

Bioinformática

Curso de Especialização na Área da Saúde – 2025/2026
Bases Técnicas em Vacinas e Biofármacos

PhD Flavio Lichtenstein

Bioinformatics, Systems Biology, and Biostatistics

Instituto Butantan – CENTD - Bioinformática



abril/2025 ~ maio/2023

O que vimos até aqui?

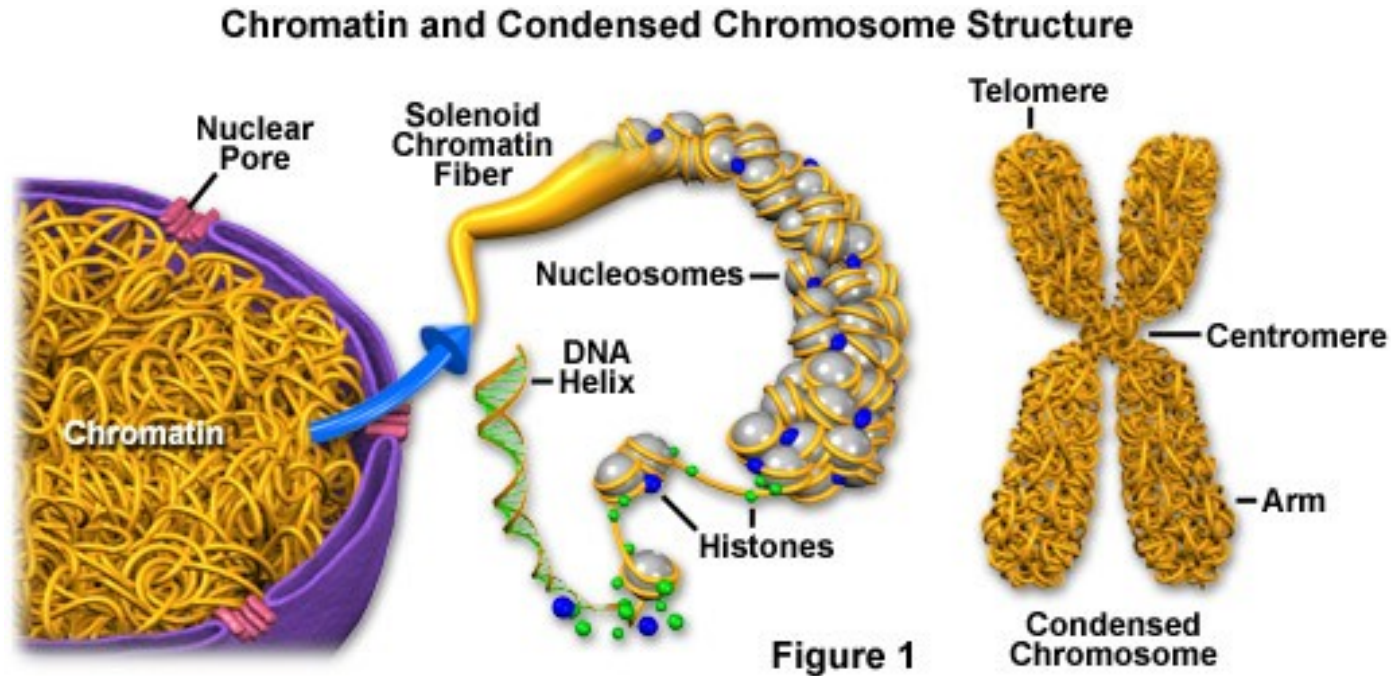
Constituição de um organismo eucarioto

- O corpo é constituído de órgãos
- Os órgãos não são homogêneos e divididos em diferentes tecidos
- Os tecidos são constituídos de células
- Há diferentes tipos de células
- As células têm vários organelas
 - Núcleo
 - Mitocôndria
 - Retículo endoplasmático
 - Lisossomo
 - Endossomo
 - Vacúolos
 - Golgi
 - etc

O que não vimos!

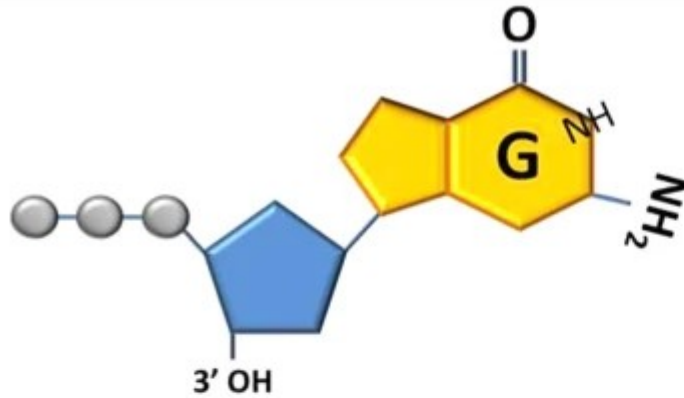
- O cromossomo é uma fita contendo material genético
- Os cromossomos estão enovelados dentro do núcleo
- Cada organismo (homem, coelho, abelha, etc) tem um número de cromossomos diferente
- O cromossomo tem milhões de nucleotídeos
- O cromossomo é formado de 2 fitas em forma de hélice
- Eles se pareiam via interações fracas ligando a fita dupla de DNA.

Os cromossomos estão enovelados dentro do núcleo

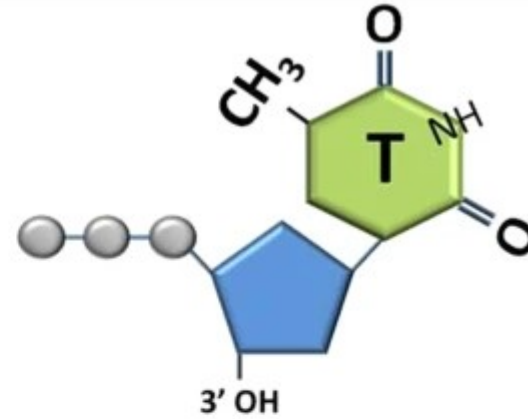


Os quatro nucleotídeos existentes no cromossomo

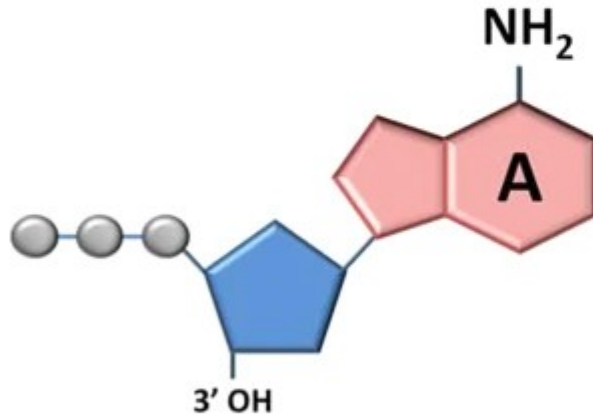
Guanina



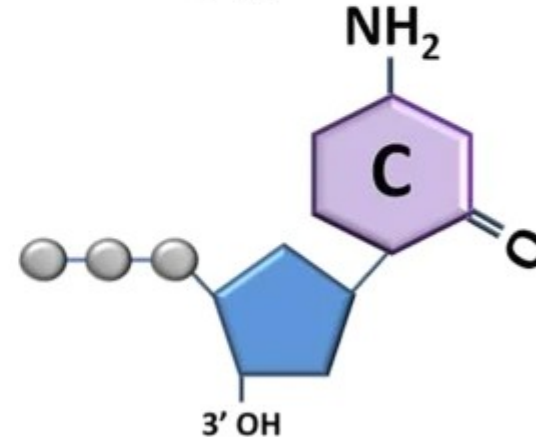
Timina



Adenina



Citosina

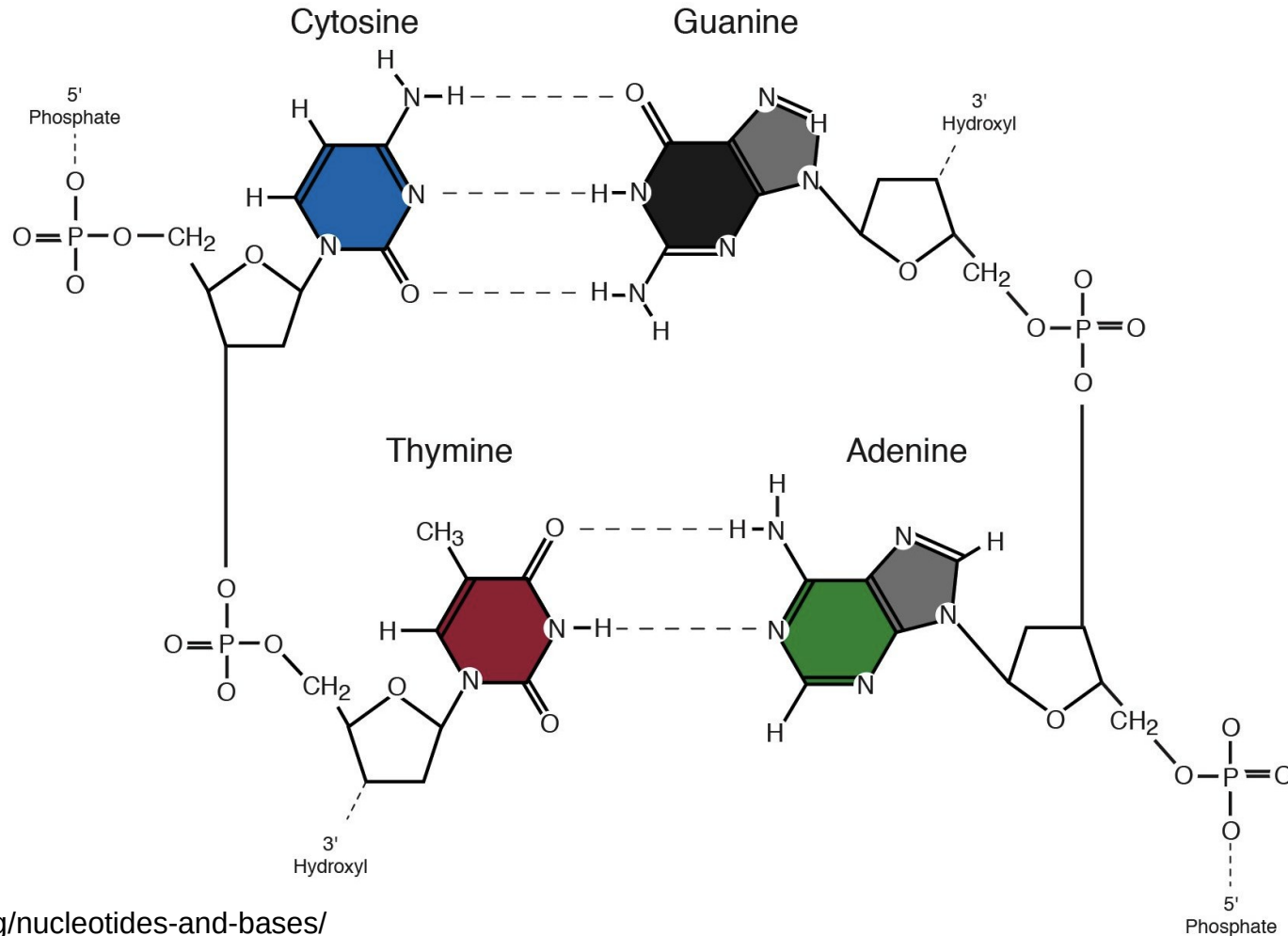


Purine
Deoxyribonucleotides

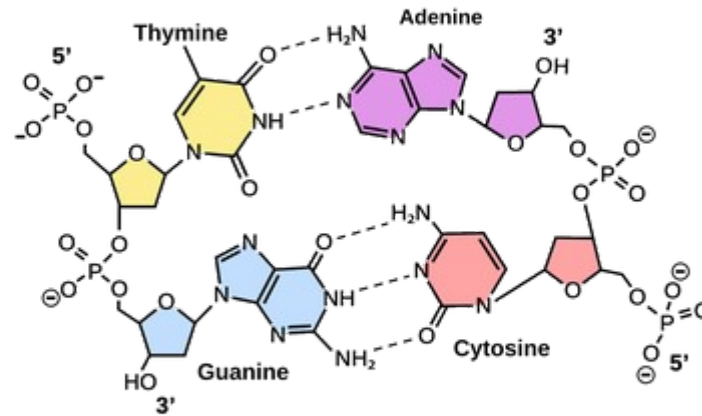
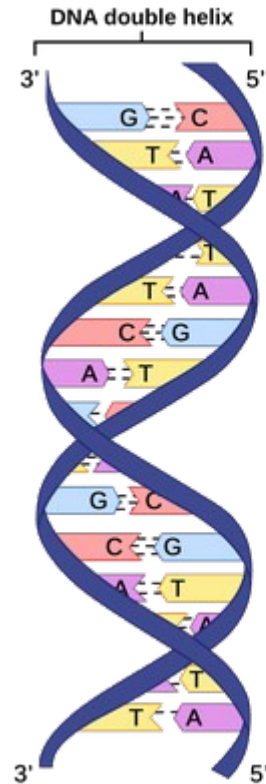
Pyrimidine
Deoxyribonucleotides

Ahmad, M., Panicker, N., Rizvi, T. et al. Electrical detection and quantification of single and mixed DNA nucleotides in suspension. Sci Rep 6, 34016 (2016).
<https://doi.org/10.1038/srep34016>

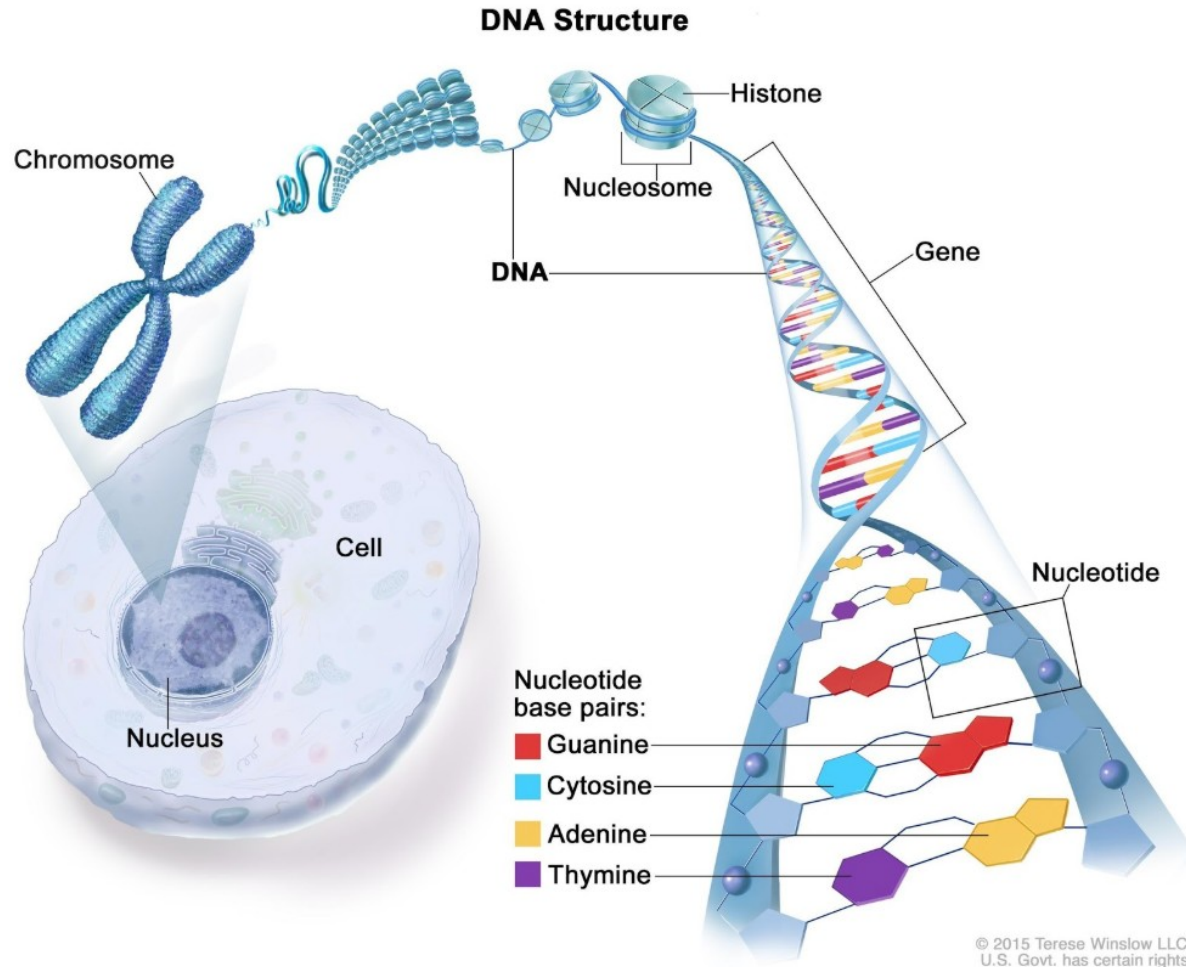
A interação fraca (ponte de hidrogênio) que segura as duas fitas de DNA



Fita dupla de DNA ~ dupla hélice



Hierarquia do cromossomo



Como estudar tudo isto?

Áreas de estudo

- Bioquímica & Química
- Biofísica
- Biologia Molecular
- Biologia Celular
- Genética
- ...
- e onde entra a matemática?

Biologia x Matemática

- Biomatemática (mathematical biology) ❌
- Matemática aplicada à Biologia de evolução ~ populações ✅
- Bioinformática (ômicas) ✅
- Imunoinformática ✅
- Biologia de Sistemas (redes – networks) ✅
- Bioestatística ✅
- Engenharia e Matemática aplicada a medicina ❌

exercício

Qual o comprimento do genoma mitocondrial?

Quantos genes existem no genoma mitocondrial?

O que é transferência horizontal e vertical de genes?

~17 mil pares de bases
num genoma circular de fita dupla

Qual o comprimento do genoma mitocondrial?

Quantos genes existem no genoma mitocondrial?

O que é transferência horizontal e vertical de genes?

37 genes, dos quais 22 são tRNA. 2 são rRNA e 13 genes são para a cadeia de transporte eletrônico da Mt

Qual o comprimento do DNA mitocondrial?

Quanto genes existem no genoma mitocondrial?

O que é transferência horizontal e vertical de genes?

A transferência horizontal de genes (THG), também conhecida como transferência lateral de genes, refere-se ao movimento de informações genéticas de um organismo para outro, que não são seus descendentes diretos. Em outras palavras, é uma forma de troca de genes entre organismos que não têm uma relação de pai-filho. Este processo é fundamental na evolução de bactérias, arqueas e, em menor medida, em eucariotas.

Quanto genes existem no genoma mitocondrial?

O que é transferência horizontal e vertical de genes?

A transferência vertical de genes refere-se à transmissão de informação genética de uma geração para a seguinte, ou seja, de pais para filhos, através da reprodução sexual ou assexuada. É um processo natural e fundamental na evolução das espécies, onde um organismo passa parte ou todo o seu genoma para os seus descendentes.

Quanto genes existem no genoma mitocondrial?

O que é transferência horizontal e vertical de genes?

O que é bioinformática?

O que é imunoinformática?

O que é bioinformática

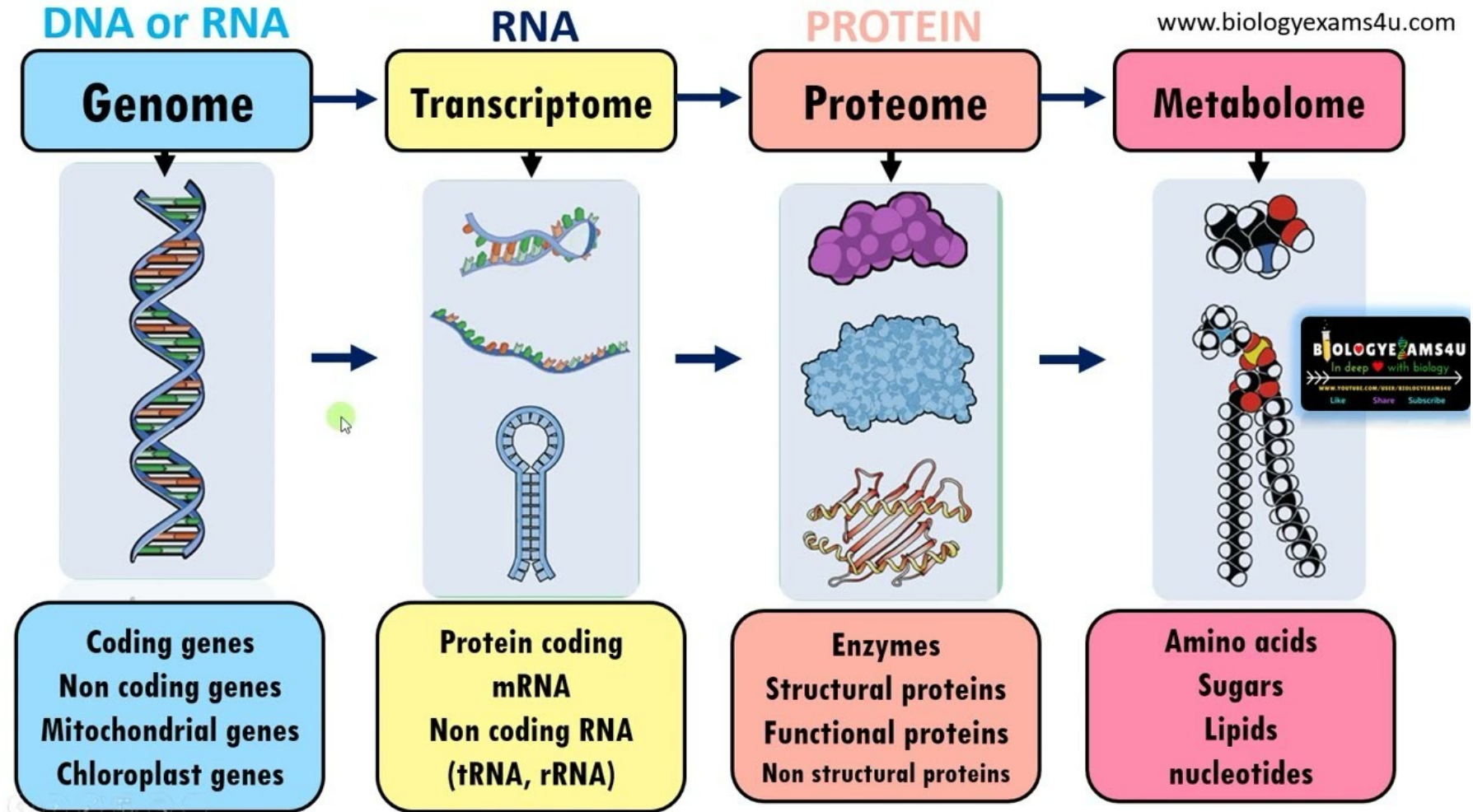
Bioinformática é o estudo de algoritmos matemáticos, aprendizado de máquina e bioestatística aplicados para melhor compreender sequências de nucleotídeos e proteínas e os resultados de experimentos das ômicas.

O que são ômicas? São estudos e experimentos relacionados a:

- Genômica
- Exômica
- Transcriptômica
- Epigenética
- Proteômica
- Metabolômica
- Single-cell



Qual a diferença entre Genoma, Transcriptoma, Proteoma and Metaboloma?



... me contaram que o COX2 é um gene interessante, será?

... me contaram que o COX2 é um gene interessante, será?

O que posso fazer?

Busca de genes / proteínas

Posso perguntar:

- Onde tem uma boa descrição deste gene?
- Em que vias ele participa?
- Como posso localizá-lo?

Busca de genes / proteínas

Posso perguntar:

- Onde tem uma boa descrição deste gene?

Genecards <https://www.genecards.org/>

- Em que vias ele participa?

KEGG: procure no google “COX2 pathway KEGG”

https://www.genome.jp/dbget-bin/www_bget?dha:cox2

- Como posso localizá-lo?

UCSC genome browser <https://genome.ucsc.edu/>

COX2 na verdade é MT-CO2

Genecard

genecards.org/cgi-bin/carddisp.pl?gene=MT-CO2&keywords=COX2

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

WEIZMANN INSTITUTE OF SCIENCE LifeMap SCIENCES

Search GeneCards (supports boolean, parenthesis and quotes) [Advanced](#)

Home | User Guide | Analysis Tools | Release Notes | About | Data Access | GeneCards Team | My Genes | Log In / Sign Up

MT-CO2 Gene - Mitochondrially Encoded Cytochrome C Oxidase II

Protein Coding (Updated: Mar 21, 2023 ; GCMP007587 ; GIFTS: 40)

Search in Gene [Follow Gene](#)

Jump to section	Aliases Paralogs	Disorders Pathways	Domains Products	Drugs Proteins	Expression Publications	Function Sources	Genomics Summaries	Localization Transcripts	Orthologs Variants
-----------------	---------------------	-----------------------	---------------------	-------------------	----------------------------	---------------------	-----------------------	-----------------------------	-----------------------

R&D Proteins Primary Antibodies
ELISAs Antibody Arrays
Activity Assays

VectorBuilder Online Vector Design Platform
Virus Packaging (AAV/Lenti)
CRISPR Library Construction

ORIGENE Proteins Antibodies Assays
Genes shRNA Primers
CRISPR Lentiviral Particles

SYNTHEGO CRISPR Knockout Kit sgRNA
KO Pools iPSC SNV Clone
Free Bioinformatics Tools

Aliases for MT-CO2 Gene

Aliases for MT-CO2 Gene

GeneCards Symbol: **MT-CO2** ² [?]

Mitochondrially Encoded Cytochrome C Oxidase II ² ⁵

COX2 ² ³ ⁴ ⁵

MTCO2 ³ ⁴ ⁵

COII ³ ⁴

CO2 ² ⁵

Cytochrome C Oxidase Polypeptide II ⁴

Cytochrome C Oxidase Subunit II ³

Cytochrome C Oxidase Subunit 2 ⁴

Cytochrome C Oxidase II ²

EC 7.1.1.9 ⁴

EC 1.9.3.1 ⁴⁸

COXII ⁴

External Ids for MT-CO2 Gene

HGNC: 7421 NCBI Entrez Gene: 4513 Ensembl: ENSG00000198712 UniProtKB/Swiss-Prot: P00403

GeneALaCart
GENECARDS BATCH QUERIES

> 190 Integrated Biomedical Sources

API, JSON, CSV, EXCEL



<https://www.genecards.org/>

Entry	cox2 CDS T01026
Name	(RefSeq) cytochrome c oxidase subunit 2
KO	K02261 cytochrome c oxidase subunit 2
Organism	dha Debaryomyces hansenii
Pathway	dha00190 Oxidative phosphorylation dha01100 Metabolic pathways
Module	dha_M00154 Cytochrome c oxidase
Brite	KEGG Orthology (KO) [BR:dha00001] 09100 Metabolism 09102 Energy metabolism 00190 Oxidative phosphorylation cox2 09180 Brite Hierarchies 09182 Protein families: genetic information processing 03029 Mitochondrial biogenesis [BR:dha03029] cox2 Mitochondrial biogenesis [BR:dha03029] Mitochondrial DNA transcription, translation, and replication fac Mitochondrial DNA-encoded proteins Cytochrome c oxidase cox2
SSDB	Ortholog Paralog Gene cluster GFIT
Motif	Pfam: COX2 COX2_TM DUF6169 Motif
Other DBs	NCBI-GeneID: 5845854 NCBI-ProteinID: YP_001621413 UniProt: A9RAG1
Position	MT:2603..3343 Genome browser
AA seq	246 aa AA seq DB search MIWTDVPTPWGMRFQDAATPNAEGMHLYDHMMYYLALMLGLVSYMLYVMMKDYKNNTFA YKYIKHGQTLIIMWTFPAVMLLLMAFP SFMLLYLCDEVLTPAMTVKVVGLQWYWKYEYS DFVSETGETVEYESYVMPEDMLEEGQLRLLDTDTSMVVPVDTHVRFMVTANDVLHCFTMP SLGIKVDACPGRLNQVSALMQRTGVYYGQCSELCGVNHGLMPIKTECVPIGDFVEWLGEQ ENVYVA
NT seq	741 nt NT seq +upstream 0 nt +downstream 0 nt atgatttgactagatgtaccaacaccttgaggaatagctttccaagatgccgcaacacct

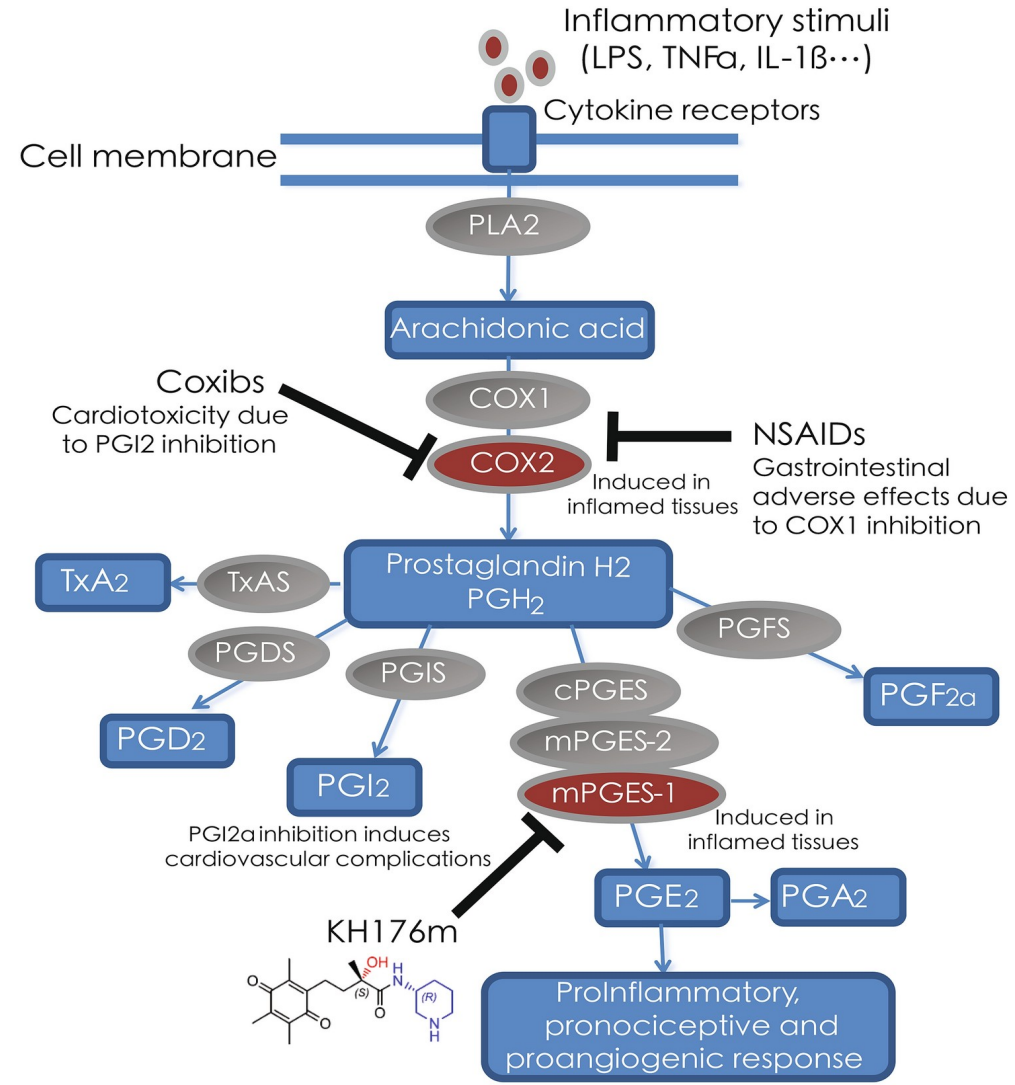
All links

Ontology (2)
 KEGG BRITE (2)
 Pathway (3)
 KEGG PATHWAY (2)
 KEGG MODULE (1)
 Genome (1)
 KEGG GENOME (1)
 Gene (17)
 KEGG ORTHOLOGY (1)
 RefGene (13)
 NCBI-PROTEINID (1)
 NCBI-Gene (1)
 OC (1)
 Protein sequence (3)
 UniProt (1)
 SWISS-PROT (1)
 RefSeq(pep) (1)
 Protein domain (3)
 Pfam (3)
 All databases (29)

[Download RDF](#)

Entry	K11987	K0
Symbol	PTGS2, COX2	
Name	prostaglandin-endoperoxide synthase 2 [EC:1.14.99.1]	
Pathway	<p>map00590 Arachidonic acid metabolism ←</p> <p>map01100 Metabolic pathways ←</p> <p>map04064 NF-kappa B signaling pathway ←</p> <p>map04370 VEGF signaling pathway ←</p> <p>map04625 C-type lectin receptor signaling pathway ←</p> <p>map04657 IL-17 signaling pathway ←</p> <p>map04668 TNF signaling pathway ←</p> <p>map04723 Retrograde endocannabinoid signaling</p> <p>map04726 Serotonergic synapse</p> <p>map04913 Ovarian steroidogenesis</p> <p>map04921 Oxytocin signaling pathway</p> <p>map04923 Regulation of lipolysis in adipocytes</p> <p>map05010 Alzheimer disease</p> <p>map05022 Pathways of neurodegeneration - multiple diseases</p> <p>map05140 Leishmaniasis</p> <p>map05163 Human cytomegalovirus infection</p> <p>map05165 Human papillomavirus infection</p> <p>map05167 Kaposi sarcoma-associated herpesvirus infection</p> <p>map05200 Pathways in cancer</p> <p>map05204 Chemical carcinogenesis - DNA adducts</p> <p>map05206 MicroRNAs in cancer</p> <p>map05222 Small cell lung cancer</p>	
Disease	<p>H00017 Esophageal cancer</p> <p>H00025 Penile cancer</p> <p>H00046 Cholangiocarcinoma</p>	
Brite	<p>KEGG Orthology (K0) [BR:ko00001]</p> <p>09100 Metabolism</p> <p>09103 Lipid metabolism</p> <p>00590 Arachidonic acid metabolism</p> <p>K11987 PTGS2, COX2; prostaglandin-endoperoxide synthase 2</p> <p>09130 Environmental Information Processing</p> <p>09132 Signal transduction</p> <p>04370 VEGF signaling pathway</p> <p>K11987 PTGS2, COX2; prostaglandin-endoperoxide synthase 2</p> <p>04064 NF-kappa B signaling pathway</p> <p>K11987 PTGS2, COX2; prostaglandin-endoperoxide synthase 2</p> <p>04668 TNF signaling pathway</p> <p>K11987 PTGS2, COX2; prostaglandin-endoperoxide synthase 2</p> <p>09150 Organismal Systems</p> <p>09151 Immune system</p> <p>04625 C-type lectin receptor signaling pathway</p>	

Exemplo de um pathway contendo COX2



Jiang, X., Renkema, H., Pennings, B. et al. Mechanism of action and potential applications of selective inhibition of microsomal prostaglandin E synthase-1-mediated PGE₂ biosynthesis by sonlicromanolol's metabolite KH176m. Sci Rep 11, 880 (2021). <https://doi.org/10.1038/s41598-020-79466-w>

Vamos achar genes? e regiões intergênicas...

UCSC genome browser



← → ↻ 🔒 genome.ucsc.edu

UNIVERSITY OF CALIFORNIA SANTA CRUZ Genomics Institute UCSC Genome Browser

Genomes Genome Browser Tools Mirrors Downloads My Data Projects Help About Us

SAT2 SHBG ATP1B2 TP53 WRAP53

Tools

-  **Genome Browser** - Interactively visualize genomic data
- BLAT** - Rapidly align sequences to the genome
- In-Silico PCR** - Rapidly align PCR primer pairs to the genome
-  **Table Browser** - Download and filter data from the Genome Browser
- LiftOver** - Convert genome coordinates between assemblies
- REST API** - Returns data requested in JSON format
- Variant Annotation Integrator** - Annotate genomic variants
- More tools...**

News

May 04, 2023 - **EVA SNP release 4 for 36 assemblies**

Apr. 24, 2023 - **New DGV Gold Standard track for hg38**

Apr. 14, 2023 - **Problematic Regions for hg38 and a new public hub**

Apr. 10, 2023 - **New Cross Tissue Nuclei track for hg38**

Mar. 29, 2023 - **Updated DGV tracks for hg19/hg38**

Mar. 08, 2023 - **TOGA Gene Prediction tracks for 41 assemblies**

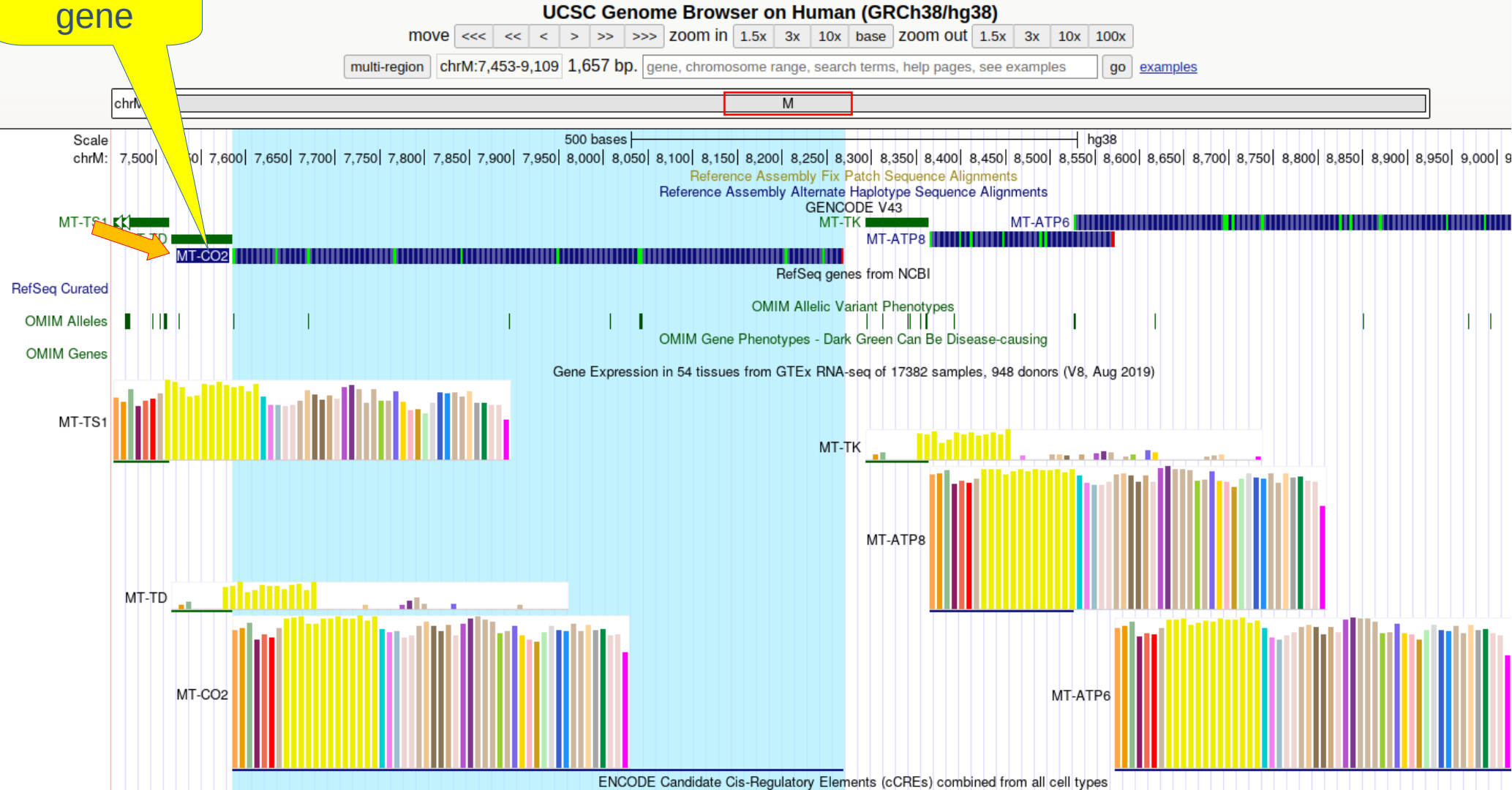
[More news...](#) [Subscribe](#)



<https://genome.ucsc.edu/>

COX2 - MT-CO2

achei o
gene



COX2 - MT-CO2

navegar
frente-trás

UCSC Genome Browser on Human (GRCh38/hg38)

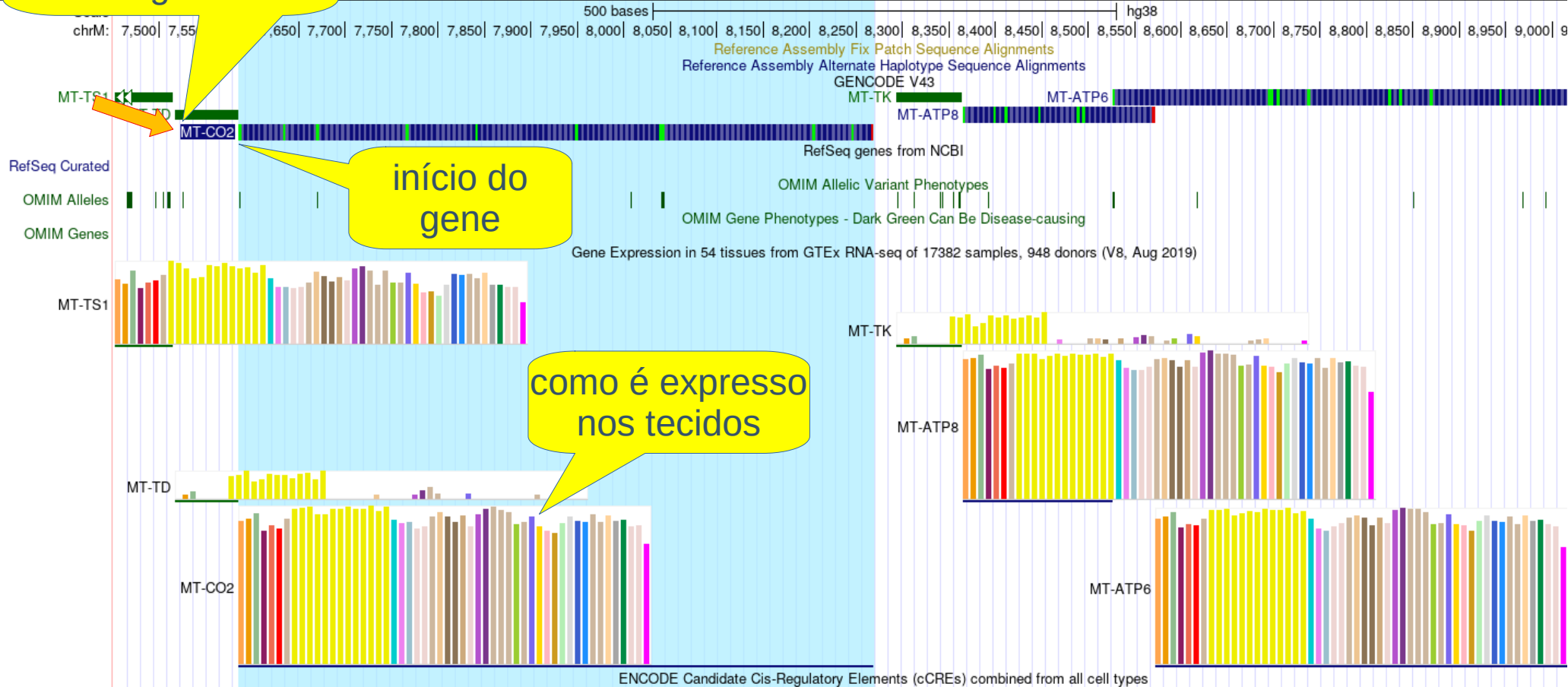
ampliar
reduzir

label do gene
dê um right-mouse

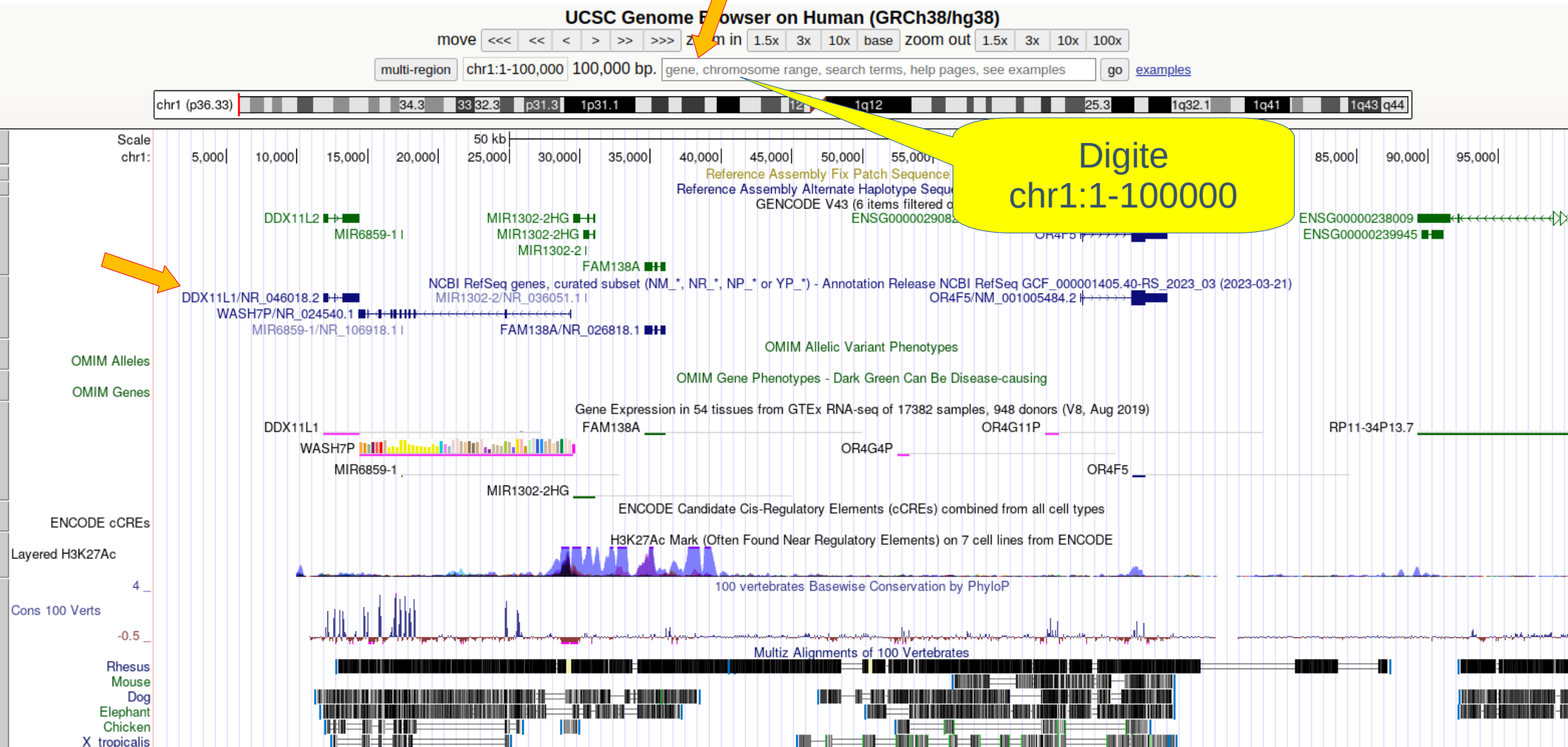
move <<< << < > >> >>> zoom in 1.5x 3x 10x base zoom out 1.5x 3x 10x 100x
multi-region chrM:7,453-9,109 1,657 bp. gene, chromosome range, search terms, help pages, see examples go examples

início do
gene

como é expresso
nos tecidos



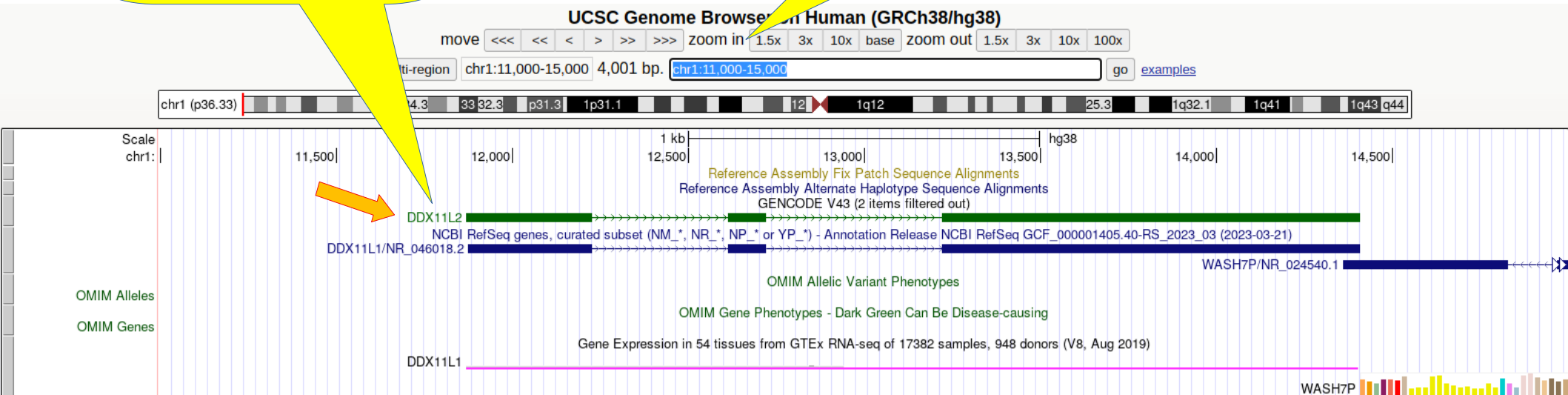
Cromossomo 1 de 1:100000



Zoom: chr1:11,000-15,000

Right-mouse
get DNA

Procure por
DDX11



Vamos fazer um BLAST de nucleotídeos

Basic Local Alignment Search Tool



<https://blast.ncbi.nlm.nih.gov/Blast.cgi>

BLAST - nucleotides

Basic Local Alignment Search Tool

BLAST finds regions of similarity between biological sequences. The program compares nucleotide or protein sequences to sequence databases and calculates the statistical significance.

[Learn more](#)

NEWS

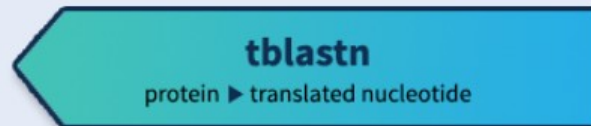
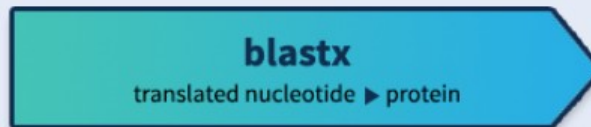
BLAST+ 2.14.0 is here!

BLASTP, BLASTX, and TBLASTN are faster than before.

Fri, 28 Apr 2023

[More BLAST news...](#)

Web BLAST



<https://blast.ncbi.nlm.nih.gov/Blast.cgi>

BLAST – copie a sequência e dê um blast

Standard Nucleotide BLAST

BLASTN programs search nucleotide databases using a nucleotide query. [more...](#)

Enter Query Sequence

Enter accession number(s), gi(s), or FASTA sequence(s) [?](#) [Clear](#) Query subrange [?](#)

From

To

Or, upload file No file chosen [?](#)

Job Title

Enter a descriptive title for your BLAST search [?](#)

☐ Align two or more sequences [?](#)

Choose Search Set

Database ☒ Standard databases (nr etc.): ☐ rRNA/ITS databases ☐ Genomic + transcript databases ☐ Betacoronavirus

[New](#) ☐ Experimental databases [Try experimental taxonomic nt databases](#) [Download](#)

For more info see [What are taxonomic nt databases?](#)

Nucleotide collection (nr/nt) [?](#)

Organism Optional ☐ exclude [Add organism](#)

Enter organism common name, binomial, or tax id. Only 20 top taxa will be shown [?](#)

Exclude Optional ☐ Models (XM/XP) ☐ Uncultured/environmental sample sequences

Limit to Optional ☐ Sequences from type material

Entrez Query Optional [YouTube](#) [Create custom database](#)

Enter an Entrez query to limit search [?](#)

Program Selection

Optimize for ☒ Highly similar sequences (megablast) ☐ More dissimilar sequences (discontiguous megablast) ☐ Somewhat similar sequences (blastn)

Choose a BLAST algorithm [?](#)

[BLAST](#) Search database Nucleotide collection (nr/nt) using Megablast (Optimize for highly similar sequences)

☐ Show results in a new window

copie aproximadamente os primeiros 60 nucleotídeos

BLAST achou DDX11 mais similares com a sequência enviada

Job Title	Nucleotide Sequence		
RID	6ANVHZR6013	Search expires on 05-19 06:38 am	Download All ▼
Program	BLASTN ?	Citation ▼	
Database	nt	See details ▼	
Query ID	lcl Query_59665		
Description	None		
Molecule type	dna		
Query Length	1450		
Other reports	Distance tree of results	MSA viewer	?

Filter Results

Organism only top 20 will appear ☐ exclude

[+ Add organism](#)

Percent Identity
 to

E value
 to

Query Coverage
 to

[Filter](#) [Reset](#)

- Descriptions
- Graphic Summary
- Alignments
- Taxonomy

Sequences producing significant alignments

Download ▼

Select columns ▼

Show ▼ [?](#)

☒ select all 100 sequences selected

[GenBank](#) [Graphics](#) [Distance tree of results](#) [MSA Viewer](#)

	Description ▼	Scientific Name ▼	Max Score ▼	Total Score ▼	Query Cover ▼	E value ▼	Per. Ident ▼	Acc. Len ▼	Accession
<input checked="" type="checkbox"/>	Homo sapiens DEAD/H-box helicase 11 (DDX11), RefSeqGene on chromosome 12	Homo sapiens	2678	2678	100%	0.0	100.00%	37947	NG_023352.1
<input checked="" type="checkbox"/>	Homo sapiens 12 BAC RP11-551L14 (Roswell Park Cancer Institute Human BAC Library) complete seque...	Homo sapiens	2678	2678	100%	0.0	100.00%	168986	AC008013.8
<input checked="" type="checkbox"/>	Homo sapiens chromosome 12 clone CH17-321K24, complete sequence	Homo sapiens	2632	2632	100%	0.0	99.45%	210014	AC277891.1
<input checked="" type="checkbox"/>	Homo sapiens FOSMID clone ABC13-48569500C5 from chromosome 12, complete sequence	Homo sapiens	2503	2503	100%	0.0	97.80%	37211	AC215997.4
<input checked="" type="checkbox"/>	Homo sapiens BAC clone CH17-138H12 from chromosome 12, complete sequence	Homo sapiens	2497	2497	100%	0.0	97.73%	211866	AC254631.1
<input checked="" type="checkbox"/>	Homo sapiens 12p13 PAC RPC11-726G1 (Roswell Park Cancer Institute Human PAC Library) complete s...	Homo sapiens	2486	2486	100%	0.0	97.59%	180685	AC006432.15

BLAST achou DDX11 mais similares

Job Title	Nucleotide Sequence
RID	6ANVHZR6013 Search expires on 05-19 06:38 am Download All ▼
Program	BLASTN ? Citation ▼
Database	nt See details ▼
Query ID	lcl Query_59665
Description	None
Molecule type	
Query Length	
Other reports	Viewer ?

Estatística:
Score: nucleotídeos alinhados
E-value: ~ p-valor (se pequeno é significativo)
Porcentagem de identidades
Comprimento
Accession Number – link para o registro

Sequências encontradas

Organismo

accession ID

Descriptions		Graphic Summary		Alignments	Taxonomy			
Sequences producing significant alignments								
<input checked="" type="checkbox"/> select all 100 sequences selected		Download		Select columns	Show 100			
		GenBank		Graphics	Distance tree of results			
		MSA Viewer						
Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
<input checked="" type="checkbox"/> Homo sapiens DEAD/H-box helicase 11 (DDX11), RefSeqGene on chromosome 12	Homo sapiens	2678	2678	100%	0.0	100.00%	37947	NG_023352.1
<input checked="" type="checkbox"/> Homo sapiens 12 BAC RP11-551L14 (Roswell Park Cancer Institute Human BAC Library) complete sequence	Homo sapiens	2678	2678	100%	0.0	100.00%	168986	U08013.8
<input checked="" type="checkbox"/> Homo sapiens chromosome 12 clone CH17-321K24, complete sequence	Homo sapiens	2632	2632	100%				
<input checked="" type="checkbox"/> Homo sapiens FOSMID clone ABC13-48569500C5 from chromosome 12, complete sequence	Homo sapiens	2503	2503	100%				
<input checked="" type="checkbox"/> Homo sapiens BAC clone CH17-138H12 from chromosome 12, complete sequence	Homo sapiens	2497	2497	100%	0.0	97.73%	211866	AC254631.1
<input checked="" type="checkbox"/> Homo sapiens 12p13 PAC RPC11-726G1 (Roswell Park Cancer Institute Human PAC Library) complete sequence	Homo sapiens	2486	2486	100%	0.0	97.59%	180685	AC006432.15

clique aqui

DDX11 registro do NG_023352 no Genbank do NCBI

GenBank ▾

Send to: ▾

Homo sapiens DEAD/H-box helicase 11 (DDX11), RefSeqGene on chromosome 12

NCBI Reference Sequence: NG_023352.1

[FASTA](#) [Graphics](#)

[Go to:](#) ☐

LOCUS	NG_023352	37947 bp	DNA	linear	PRI 25-MAR-2023
DEFINITION	Homo sapiens DEAD/H-box helicase 11 (DDX11), RefSeqGene on chromosome 12.				
ACCESSION	NG_023352				
VERSION	NG_023352.1				
KEYWORDS	RefSeq; RefSeqGene.				
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo.				
REFERENCE	1 (bases 1 to 37947)				
AUTHORS	Alkhunaizi,E., Brosh,R.M. Jr., Alkuraya,F.S. and Chitayat,D.				
TITLE	Warsaw Syndrome				
JOURNAL	(in) Adam MP, Mirzaa GM, Pagon RA, Wallace SE, Bean LJH, Gripp KW and Amemiya A (Eds.); GENEREVIEWS(R); (1993)				
PUBMED	31169992				
COMMENT	REVIEWED REFSEQ : This record has been curated by NCBI staff. The reference sequence was derived from AC008013.8 . This sequence is a reference standard in the RefSeqGene project.				

Summary: DEAD box proteins, characterized by the conserved motif

Aonde estão os Éxons?

Onde começa este gene? start-codon ATG

```
>hg38 dna range=chr12:31073860-31104799 5'pad=0 3'pad=0 strand=+ repeatMasking=none
GTTGTTCCGGCTGCCTTTCACTGAGGGGACCCGCCAGTTTCTAACTCAGT
GGCGTTTGCCCTGATTCCCGGGGCTGGCTTTCAGCGTAGCAATTCTGCC
GGCGAAGAAGGTGAGCGCAGTGCTGTGTGGCAGCAGAGCTCCTTAGGACG
AGGAGCAGCGGGACGAGGAAGGGCAGACTGGTGAATCGCAAACCTGGGCG
TCTGTTCCGGCGCCGGACCCCTATTTGCAAAGGTGGGTGGCCCGTCCGGA
GCGGGAAAAACATTCCGGAAAGTGGAGGGCCGGGCCAGCGTGATTGACAAGC
GGGAACCCCTGTGTGGGACGGGTAGGCCTAGGAAGGTTGTGCCTGCGGT
GGAACCTGGGCGGTGCGCAGAAGTGGGCATTAAACAGCAGCCGCGTGTCTGG
GTCTTAGATTTGGCCAGCTGTGTTGAGCTTTTCATGGAATTATCTTACGT
AGATAAGACACTGCAACAGTGAGTGAGCGCTTGTAAACCACTGTCTCTT
ACGGAAACTGAGCCCCAGCGATGCTAACTTTAGCAAGGATACAGCTGGGA
TCCTAAACTTGGCAATCAGAGCCAGAGCCGATAAAAGTTAGCTGCATGAG
TCTAGCTTCCCCAGGGCGGGAATCGAGGCGGAGCAGGGTACAGTACGGA
GGCCAGGAATGAGTGCACTTGACCAGATTGTTGACGGAAGTGTCATAAAA
ATGGACTTAAATGCTGATAAGCAGCTGTTTGGGTTCCACACAGGGTGCGA
GCTCAGGATGCACGTTGGAGGGGACACAGGGCCAGAGCAAGGTGGGAATG
CGGGTATTATGGGCCAGGCCATCCTCCAGCTGGTGGAGCACCGCAGTACT
GCAGTGTGTGGCCCTGGCTTAACAGCAGTGCGGAAAAAGCTTTTTCTTGG
GGCTGTGGTGCTTCCAGGTGTGTGAAGTAAAAACATTTTAGGGCCATAC
GGTAGATAGTACCTGCCACATAGTTGTCTTGGTCAGCCCCGGCTGCCATG
ACAAAATACCATAAACTGGGTGGCTTAGACACAGAAATGTTTCTCAC
AGTCCAGAGGCTTGGGATGCCAAGATCAATGTTGGGAGGACACAACCTG
AGCCCAACAACAATAGTCTGTACTCAGTAATGAGTGATAATTGGGGACTGA
AGAAAATGAACACGTAAAGAACTAATATGTTCTGAAGTGCTTTCACAAC
TCTAACCTCGTTTTATGAGCGTGAGCTTTGCTGCTGGTGTGTGCTTG
GCACTGGGAGGTGATGTTGTCTCCACACAGCCAACCTGAAGAGGGCTG
AACAGTCACTGCAAATGTTTTAATAGGGCTTAGTGAATCCGTTATACT
CAGATTTATCTAAACCTCTATGATTTAGCCTGTGCTGCTTCTGGAATAAT
GAGATCCATAATTACCACTGATGGGAAGTGAAATAATACTTACGTTTCT
TCTGTATTGTTAGTGTTGTCTTCTAGAGTTTGGTAACCAAGTCTGTCT
TTTTATCCCACTTATCCTGGTGGGAGGAAAAAGTGAGGAGATAGAAAGTT
TCAGGTGGCTTGGGGGTCTGGCAGATGTGGTTCAAATCCTGAGTTCAAGC
ACTTGCTGAGTGACCTTGGGCAAGTCATATAAGTTTACTGAGTCTCAGTT
TCTTCTCTATAAAATAGAGCTTATAAAATACCACACAGGGTTTTGTG
GGGTTATCTGAAAAACGGTTTGAAACCATTAAGAACTGGCCATTTAACTA
```

Aonde estão os Éxons?

Onde começa este gene? start-codon ATG

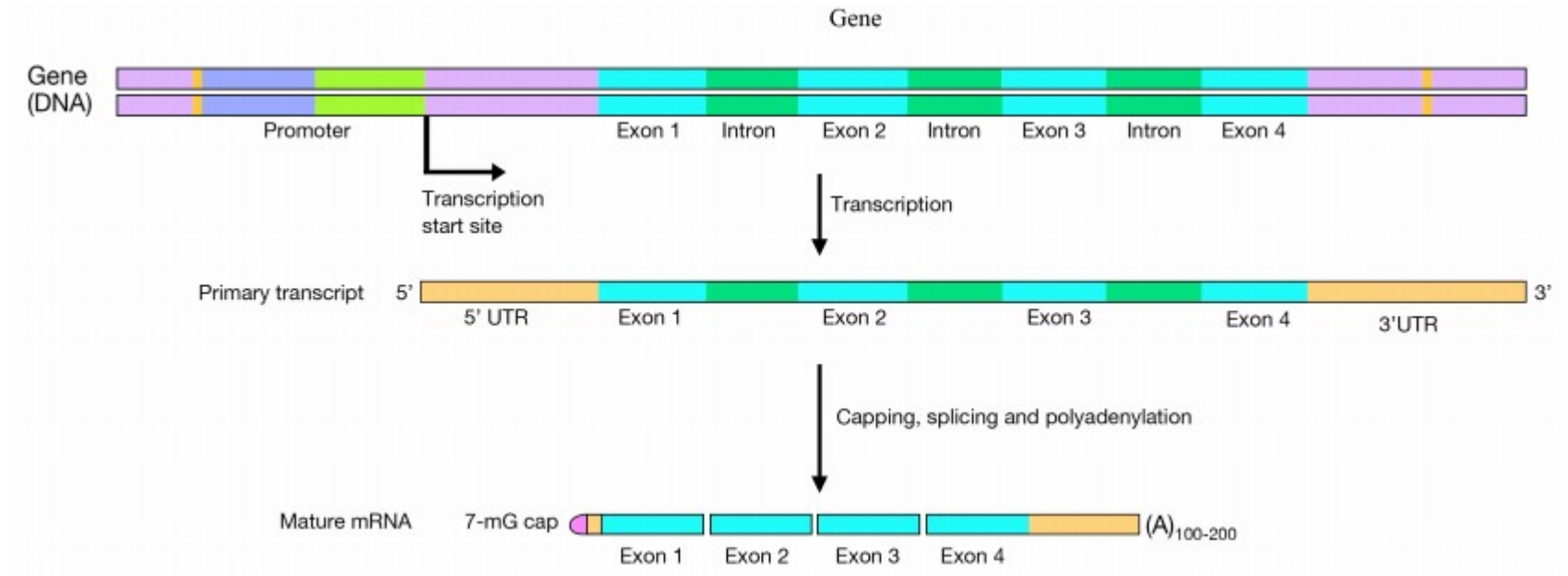
>hg38 dna range=chr12:31073860-31104799 5'pad=0 3'pad=0 strand=+ repeatMasking=none

GTGTTCCGGCTGCCTTTCACTGAGGGGACCGCCAGTTTCTAAGTCACT
GGCGTTTGCCCTGATTCCCGGGGCTGGCTTTCAGCGTAGCAATTCTGCC
GGCGAAGAAGGTGAGCGCAGTGTGTGTGGCAGCAGAGCTCCTTAGGACG
AGGAGCAGCGGGACGAGGAAGGGCAGACTGGTGAATCGCAAACCTGGGCG
TCTGTTCCGGCGCCGGACCCCTATTTGCAAAGGTGGGTGGCCCGTCCGGA
GCGGGAAAAACATTCCGGAAGTGGAGGGCCGGGCGCAGCGTATTGACAAGC
GGGAACCCCTGTGTGGGACGGGTAGGCCTAGGAAGGTTGTGCTGCGGT
GGAACCTGGGCGGTGCGCAGAAGTGGGCATTAAACAGCAGCGCGTGTCTGG
GTCTTAGATTGGCCAGCTGTGTTGAGCTTTTCATGGAATTATCTTACGT
AGATAAGACACTGCAACAGTGAGTGAGCGCTTGAACCCACCTGTCTCTT
ACGGAACTAGACCCAGCGATGCTAACTTTAGCAAGGATACAGCTGGGA
TCCTAAACTTGGCAATCAGAGCCAGAGCCGATAAAGTTAGCTGCATGAG
TCTAGCTTCCCCAGGGCGGGAATCAGGGCGGAGCAGGGTACAGTACGGA
GGCCAGGAATGAGTGCACCTTGACCAGATTGTTGACGGAAGTGTCAAAAA
ATGGACTTAAATGCTGATAAGCAGCTGTTTGGGTTCCACACAGGGTGCAG
GCTCAGGATGACGTTGGAGGGGACACAGGGCCAGAGCAAGGTGGGAATG
CGGGTATTATGCGGCGAGCCATCCTCCAGCTGGTGGAGCACCGCAGTACT
GCAGTGTGTGGCCCTGGCTTAAACAGCAGTGCGGAAAAAGCTTTTCTTGG
GGCTGTGGTGCTTCCAGGTGTGTGAAGTAAAAACATTTTAGGGCCATAC
GGTAGATAGTACCTGCCACATAGTTGTCTTGGTCAGCCCCGGCTGCCATG
ACAAAATACCATAAAGTGGGTGGCTTAGACACAGAAATGTTTCTCAC
AGTCCCAGAGGCTTGGGATCCAGATCAATGTTGGGAGGACACAAGT
AGCCCAACAACATAGTCTGTAAGTATGAGTGAATAATGGGGACTGA
AGAAAATGAACACGTTAAGAACTAATATGTTCTCTGAAGTGCTTTCACAAC
TCTAACCTCGTTTATGAGCGTGAGCTTGTGTCTGGTGTGTGCTTGG
GCACTGGGAGGTGATGTTGTCTCCACACAGCCAACCTGAAGAGGGCTG
AACAGATCACTGCAGAAATGTTTTTAATAGGGCTTAGTGAATCCGTTAAT
CAGATTTATCTAAACCTCTATGATTTAGCCTGTGCTGCTTCTGGAATAAT
GAGATCCATAATTACCACTGATGGGAAGTGAAATAAATACTTACGTTTCT
TCTGTATTGTTTAGTGTGTCATTCTAGAGTTTGGTAACCAAGTCTGTCT
TTTTATCCCACTTATCCTGGTGGGAGGAAAAAGTGAGGAGATAGAAAGTT
TCAGGTGGCTTGGGGGTCTGGCAGATGTTGTTCAAATCCTGAGTTCAAGC
ACTTGCTGAGTGACCTTGGGCAAGTCATATAAGTTTACTGAGTCTCAGTT
TCTTCTCTATAAAATAGAGCTTATAAAATACCACACAGGGTTTTTGTG
GGGTTATCTGAAAAACGGTTTGAACCATTAAGAACTGGCCATTTAACTA

Está lá para baixo.
Mas como foi achado?

ACACGGTGAACCCCGTCTCTACCAAAAAAAAAAAAAACGTTGGT
GGCGGGCACCTGTAGTCCCAGCTACTCGGGAGGCTTGAAGATGG
TGTGAACCCGGGAGGCGGAGCGTGCAGTGAGCCGATCACTGCA
CTCCAGCCTGGGCGACAGAGTGAGACTCTGTCTCGAATAAAAAA
AAGATAGGGACGTGTATGTTAACTTGCCTATCCAAGAGCTGTGC
TTATGGTCTCTGCTGTGCGTCATGATTTTCCAGGACAACGGGA
TAAAGTGAAGTAGCTTCGGCTTGTGAATGTGCATTGACGTGGGAG
AAGAAAGCTGCAAAAGTCATTATAAGCAACACCCTTGTAGGGGTGC
TGGTCTGTGAGAAGAGAGCTCTAAGCCTTTTGTAGAGTTAGAAAGGT
ATCGGAGCCACGGTGAATGCAGGGGAGATTGGGTTTGGGCTTTCCTG
GTCTGCATTCTGCTACAGCCGTTAAATGCCGCTAGATAGTGCGTGATT
CTGGTATGGCCTCACGTGGACCTGCTGCGAAGGATGGAGAACATGGTC
TCTGCTTCCAGAAAAAAGGAGAAATTTGGTAATAAGTGTGGAGACTGCT
CTTAAATAATGCTCCAGATTTCAAGCCACTTCTTCTGGACCATGAGAGA
GCTCCCTAATGTTGTATTTATTTTTCTAGGTCCATGGCTAATGAAACAC
AGAAGGTTGGTGCCATCCATTTTCTTTTCCCTTACACCCTATTCCATC

Aonde estão os Éxons? Onde começa este gene?



Localizando os éxons no texto do Genbank

mRNA

```
join(5016..5247,9546..9693,14969..15217,15739..15825,  
16125..16282,19094..19139,20200..20307,20559..20646,  
21042..21250,22875..23027,24002..24048,24401..24480,  
25746..25790,25911..25978,27497..27535,27793..27901,  
28015..28146,29041..29153,31791..31863,32183..32286,  
32989..33138,33399..33467,33583..33683,34092..34176,  
34473..34551,34733..34887,34963..35955)  
/gene="DDX11"  
/gene_synonym="CHL1; CHLR1; KRG2; WABS"  
/product="DEAD/H-box helicase 11, transcript variant 1"  
/transcript_id="NM_030653.4"  
/db_xref="GeneID:1663"  
/db_xref="HGNC:HGNC:2736"  
/db_xref="MIM:601150"
```

exon

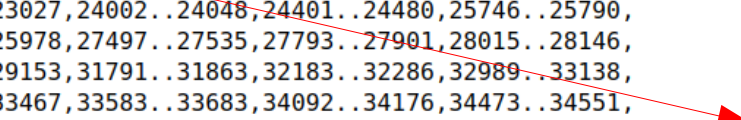
```
5016..5247  
/gene="DDX11"  
/gene_synonym="CHL1; CHLR1; KRG2; WABS"  
/inference="alignment:Splign:2.1.0"  
/number=1
```

exon

```
9546..9693  
/gene="DDX11"  
/gene_synonym="CHL1; CHLR1; KRG2; WABS"  
/inference="alignment:Splign:2.1.0"  
/number=2
```

CDS

```
join(9550..9693,14969..15217,15739..15825,16125..16282,  
19094..19139,20200..20307,20559..20646,21042..21250,  
22875..23027,24002..24048,24401..24480,25746..25790,  
25911..25978,27497..27535,27793..27901,28015..28146,  
29041..29153,31791..31863,32183..32286,32989..33138,  
33399..33467,33583..33683,34092..34176,34473..34551,  
34733..34887,34963..34992)  
/gene="DDX11"  
/gene_synonym="CHL1; CHLR1; KRG2; WABS"  
/EC_number="3.6.4.12"
```



9541	cctaggtcca	atgggctaata	aacacagaag	gttgggtgcc	tccattttcc	ttttcccttc
9601	acaccctatt	ccatccagga	agacttcatg	gcagagctgt	accggggttt	ggaggctggc
9661	aagattggga	tatttgagag	tccaactggc	actgtgagta	tgaacagtga	gagatactga
9721	aaaggacaac	ttaacagcag	ccgtactagc	ttttcctgtt	tgccatcca	gagattttca

Parte do registro do Genbank refere-se a anotações

```
/EC_number="3.6.4.12"  
/note="isoform 1 is encoded by transcript variant 1;  
DEAD/H (Asp-Glu-Ala-Asp/His) box polypeptide 11 (CHL1-like  
helicase homolog, S. cerevisiae); probable ATP-dependent  
RNA helicase DDX11; CHL1-related helicase gene-1; DEAD/H  
(Asp-Glu-Ala-Asp/His) box helicase 11; KRG-2; hCHLR1;  
DEAD/H box protein 11; CHL1-related protein 1;  
keratinocyte growth factor-regulated gene 2 protein;  
probable ATP-dependent DNA helicase DDX11"  
/codon_start=1  
/product="ATP-dependent DNA helicase DDX11 isoform 1"  
/protein_id="NP_085911.2"  
/db_xref="CCDS:CCDS41767.1"  
/db_xref="GeneID:1663"  
/db_xref="HGNC:HGNC:2736"  
/db_xref="MIM:601150"  
/translation="MANETQKVGAIHFPPFTPYSIQEDFMAELYRVLEAGKIGIFES  
PTGTGKSLSLICGALSWLRDFEQKKREEEARLLETGTGPLHDEKDESLCLSSSCEGAA  
GTPRPAGEPAWVTQFVQKKEERDLVDRLKAEQARRKQREERLQQLQHRVQLKYAAKRL  
RQEEEEERENLLRLSRELETGPEAERLEQLESGEELVLAYESDEEKKVASRVEDE  
DDLEEEHITKIYYCSRTHSQLAQFVHEVKKSPFGKDVRLVSLGSRQNLVNEDEVKSLG  
SVQLINDRCVDMQRSRHEKKKGAEKKPKRRRQEKQAACPFYNHEQMGLLRDEALAEV  
KDMEQLLALGKEARACPYYSRLAIPAAQLVVLPHYQMLLHAATRQAAGIRLQDQVVII  
DEAHNLIDTITGMHSEVSGSQLCAHSQLLQYVERYGKRLKAKNLMYKQILYLLEK  
FVAVLGGNIKQNPNTQSLSQGTGTELKTINDFLFQSQIDNINLFKVQRYCEKSMISRL  
FGFTERYGAVFSSREQPKLAGFQQFLQSLQPRTEALAAPADESQASTLRPASPLMHI  
QGFLAALTANQDGRVILSRQGSLSQSTLKFLLNPAVHFAQVVKECRAVVIAGGTMQ  
PVSDFRQQLLACAGVEAERVVEFSCGHVIPPDNILPLVICSGISNQPLEFTFQKREL  
QMMDEVGRILCNLCGVVPGGVVCFPPSYEYLRQVHAHWEKGGLLGRLAARKKIFQEPK  
SAHQVEQVLLAYSRCIQACGQERGQVTGALLLSVVGGMSEGINFSDNLGRCVVMVGM  
PPFNIRSAELQEKMAYLDQTLPRAPGQAPP GKALVENLCMKAVNQSIGRAIRHQKDF  
SVVLLDQRYARPPVLAKLPWIRARVEVKATFGPAIAAVQKHFHREKSASS"
```

outras bases
do NCBI

Protein
CDS
HGNC (Hugo)
MIM



clique aqui para
ir para as anotações
da proteína

clique aqui também
para ver o coding
segment CDS

Salvar a proteína em fasta no site

GenPept ▾

Send to: ▾

ATP-dependent DNA helicase DDX11 isoform 1 [Homo sapiens]

NCBI Reference Sequence: NP_085911.2

[Identical Proteins](#) [FASTA](#) [Graphics](#)

Go to: ☐

LOCUS NP_085911 906 aa linear PRI 12-MAR-2023
DEFINITION ATP-dependent DNA helicase DDX11 isoform 1 [Homo sapiens].
ACCESSION NP_085911
VERSION NP_085911.2
DBSOURCE REFSEQ: accession [NM_030653.4](#)
KEYWORDS RefSeq; MANE Select.
SOURCE Homo sapiens (human)
ORGANISM [Homo sapiens](#)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini;
Catarrhini; Hominidae; Homo.
REFERENCE 1 (residues 1 to 906)
AUTHORS Saeed M, Ibanez-Costa A, Patino-Trives AM, Munoz-Barrera L,
Collantes Estevez E, Aguirre MA and Lopez-Pedrerera C.
TITLE Expression of DDX11 and DNMT1L at the 12p11 Locus Modulates Systemic
Lupus Erythematosus Susceptibility
JOURNAL Int J Mol Sci 22 (14), 7624 (2021)
PUBMED [34299244](#)
REMARK GeneRIF: Expression of DDX11 and DNMT1L at the 12p11 Locus Modulates
Systemic Lupus Erythematosus Susceptibility.

Export:
File
Fasta

<https://www.ncbi.nlm.nih.gov/protein/100913202>

Copie as sequências CDS no conversor do EMBOSS

<https://www.ebi.ac.uk/>

<https://www.embl.org/>

EMBL

The European
Molecular
Biology
Laboratory

EMBOSS Transeq

[Input form](#)[Web services](#)[Help & Documentation](#)[Bioinformatics Tools FAQ](#)[Feedback](#)

Tools > Sequence Translation > EMBOSS Transeq

Service Announcement

The new Job Dispatcher Services beta website is now available at <https://wwwdev.ebi.ac.uk/Tools/jdispatcher>. We'd love to hear your feedback about the new webpages!

EMBOSS Transeq

EMBOSS Transeq translates nucleic acid sequences to their corresponding peptide sequences. It can translate to the three forward and three reverse frames, and output multiple frame translations at once.

STEP 1 - Enter your input sequence

Enter or paste a DNA/RNA sequence in any supported format:

```
ATGGCTAATGAAACACAGAAGGTTGGTGCCATCCATTTTCCTTTCCCTTCACACCCTATTCCATCCAGG
AAGACTTCATGGCAGAGCTGTACCGGGTTTTGGAGGCTGGCAAGATTGGGATATTTGAGAGTCCAACCTGG
CACTGGGAAGTCCTTAAGTCTTATTTGTGGGGCCCTCTCTTGGCTCCGTGACTTTGAACAGAGAAGAAGCGT
GAAGAAGAGGGCAGCACTCCTTGAAACTGGAACCTGGCCCCCTACATGATGAGAAAGATGAATCCCTGTGTC
TGTCTTCTTCTCGCAAGGGGCTGCAGGCACCCGAGGCCTGCTGGAGAACCGGCCTGGGTACTCAGTT
TGTGCAGAGAAGAAAGAGAGAGGGACCTGGTGGACCGACTAAAGGCGGAGCAGGCCAGGAGGAAGCAGCGA
GAAGAACGCCTGCAGCAGCTGCAGCACAGGGTGACGCTCAAGTATGCAGCCAAGCGCCTGAGGCAGGAAG
AAGACAAACAGACAGATCTCTCGCCCTCAGCAGCCAGATCTAGACAGAGCCGCGCAGGCTGAGCCCT
```

Or, upload a file: No file chosen

[Use a example sequence](#) | [Clear sequence](#) | [See more example input](#)

STEP 2 - Select Parameters

FRAME

1

CODON TABLE

Standard Code

Copie o CDS
aqui

https://www.ebi.ac.uk/Tools/st/emboss_transeq/

Resultado do conversor do EMBOSS

<https://www.ebi.ac.uk/Tools/emboss/>

EMBL

The European
Molecular
Biology
Open Source
Software

Results for job emboss_transeq-l20230518-184221-0857-82638710-p2m

Tool Output

Submission Details

Download

Show Colors

```
>EMBOSS_001_1
MANETQKVGAIHFPPFTPYSIQEDFMAELYRVLEAGKIGIFESPTGTGKSLSLICGALS
WLRDFEQKKREEEARLLETGTGPLHDEKDESLCLSSSCGAAGTPRPAGEPAWVTQFVQK
KEERDLVDRLKAEQARRKQREERLQQLQHRVQLKYAAKRLRQEEEEERENLLRLSREMLET
GPEAERLEQLESGEEELVLAYESDEEKKVASRVDEDEDDLEEEHITKIYYCSRTHSQA
QFVHEVKKSPFGKDVRLVSLGSRQNL CVNEDVKSLGSLVQLINDRCVDMQRSRHEKKKGAE
EEKPKRRRQEKQAACPFYNHEQMGLLRDEALAEVKDMEQLLAGKEARACPYYSRLAIP
AAQLVVLPHYQMLLHAATRQAAGIRLQDQVVIIDEAHNLIDTITGMHSVEVSGSQLCQHS
QLLQYVERYGKRLKAKNLMYLKQILYLLEKFVAVLGGNIKQNPNTQSLSQTGTELKTIND
FLFQSQIDNINLFKVQRYCEKSMISRLFGFTERYGAVFSSREQPKLAGFQQFLQSLQPR
TTEALAAPADESQASTLRPASPLMHIQGFALAALTANQDGRVILSRQGSLSQSTLKFLLL
NPAVHFAQVVKECRAVVIAGGTMQPVSDFRQQLACAGVEAERVVEFSCGHVIPPDNILP
LVICSGISNQPLEFTFQKRELQMMDEVGRILCNLCGVVPGGVVCFPSYEYLRQVHAHW
EKGGLLGRLAARKKIFQEPKSAHQVEQVLLAYSRCIQACGQERGQVTGALLLSVVGGKMS
EGINFSDNLGRCVVMVGMPFPNIRSAELQEKMAYLDQTLPRAPGQAPPKALVENLCMKA
VNQSIGRAIRHQKDFASVLLDQRYARPPVLAKLPWIRARVEVKATFGPAIAAVQKFHR
EKSASS*
```


exercício

Exercício fazer um BLAST de amino ácidos

Pegue os 30 primeiros aa de DDX11 e faça um BLAST



<https://blast.ncbi.nlm.nih.gov/Blast.cgi>

Mas, na verdade eu gosto do Seaview



PRABI-Doua
Pôle Rhône-Alpes de Bioinformatique Site Doua

Download

<https://doua.prabi.fr/software/seaview>

Presentation

Online Services

Databases

Software and packages

Miscellaneous

Legacy

Partners



Use Linux,
por favor

SeaView - Multiplatform GUI for molecular phylogeny

Version 5.0.5

NEW: seaview performs reconciliation between gene and species trees using **Treerecs** version 1.2
NEW: bootstrap support optionally with the "Transfer Bootstrap Expectation" method
NEW: trimming-rule to shorten long sequence names in phylogenetic trees
NEW: 64-bit version for the MS Windows platform
NEW: multiple-tree windows
NEW: seaview uses **PHYMLIP v3.696** to compute parsimony trees
NEW: seaview can be run without GUI using a command line
NEW: seaview drives the **PhyML v3.1** program to compute maximum likelihood phylogenetic trees.
NEW: seaview drives the **Gblocks** program to select blocks of conserved sites.

SeaView is a multiplatform, graphical user interface for multiple sequence alignment and molecular phylogeny.

- SeaView reads and writes various file formats ([NEXUS](#), MSF, CLUSTAL, FASTA, PHYLIP, [MASE](#), Newick) of DNA and protein sequences and of phylogenetic trees.
- SeaView drives programs [muscle](#) or [Clustal Omega](#) for multiple sequence alignment, and also allows to use any external alignment algorithm able to read and write FASTA-formatted files.
- SeaView drives the [Gblocks](#) program to select blocks of evolutionarily conserved sites.
- SeaView computes phylogenetic trees by
 - parsimony, using PHYLIP's [dnapars/protpars](#) algorithm,
 - distance, with [NJ](#) or [BioNJ](#) algorithms on a variety of evolutionary distances,
 - maximum likelihood, driving program [PhyML](#) 3.1.
- SeaView can use the [Transfer Bootstrap Expectation](#) method to compute the bootstrap support of PhyML and distance trees.
- SeaView uses the [Treerecs](#) method to reconcile gene and species trees.
- SeaView prints and draws phylogenetic trees on screen, SVG, PDF or PostScript files.
- SeaView allows to download sequences from EMBL/GenBank/UniProt using the Internet.

Screen shots of the main [alignment](#) and [tree](#) windows. Dialog window to perform [Maximum-Likelihood](#) tree-building.
On-line [help](#) document.Old [seaview version 3.2](#)

Download SeaView



[macOS X](#)
ready for macOS 10.3 - 13.0



[32-bit Linux on x86](#)
[64-bit Linux on x86_64](#)



[MS Windows](#) 100% CLEAN
self-extractable archive



[Solaris on SPARC](#)

Bioinformática

Curso de Especialização na Área da Saúde – 2025/2026
Bases Técnicas em Vacinas e Biofármacos

PhD Flavio Lichtenstein

Bioinformatics, Systems Biology, and Biostatistics

dúvidas: flavio.lichtenstein@butantan.gov.br

abril/2025 e maio/2023