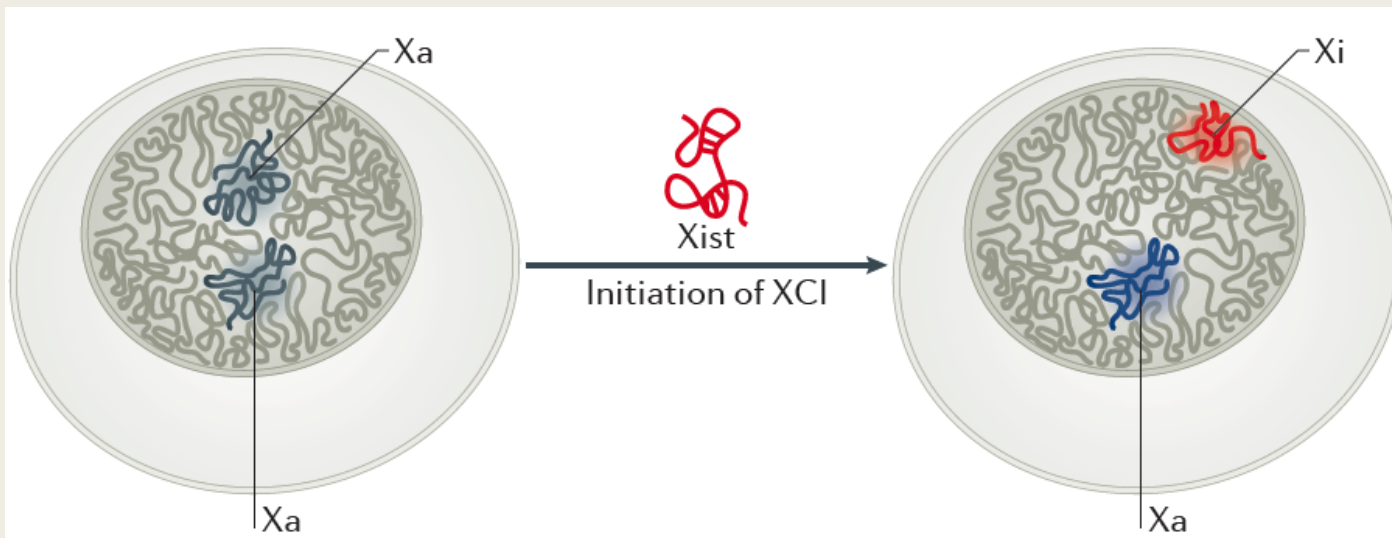
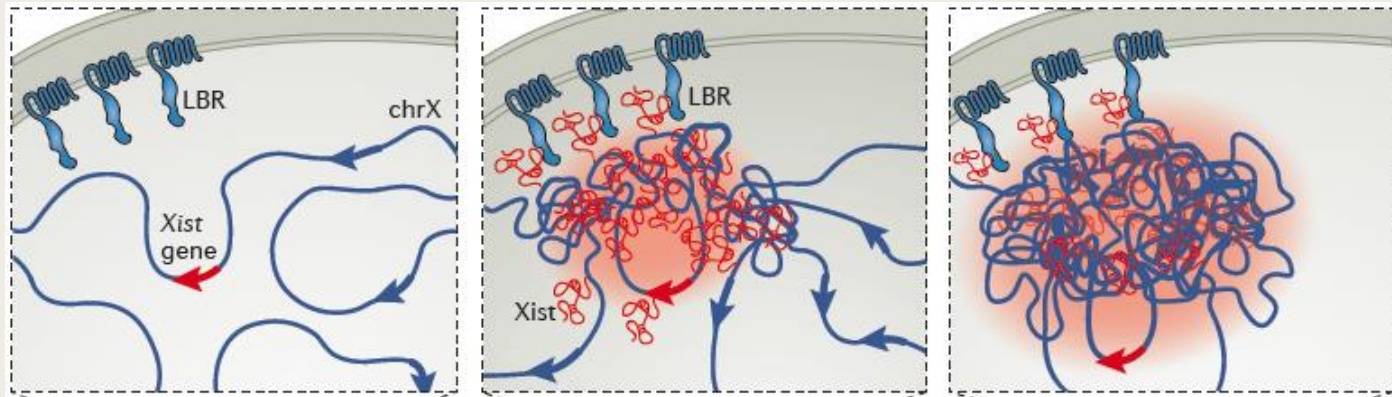


пиРНК и транспозоны



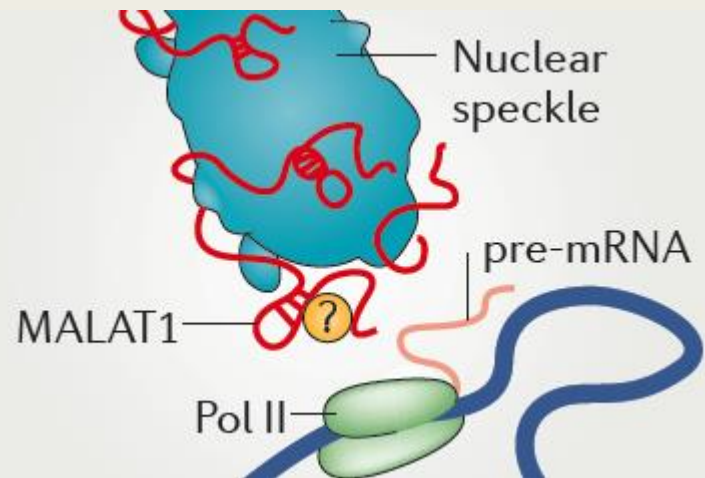
XIST

инактивация X хромосомы у самок млекопитающих

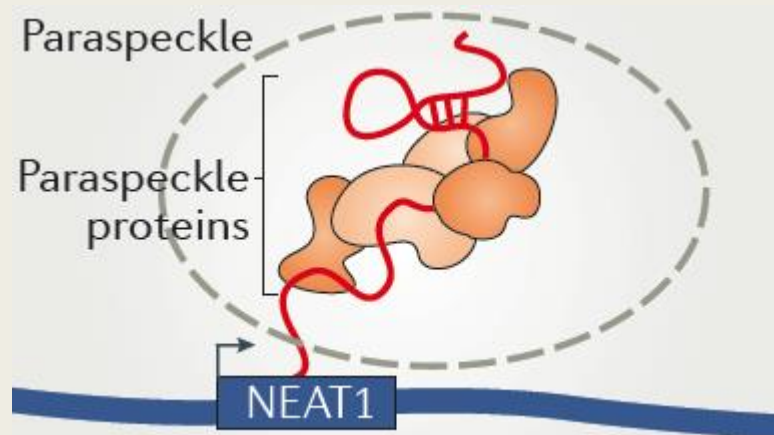


Организация ядерных структур

MALAT1 – спеклы

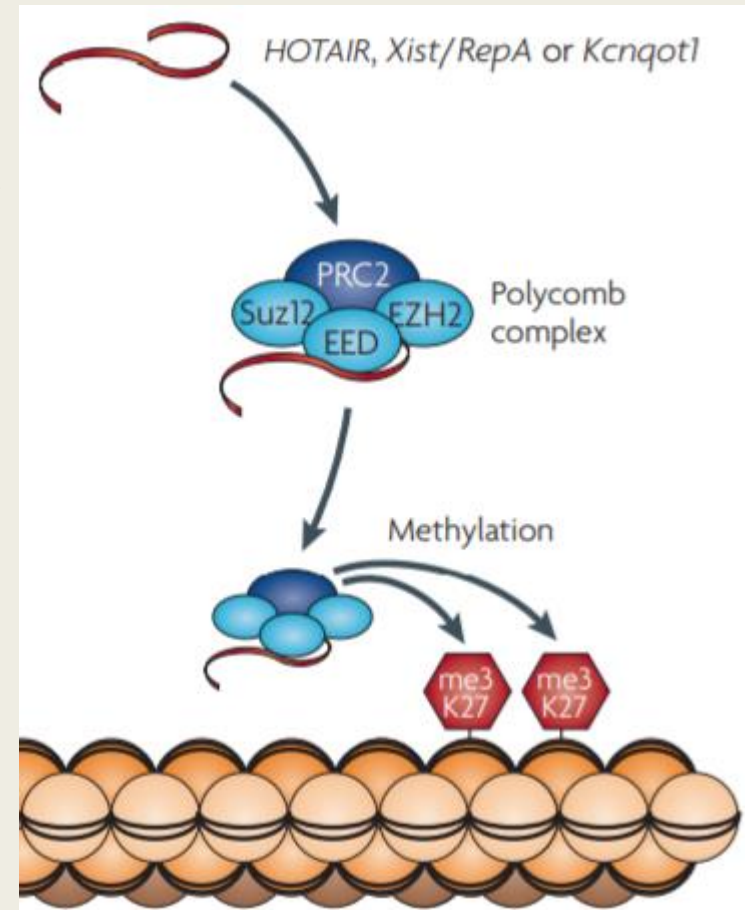


NEAT1 – параспеклы



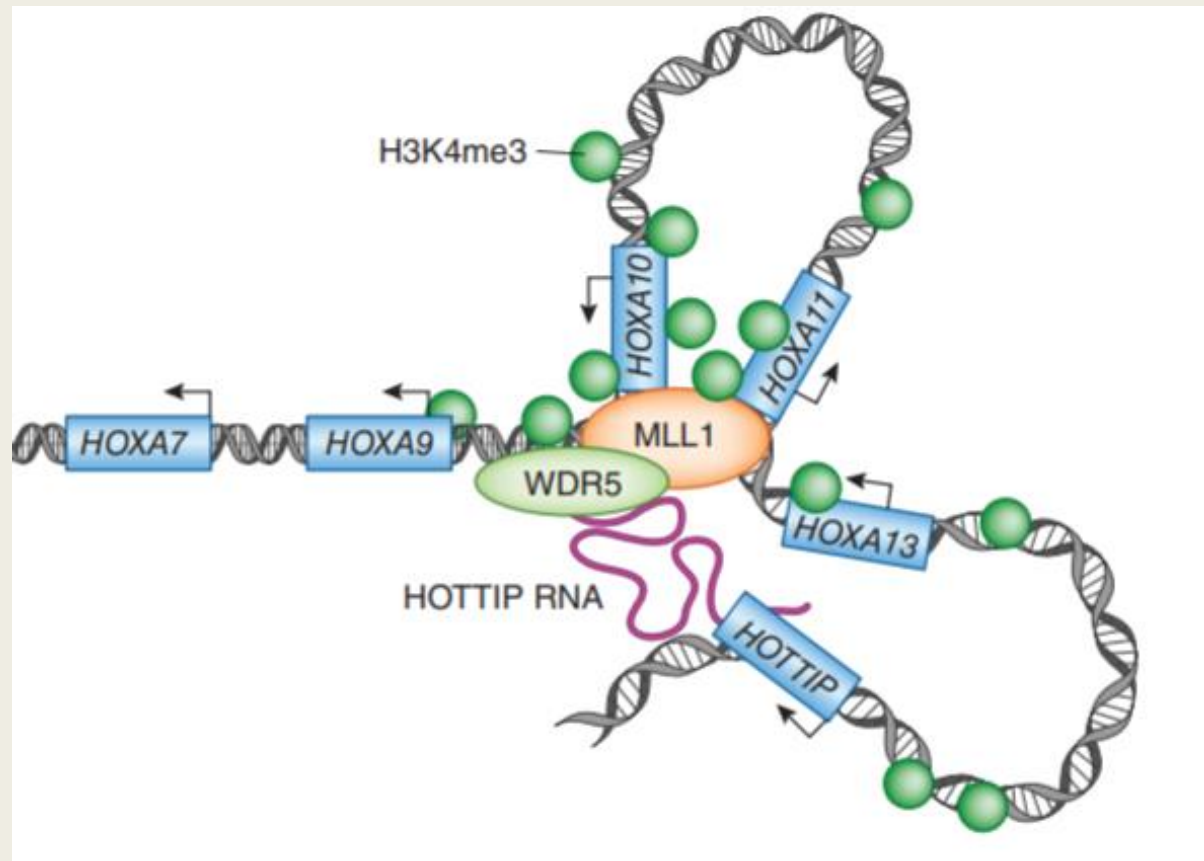
Формирование гетерохроматина

HOTAIR, Kcnqot1 –
привлечение факторов
ремоделирования хроматина

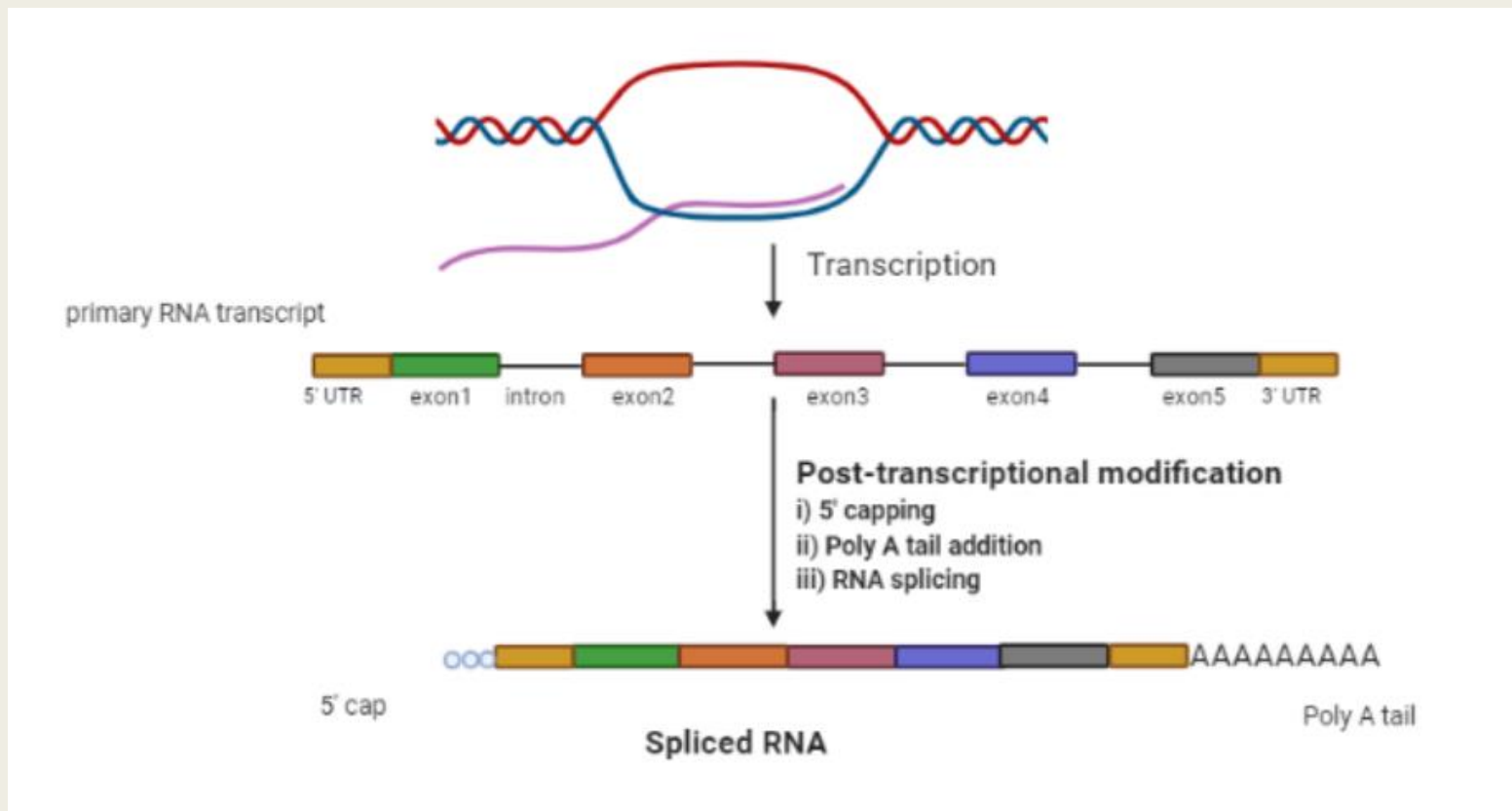


Активация транскрипции

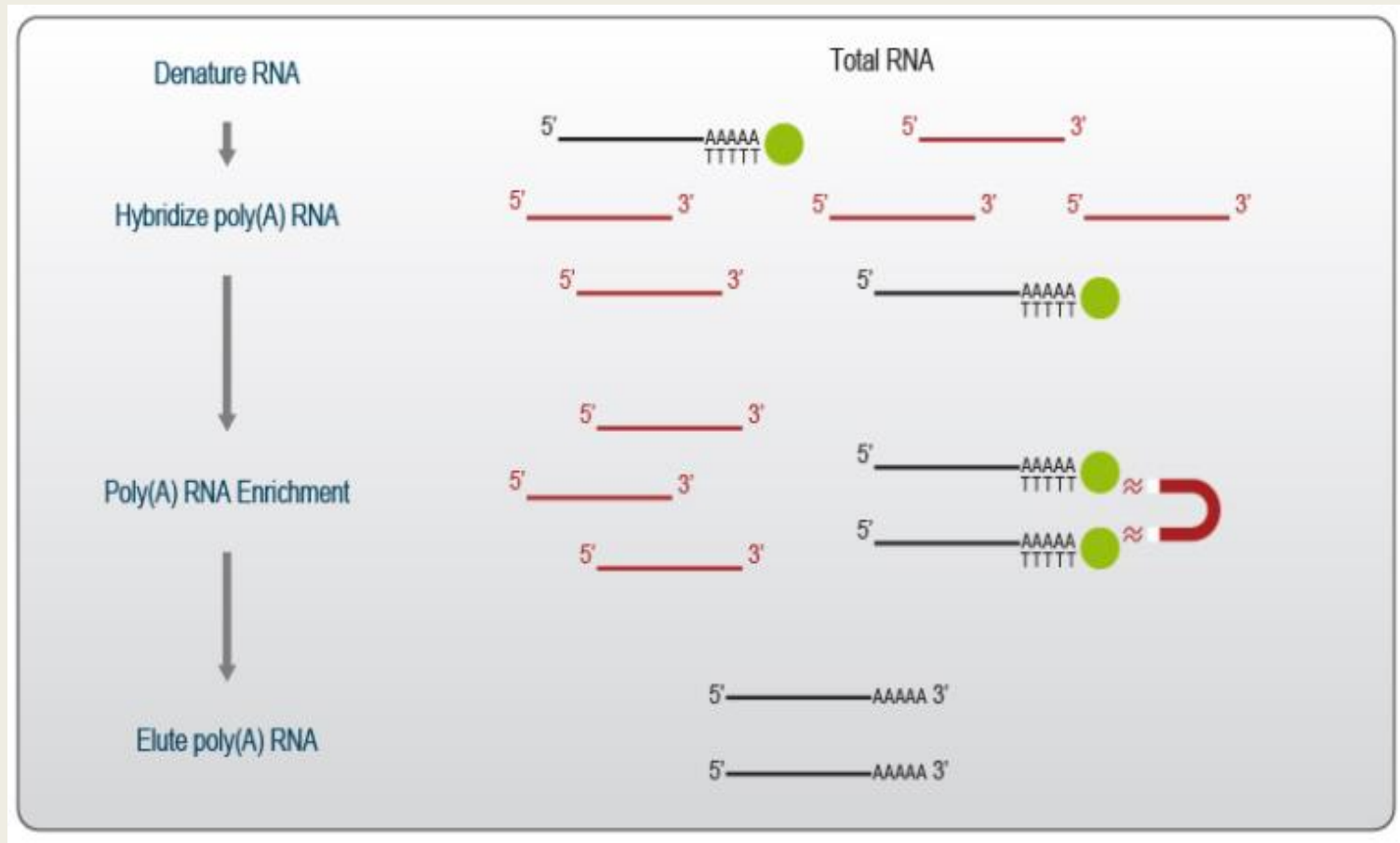
НОТТИР – активация генов *HOXA*



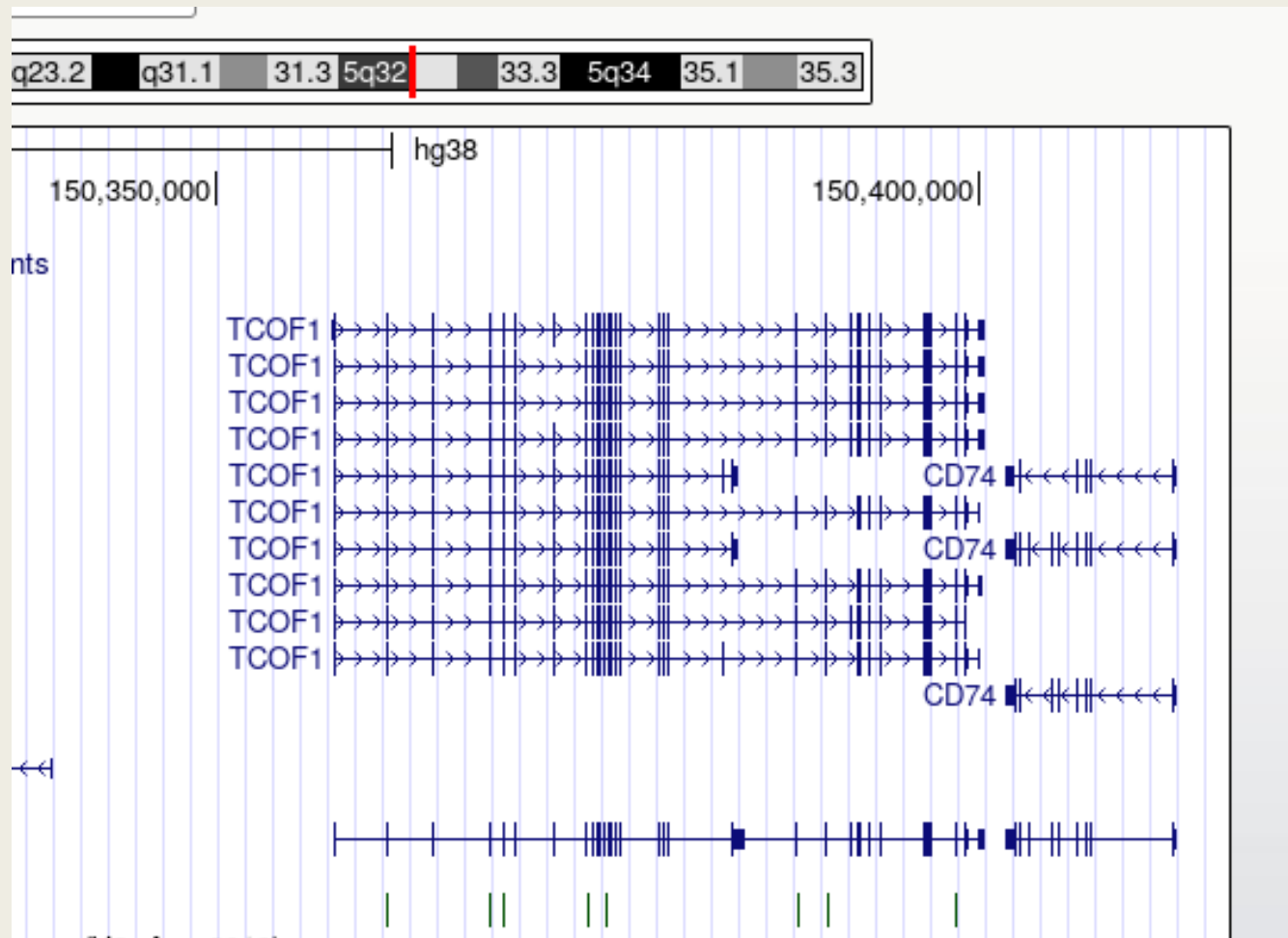
Процессинг мРНК у эукариот



Выделение фракции полиА-РНК



Ген vs транскрипт



GENECARDS



Free for academic non-profit institutions. Other users need a [Commercial license](#)



Keywords ▾

Search Term



Advanced

Home | User Guide | Analysis Tools ▾ | Release Notes | About ▾ | Data Access

My Genes | Log In / Sign Up

TCOF1 Gene - Treacle Ribosome Biogenesis Factor 1

Follow Gene ★ ✉

Protein Coding (GC05P150358 ⓘ ; GIFts: 42 ⓘ) ⓘ ⓘ

Phenotype Search

Jump to section	Aliases Paralogs	Disorders Pathways	Domains Products	Drugs Proteins	Expression Publications	Function Sources	Genomics Summaries	Localization Transcripts	Orthologs Variants
Research Products	Antibodies Cell Lines	Assays Clones	Proteins Primers	Inhib. RNA Genotyping	CRISPR	Exp. Assays	miRNA	Drugs	Animal Models

R&D	Proteins Primary Antibodies ELISAs Antibody Arrays Activity Assays	ORIGENE	Proteins Antibodies Assays Genes shRNA Primers CRISPR Lentiviral Particles	SYNTHEGO	CRISPR Knockout Kit sgRNA KO Pools iPSC SNV Clone Free Bioinformatics Tools	InVivo Biosystems	C. elegans Transgenics Zebrafish Genome Editing Humanized animal models
----------------	--	----------------	--	-----------------	---	--------------------------	---

Aliases for TCOF1 Gene

Aliases for TCOF1 Gene

GeneCards Symbol: **TCOF1** ² ⓘ

Treacle Ribosome Biogenesis Factor 1 ^{2 3 5}

Treacle ^{2 3 5}

TCS ^{2 3 5}

Treacher Collins-Franceschetti Syndrome 1 ^{2 3}

Treacher Collins Syndrome Protein ^{3 4}

Treacle Protein ^{3 4}

Nucleolar Trafficking Phosphoprotein ³

MFD1 ³

TCS1 ³

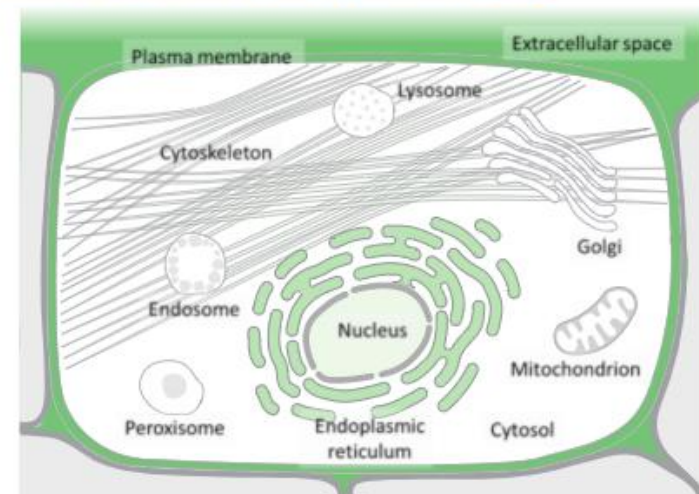
External Ids for TCOF1 Gene

HGNC: 11654 NCBI Entrez Gene: 6949 Ensembl: ENSG00000070814 OMIM®: 606847 UniProtKB/Swiss-Prot: Q13428

Previous GeneCards Identifiers for TCOF1 Gene

GC05P149806, GC05P150198, GC05P149720, GC05P149765, GC05P149717, GC05P144885

Subcellular locations from COMPARTMENTS ⓘ



Human Protein Atlas

THE HUMAN PROTEIN ATLAS

MENU HELP NEWS

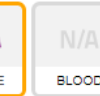
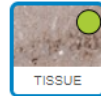
PCDHA1

Search

Fields »

Search result (5 genes): PCDHA1 | PCDHA10 | PCDHA11 ...

PCDHA1



PROTEIN SUMMARY

RNA DATA

GENE/PROTEIN

ANTIBODIES AND VALIDATION



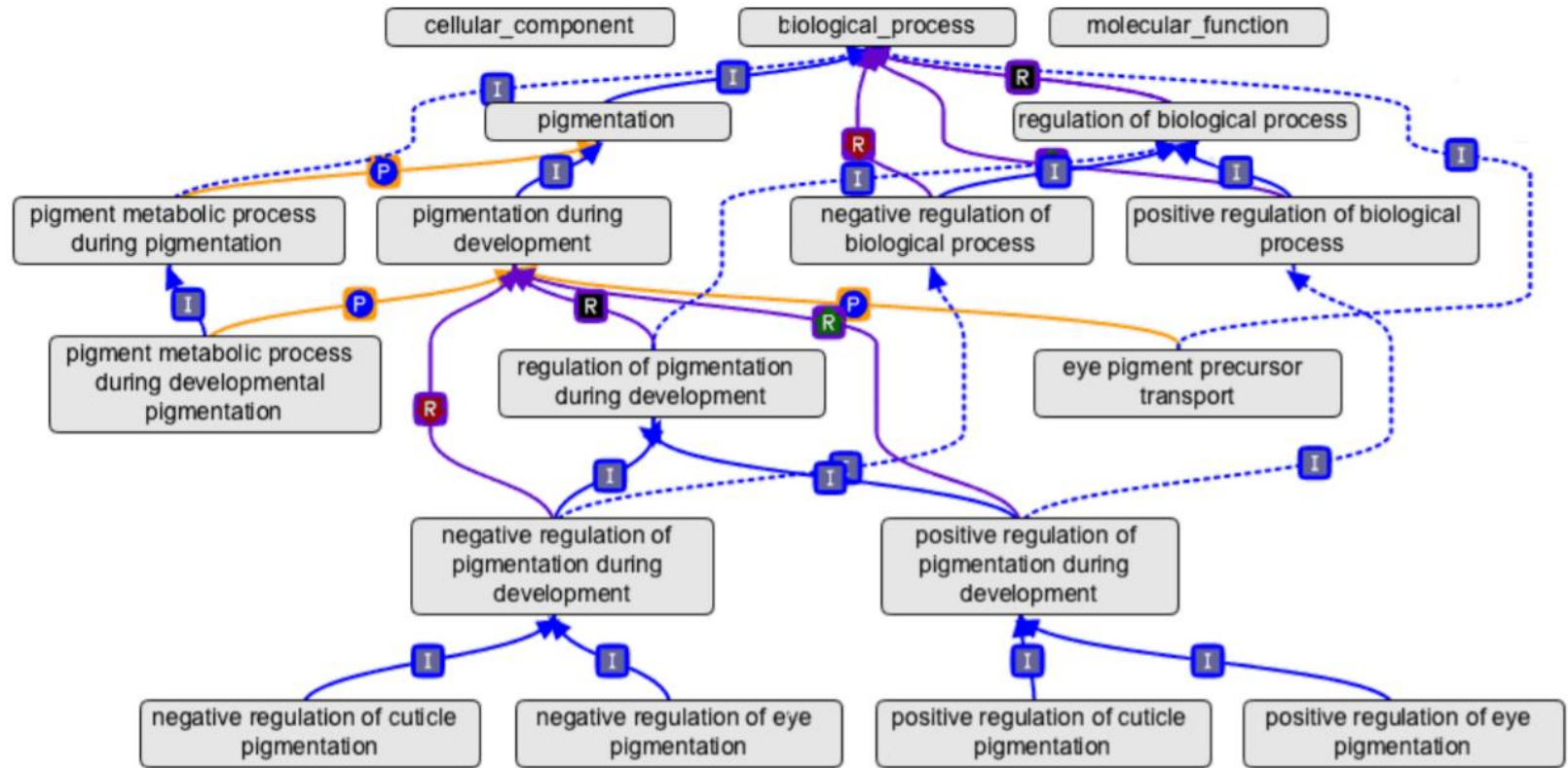
HUMAN PROTEIN ATLAS SUMMARYⁱ

Protein ⁱ	Protocadherin alpha 1
Gene name ⁱ	PCDHA1
Tissue specificity ⁱ	Tissue enhanced (brain, pituitary gland, testis)
Tissue expression cluster ⁱ	Brain - Unknown function (mainly)
Single cell type specificity ⁱ	Cell type enhanced (Sertoli cells, Alveolar cells type 1, Alveolar cells type 2, Hepatic stellate cells, Spermatogonia)
Single cell type expression cluster ⁱ	Alveolar cells - Smell perception (mainly)
Immune cell specificity ⁱ	Not detected in immune cells
Brain specificity ⁱ	Low human brain regional specificity
Cancer prognostic summary	Gene product is not prognostic
Predicted location ⁱ	Membrane, Secreted (different isoforms)
Extracellular location ⁱ	Secreted - unknown location
Subcellular summary ⁱ	Located in Plasma membrane
Protein function (UniProt) ⁱ	Potential calcium-dependent cell-adhesion protein. May be involved in the establishment and maintenance of specific neuronal connections in the brain. show less
Biological process (UniProt) ⁱ	Cell adhesion
Ligand (UniProt) ⁱ	Calcium
Gene summary (Entrez) ⁱ	This gene is a member of the protocadherin alpha gene cluster, one of three related gene clusters tandemly linked on chromosome five that demonstrate an unusual genomic organization similar to that of B-cell and T-cell receptor gene clusters. The alpha gene cluster is composed of 15 cadherin superfamily genes related to the mouse CNR genes and consists of 13 highly similar and 2 more distantly related coding sequences. The tandem array of 15 N-terminal exons, or variable exons, are followed by downstream C-terminal exons, or constant exons,

А если много генов?

- Gene Ontology
- KEGG
- Reactome
- STRING
- Panther

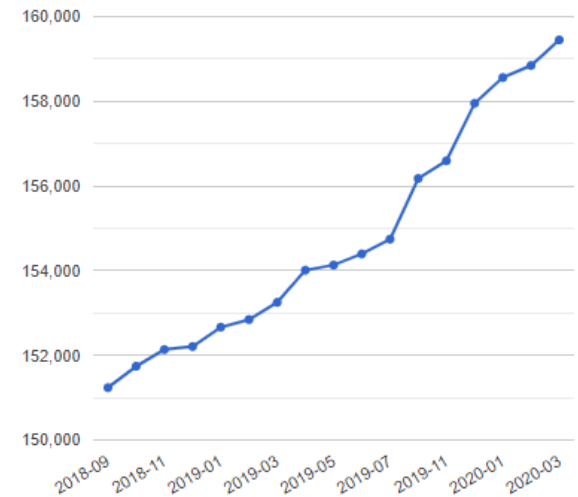
Gene Ontology - GO



База данных GO представляет собой граф биологических терминов, соединенных различными отношениями

GO

Growth of annotated scientific publications over time



Ontology

Property	Value
Valid terms	44531 ($\Delta = -48$)
Obsoleted terms	2895 ($\Delta = 40$)
Merged terms	1967 ($\Delta = 35$)
Biological process terms	29225
Molecular function terms	11124
Cellular component terms	4182

Annotations

Property	Value
Number of annotations	7,524,022
Annotations for biological process	2,927,597
Annotations for molecular function	2,325,747
Annotations for cellular component	2,270,678
Annotations for evidence PHYLO	3,610,054
Annotations for evidence IEA	2,041,235
Annotations for evidence OTHER	832,336
Annotations for evidence EXP	757,235
Annotations for evidence ND	242,830
Annotations for evidence HTP	40,332
Number of annotated scientific publications	159,450

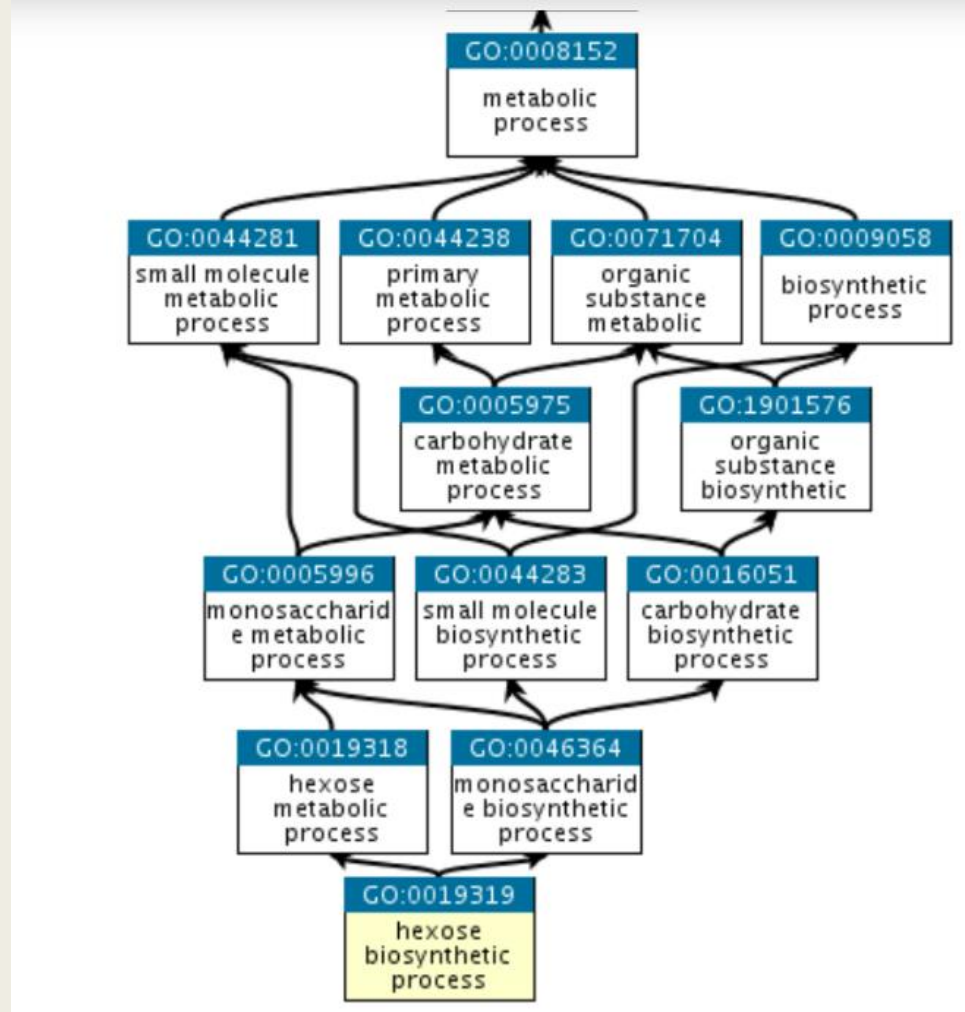
Gene products and species

Property	Value
Annotated gene products	1,405,061
Annotated species	4,593
Annotated species with over 1,000 annotations	195

GO

- **Молекулярные функции** (англ. molecular function) — специфическая активность генного продукта на молекулярном уровне, например, связывание углеводов или АТФ-азная активность.
- **Биологические процессы** (англ. biological process) — сложные явления, необходимые для жизнедеятельности организмов и происходящие благодаря осуществлению последовательности молекулярных функций, например, митоз или биосинтез пуринов.
- **Клеточные компоненты** (англ. cellular component) — части клетки или внеклеточного пространства, где осуществляется функция генного продукта, например, ядро или рибосома

GO



Типы достоверности

Inferred from Electronic Annotation (**IEA**) →

Inferred from Sequence or structural Similarity (**ISS**)

- Inferred from Sequence Orthology (**ISO**)
- Inferred from Sequence Alignment (**ISA**)
- Inferred from Sequence Model (**ISM**)

Inferred from Experiment (**EXP**)

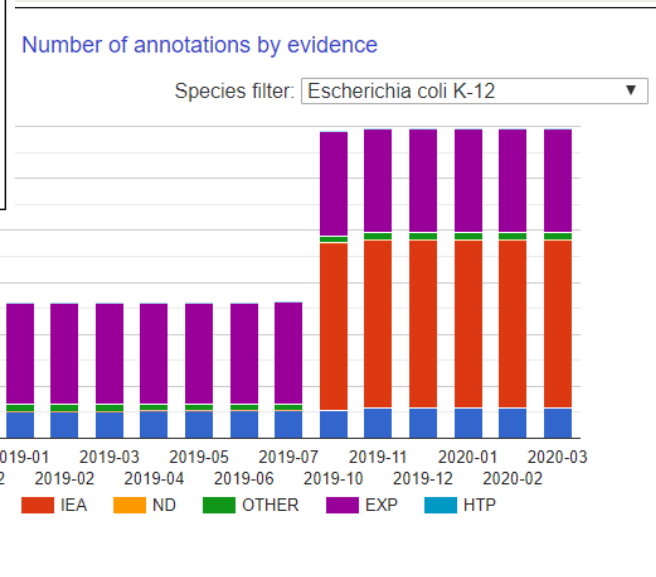
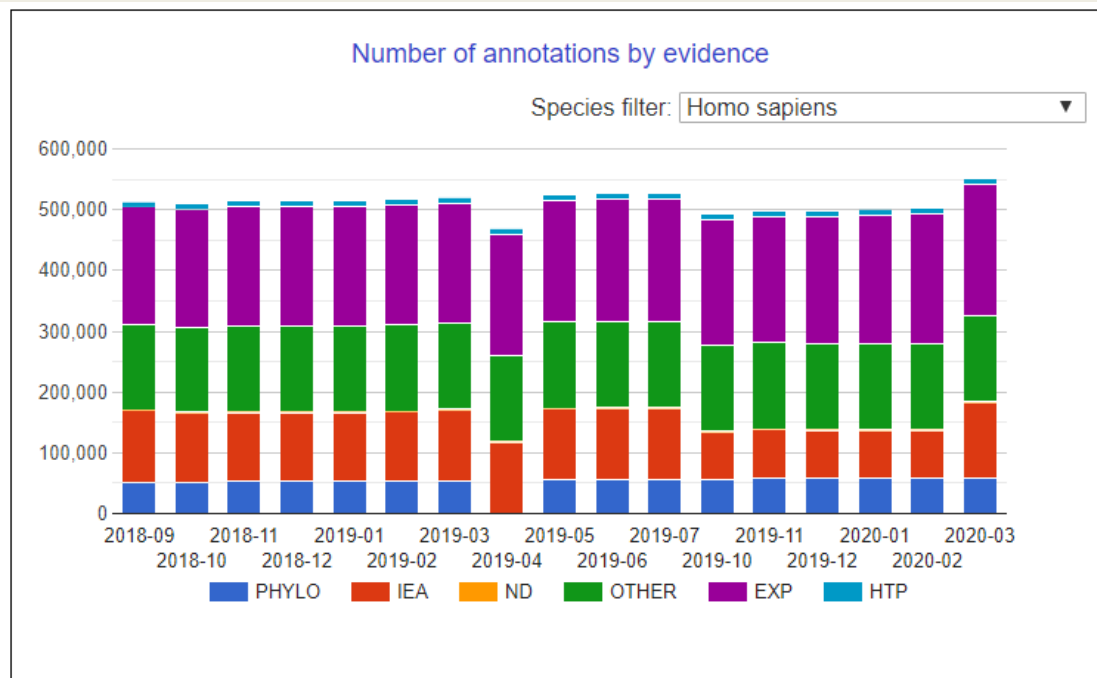
- Inferred from Direct Assay (**IDA**)
- Inferred from Physical Interaction (**IPI**)
- Inferred from Mutant Phenotype (**IMP**)
- Inferred from Genetic Interaction (**IGI**)
- Inferred from Expression Pattern (**IEP**)

Traceable Author Statement (**TAS**)

Non-traceable Author Statement (**NAS**)

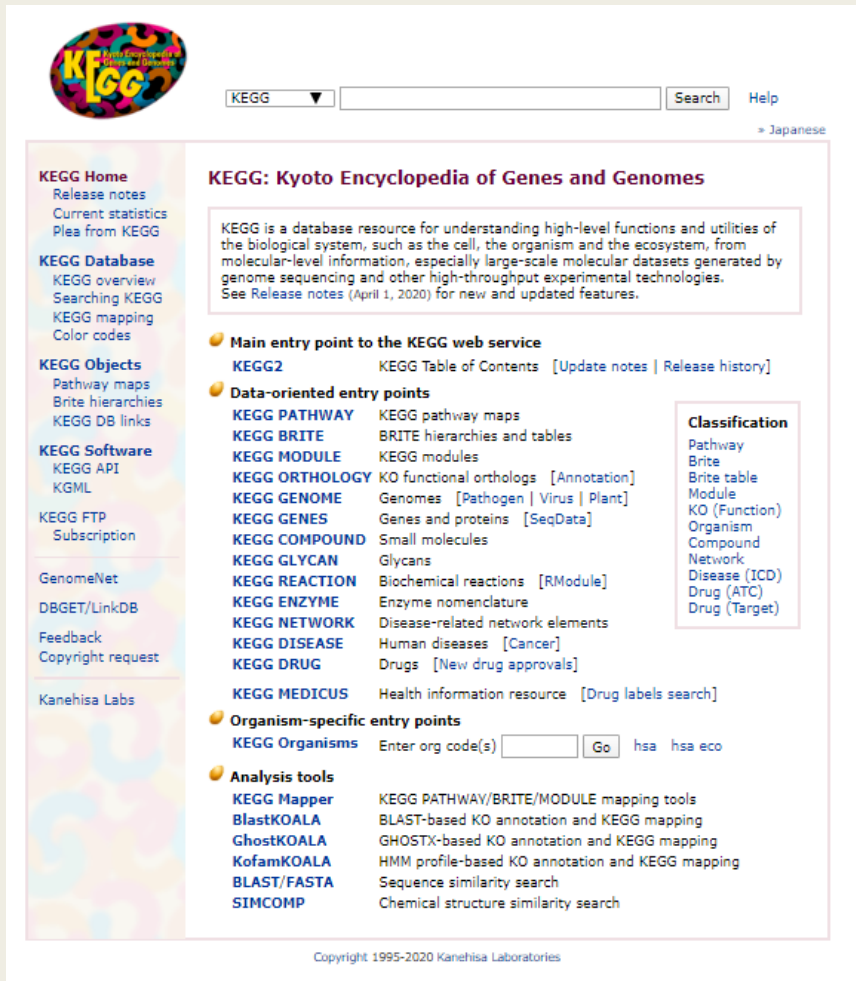


Типы достоверности

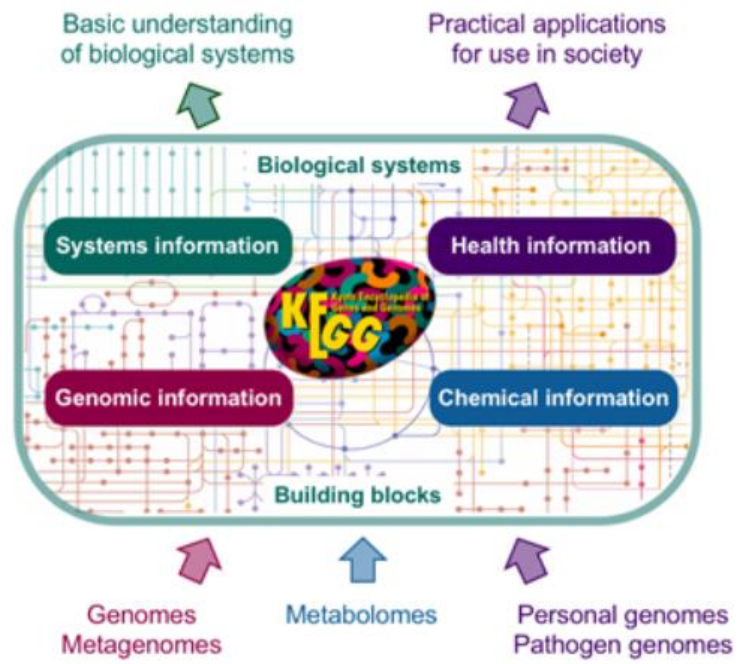


KEGG – база данных метаболических путей

https://www.genome.jp/kegg/



The screenshot shows the KEGG website interface. At the top left is the KEGG logo. Below it is a search bar with 'KEGG' entered and a 'Search' button. To the right of the search bar are links for 'Help' and '> Japanese'. The main content area is titled 'KEGG: Kyoto Encyclopedia of Genes and Genomes'. It contains a brief description of KEGG as a database resource for understanding biological systems. Below this is a list of entry points categorized into 'Main entry point to the KEGG web service', 'Data-oriented entry points', and 'Analysis tools'. A 'Classification' sidebar lists various KEGG categories like Pathway, Brite, and Module. At the bottom of the page, there is a copyright notice: 'Copyright 1995-2020 Kanehisa Laboratories'.



KEGG

Current Statistics

KEGG Database as of 2020/4/29

Systems information

KEGG PATHWAY	Pathway maps, reference (total)	537 (705,156)
KEGG BRITE	Functional hierarchies, reference (total)	200 (247,573)
KEGG MODULE	KEGG modules	394
	Reaction modules	41

Genomic information

KEGG ORTHOLOGY	KEGG Orthology (KO) groups	23,354
KEGG GENOME	KEGG organisms (540 eukaryotes, 5659 bacteria, 318 archaea)	6,517
	KEGG selected viruses	341
KEGG GENES	Genes in KEGG organisms and other categories (including 4,057 addendum, 372,625 viral) (see annotation statistics)	31,487,727
KEGG SSDB	Best hit relations within GENES	375,636,238,942
	Bi-directional best hit relations within GENES	18,772,924,969

Chemical information

KEGG COMPOUND	Metabolites and other small molecules	18,700
KEGG GLYCAN	Glycans	11,040
KEGG REACTION	Biochemical reactions	11,415
	Reaction class	3,165
KEGG ENZYME	Enzyme nomenclature	7,736

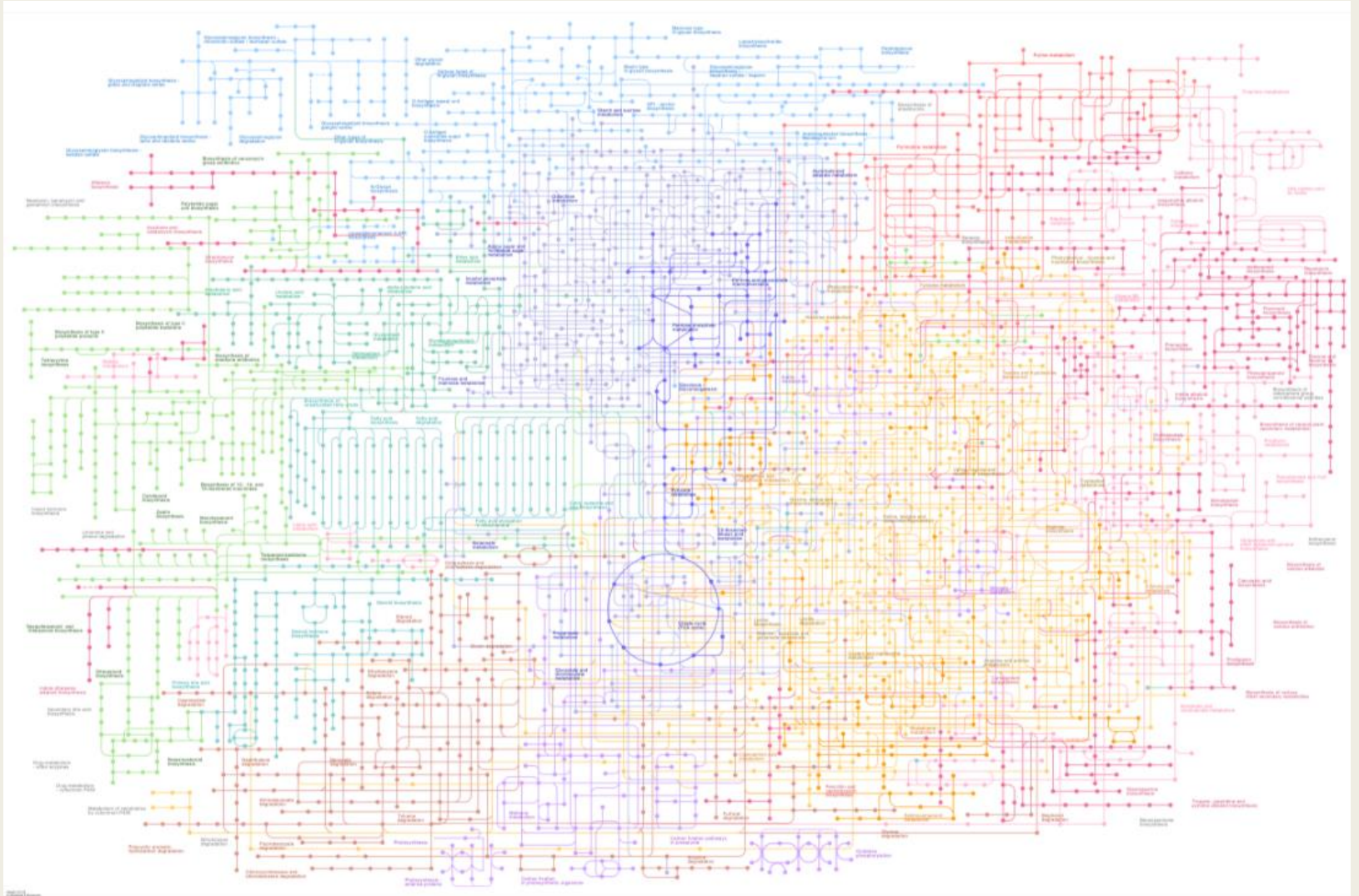
Health information

KEGG NETWORK	Disease-related network elements	1,011
	Network variation maps	114
KEGG VARIANT	Human gene variants	416
KEGG DISEASE	Human diseases	2,420
KEGG DRUG	Drugs	11,255
	Drug groups	2,276
KEGG ENVIRON	Crude drugs and health-related substances	864

Drug labels

KEGG MEDICUS	Japanese prescription drug labels from JAPIC	14,076
	Japanese OTC drug labels from JAPIC	11,011
KEGG MEDICUS	FDA prescription drug labels linked to DailyMed	27,668
	FDA OTC drug labels linked to DailyMed	36,506

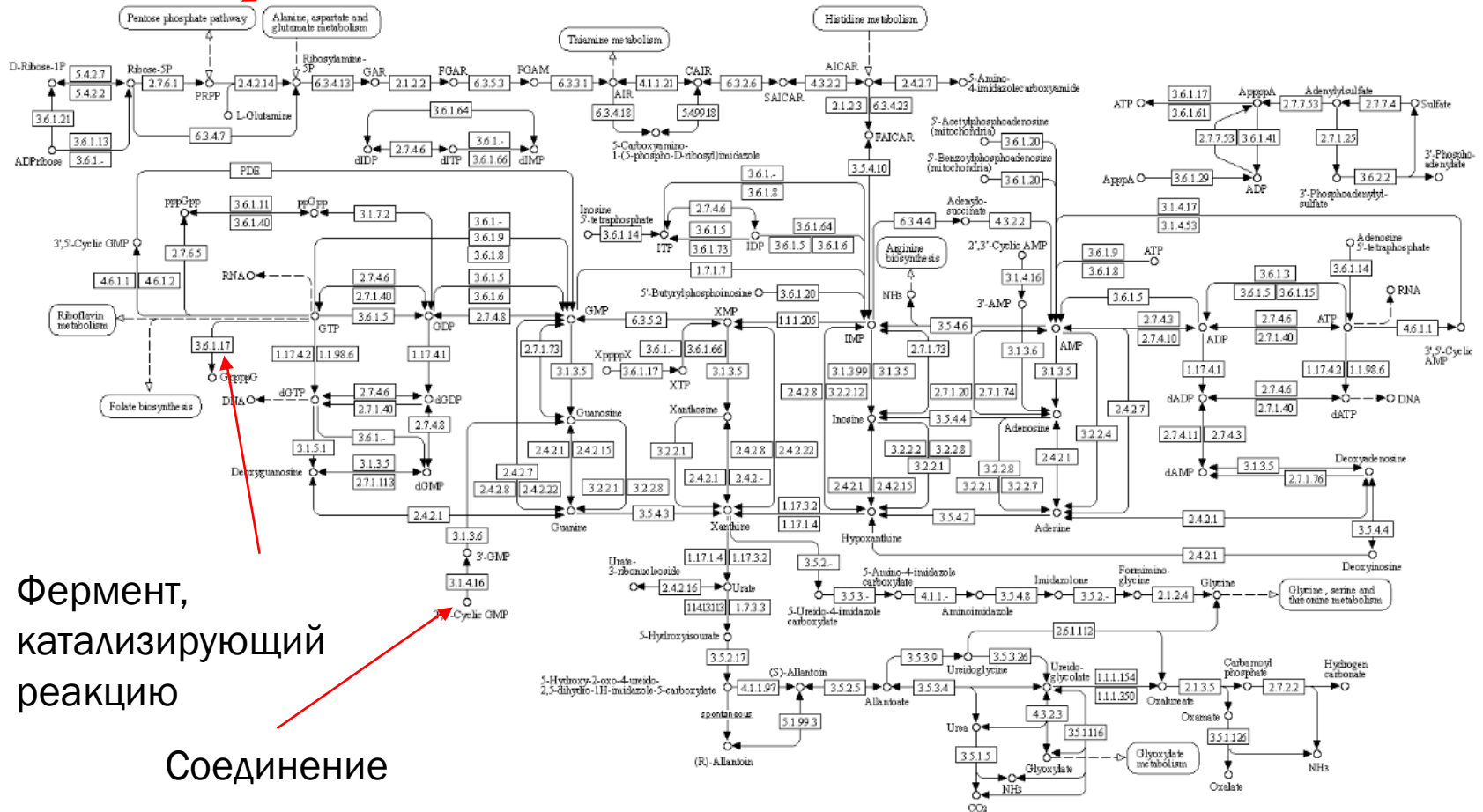
KEGG



KEGG

Другой метаболический путь

PURINE METABOLISM



Фермент,
катализирующий
реакцию

Соединение

KEGG



ENZYME: 2.7.7.4

[Help](#)

Entry	EC 2.7.7.4 Enzyme
Name	sulfate adenylyltransferase; ATP-sulfurylase; adenosine-5'-triphosphate sulfurylase; adenosinetriphosphate sulfurylase; adenylylsulfate pyrophosphorylase; ATP sulfurylase; ATP-sulfurylase; sulfurylase
Class	Transferases; Transferring phosphorus-containing groups; Nucleotidyltransferases BRITE hierarchy
Sysname	ATP:sulfate adenylyltransferase
Reaction(IUBMB)	ATP + sulfate = diphosphate + adenylyl sulfate [RN:R00529]
Reaction(KEGG)	R00529; (other) R04929 Reaction
Substrate	ATP [CPD:C00002]; sulfate [CPD:C00059]
Product	diphosphate [CPD:C00013]; adenylyl sulfate [CPD:C00224]
Comment	The human phosphoadenosine-phosphosulfate synthase (PAPS) system is a bifunctional enzyme (fusion product of two catalytic activities). In a first step, sulfate adenylyltransferase catalyses the formation of adenosine 5'-phosphosulfate (APS) from ATP and inorganic sulfate. The second step is catalysed by the adenylylsulfate kinase portion of 3'-phosphoadenosine 5'-phosphosulfate (PAPS) synthase, which involves the formation of PAPS from enzyme-bound APS and ATP. In contrast, in bacteria, yeast, fungi and plants, the formation of PAPS is carried out by two individual polypeptides, sulfate adenylyltransferase (EC 2.7.7.4) and adenylyl-sulfate kinase (EC 2.7.1.25).

All links

Pathway (16)
 KEGG PATHWAY (14)
 KEGG MODULE (2)
 Chemical substance (6)
 KEGG COMPOUND (6)
 Chemical reaction (4)
 KEGG REACTION (2)
 KEGG RCLASS (2)
 Gene (106612)
 KEGG ORTHOLOGY (5)
 KEGG GENES (8508)
 KEGG MGENES (61373)
 RefGene (36726)
 Protein sequence (77269)
 UniProt (40770)
 SWISS-PROT (457)
 RefSeq(pep) (35969)
 PDBSTR (54)
 PMD (19)
 DNA sequence (141864)
 RefSeq(nuc) (127520)
 GenBank (7103)
 EMBL (7241)
 3D Structure (24)
 PDB (24)
 Protein domain (9)
 InterPro (8)
 Pfam (1)
 All databases (325804)

[Download RDF](#)

Reactome



[About](#) [Content](#) [Docs](#) [Tools](#) [Community](#) [Download](#)

Find Reactions, Proteins and Pathways

e.g. O95631, NTN1, signaling by EGFR, glucose

Go!



Pathway Browser

Visualize and interact with Reactome biological pathways



Analysis Tools

Merges pathway identifier mapping, over-representation, and expression analysis



ReactomeFIViz

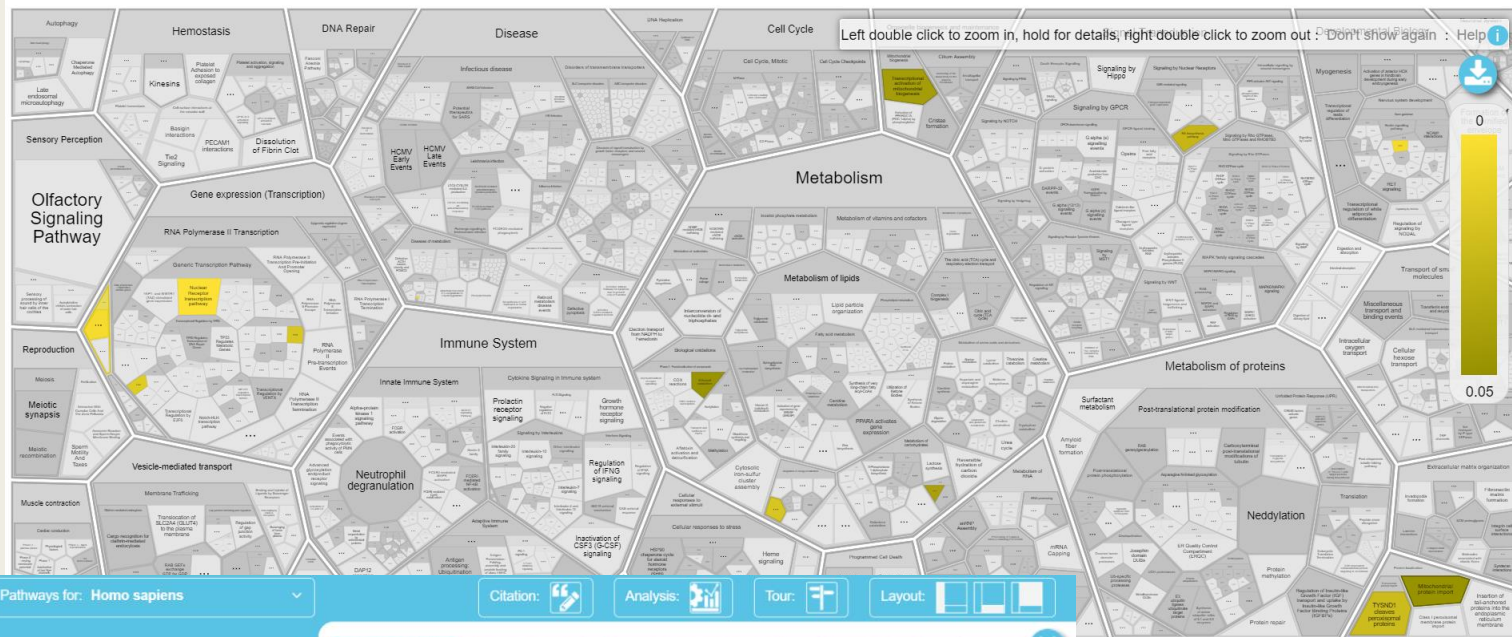
Designed to find pathways and network patterns related to cancer and other types of diseases



Documentation

Information to browse the database and use its principal tools for data analysis

Reactome



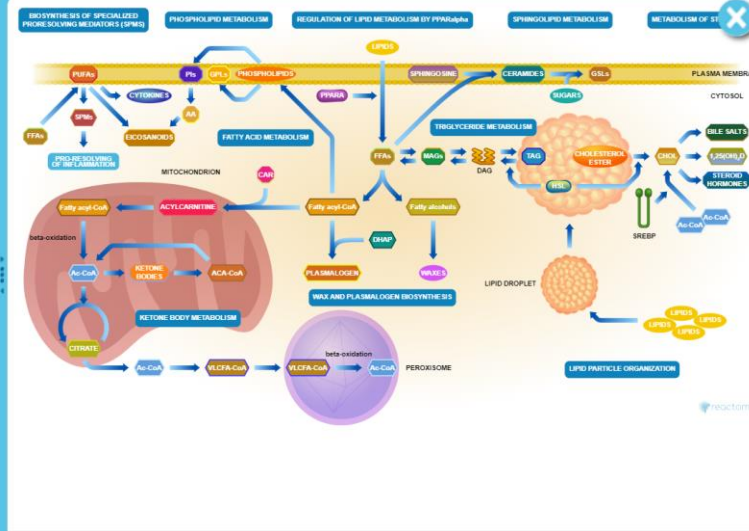
reactome 3.7 76

Pathways for: Homo sapiens

Citation: Analysis: Tour: Layout:

Event Hierarchy:

- Metabolism (773,634) FDR: 9.99E-1
 - Metabolism of carbohydrates (10/456) FDR: 8.38E-1
 - Inositol phosphate metabolism (1/90) FDR: 9.24E-1
 - Metabolism of lipids (21/1,437) FDR: 10E-1
 - Integration of energy metabolism (3/145) FDR: 7.82E-1
 - Regulation of insulin secretion (3/106) FDR: 6.62E-1
 - Glucagon signaling in metabolic regulation (2/40) FDR: 6.62E-1
 - PKA-mediated phosphorylation of key metabolic factors (2/7) FDR: 6.62E-1
 - Insulin effects increased synthesis of Xylulose-5-Phosphate
 - Activation of PP2A by Xylulose-5-phosphate
 - AMPK inhibits chREBP transcriptional activation activity
 - PP2A-mediated dephosphorylation of key metabolic factors
 - ChREBP activates metabolic gene expression
 - Metabolism of nitric oxide: NOS3 activation and regulation (1/41) FDR: 6.9E-1
 - The citric acid (TCA) cycle and respiratory electron transport (11/235) FDR: 6.62E-1
 - Metabolism of nucleotides (3/253) FDR: 9.76E-1
 - Metabolism of vitamins and cofactors (9/377) FDR: 7.48E-1
 - Metabolism of amino acids and derivatives (15/661) FDR: 8.48E-1
 - Metabolism of porphyrins (2/74) FDR: 6.62E-1
 - Biological oxidations (9/545) FDR: 9.74E-1



STRING

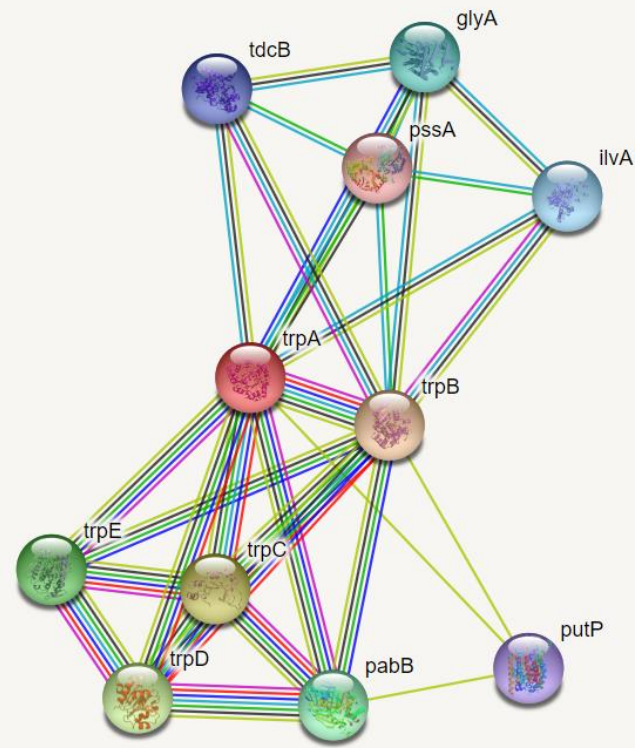


Search

Download

Help

My Data



STRING

Node Color



*colored nodes:
query proteins and first shell of interactors*



*white nodes:
second shell of interactors*

Node Content



*empty nodes:
proteins of unknown 3D structure*



*filled nodes:
some 3D structure is known or predicted*

Known Interactions



from curated databases



experimentally determined

Predicted Interactions



gene neighborhood



gene fusions



gene co-occurrence

Others



textmining



co-expression



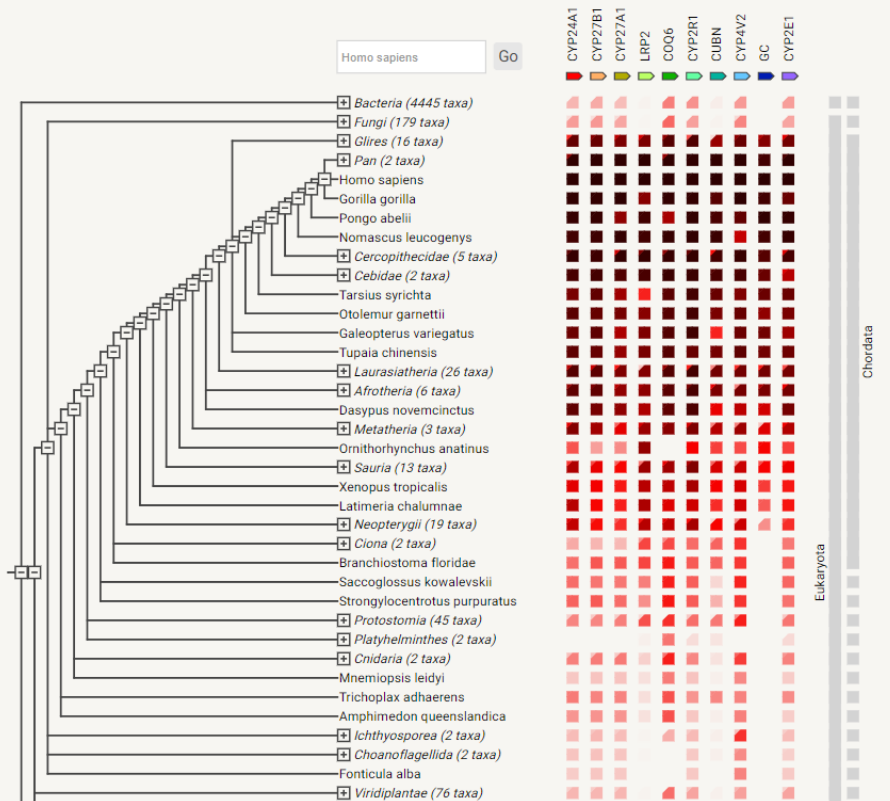
protein homology

Biological Process (Gene Ontology)	download	<i>87 GO-terms significantly enriched; file-format: tab-delimited</i>
Molecular Function (Gene Ontology)	download	<i>30 GO-terms significantly enriched; file-format: tab-delimited</i>
Cellular Component (Gene Ontology)	download	<i>5 GO-terms significantly enriched; file-format: tab-delimited</i>
Reference publications (PubMed)	download	<i>1931 publications significantly enriched; file-format: tab-delimited</i>
local network cluster (STRING)	download	<i>10 clusters significantly enriched; file-format: tab-delimited</i>
KEGG Pathways	download	<i>20 pathways significantly enriched; file-format: tab-delimited</i>
Reactome Pathways	download	<i>2 pathways significantly enriched; file-format: tab-delimited</i>
Annotated Keywords (UniProt)	download	<i>16 keywords significantly enriched; file-format: tab-delimited</i>
Protein Domains (Pfam)	download	<i>18 domains significantly enriched; file-format: tab-delimited</i>
Protein Domains and Features (InterPro)	download	<i>32 domains significantly enriched; file-format: tab-delimited</i>
Protein Domains (SMART)	download	<i>2 domains significantly enriched; file-format: tab-delimited</i>

STRING

Viewers ▾ | Legend > | Settings > | Analysis > | Table > | More + | Less -

- Network**
Summary view: shows current interactions. Nodes can be moved; popups provide information on nodes & edges.
- Cooccurrence**
Gene families whose occurrence patterns across genomes show similarities.
- Experiments**
Co-purification, co-crystallization, Yeast2Hybrid, Genetic Interactions, etc ... as imported from primary sources.
- Coexpression**
Proteins whose genes are observed to be correlated in expression, across a large number of experiments.
- Databases**
Known metabolic pathways, protein complexes, signal transduction pathways, etc ... from curated databases.
- Neighborhood**
Groups of genes that are frequently observed in each other's genomic neighborhood.
- Textmining** currently showing
Automated, unsupervised textmining - searching for proteins that are frequently mentioned together.
- Fusion**
Genes that are sometimes fused into single open reading frames.



Panther

