INSTITUTO BUTANTAN

Escola Superior
Instituto Butantan

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

centre of excellence in new target discovery

abril/2025 ~ maio/2023

O que vimos até aqui?

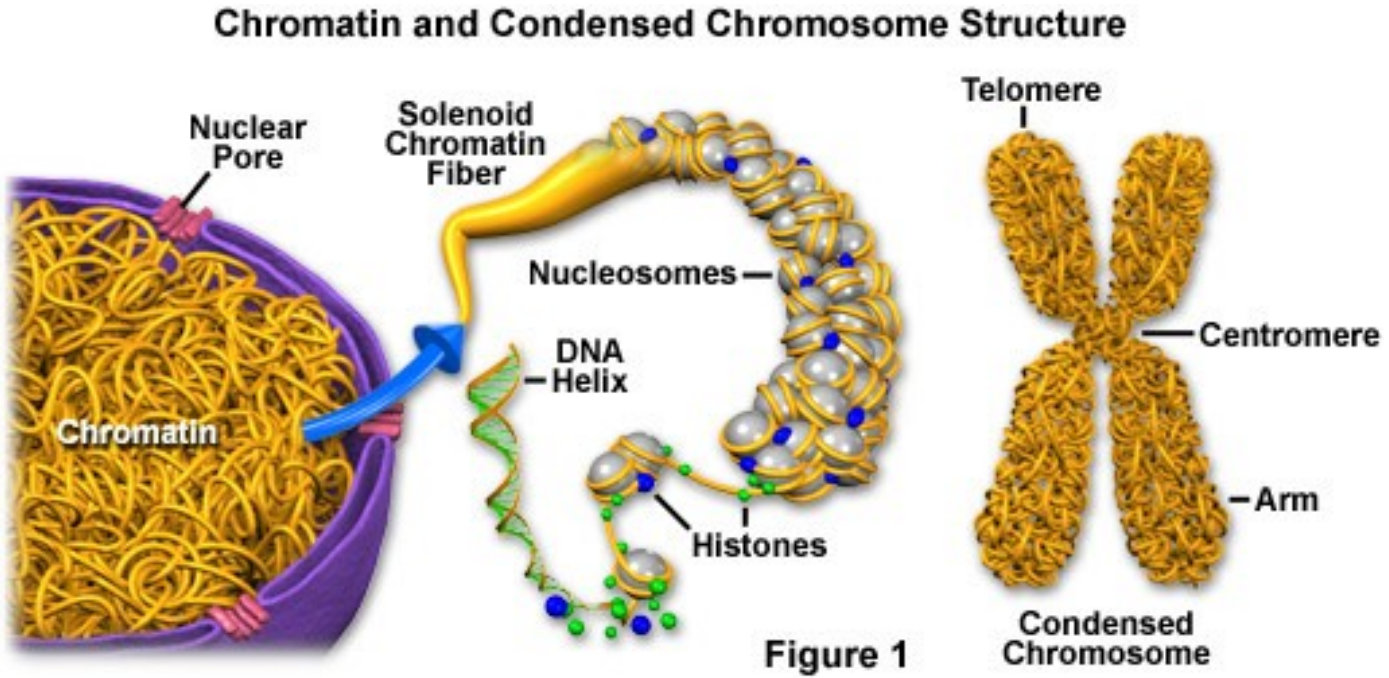
Constituição de um organismo eucarioto

- 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.

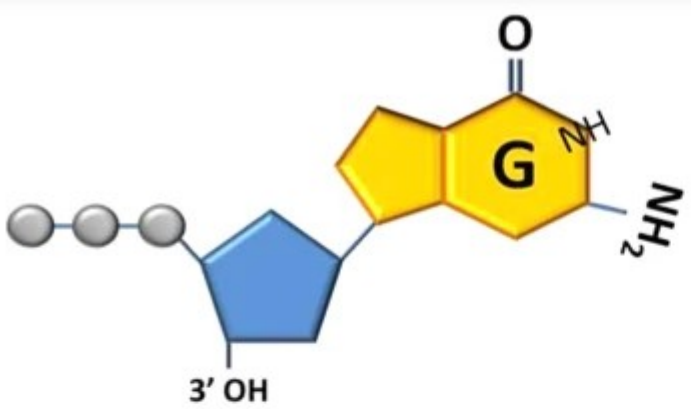
Os cromossomos estão enovelados dentro do núcleo



https://micro.magnet.fsu.edu/cells/nucleus/chromatin.html

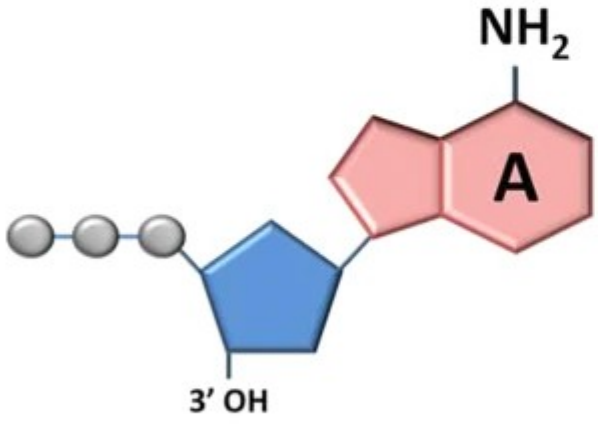
Os quatro nucleotídeos existenste no cromossomo

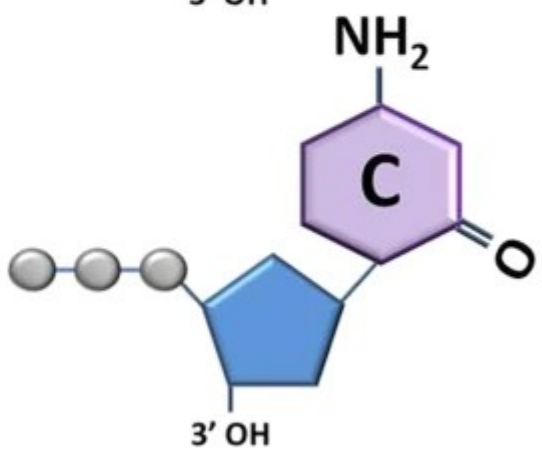
Guanina



Timina

Adenina





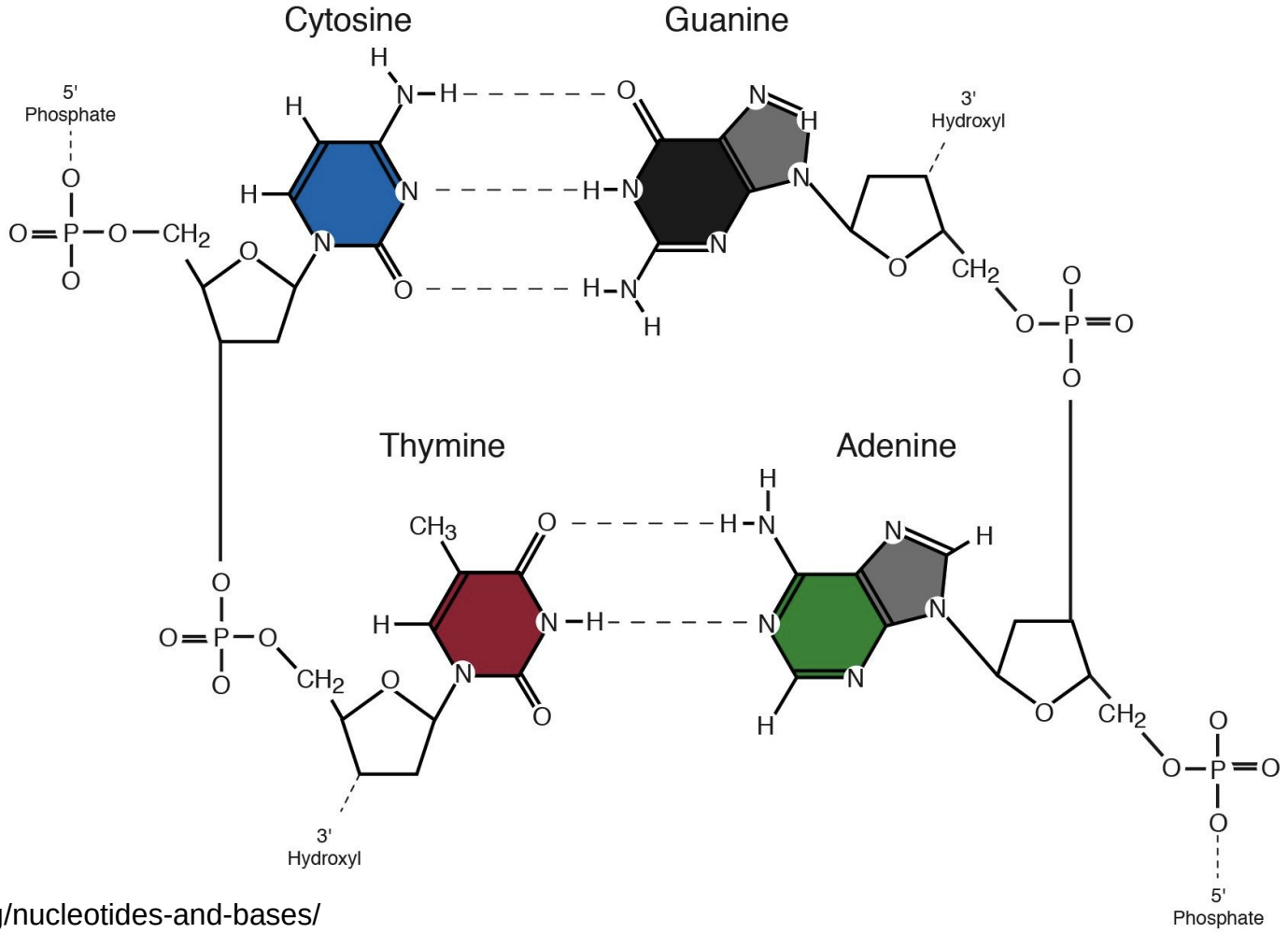
Citosina

Purine Deoxyribonucleotides

Pyrimidine Deoxyribonucleotides

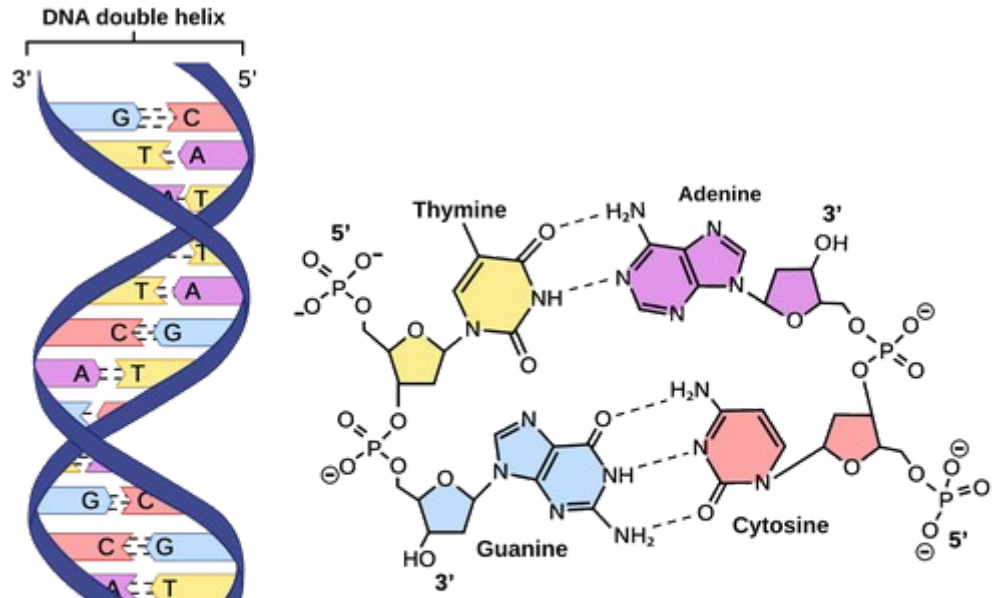
Ahmad, M., Panicker, N., Rizvi, T. et al. Electrical detection and quantification of single and mixed DNA nucleotides in suspension. Sci Rep 6, 34016 (2016). https://doi.org/10.1038/srep34016

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



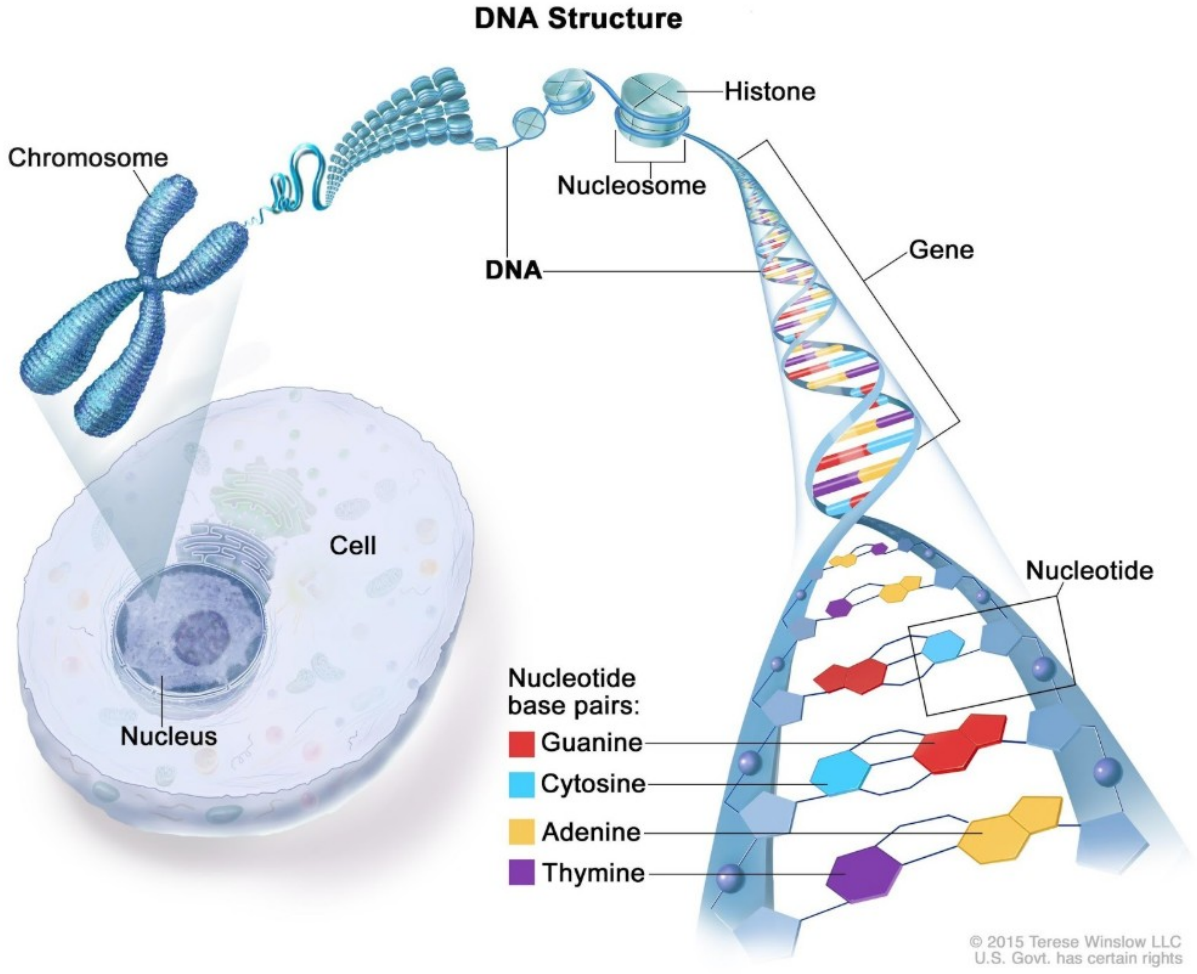
https://knowgenetics.org/nucleotides-and-bases/

Fita dupla de DNA ~ dupla hélice



https://knowgenetics.org/nucleotides-and-bases/

Hierarquia do cromossomo



https://project8p.org/chromosome-101/

Como estudar tudo isto?

Áreas de estudo

- Bioquímica & Química
- Biofísica
- Biologia Molecular
- Biologia Celular
- Genética
- …
- e onde entra a metemática?

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 eltrônico da Mt

Qual o comprimento

Quantos 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.

Qu[illegible]

Quantos genes existem no ge[illegible] 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.

Quantos genes existem no genoma [illegible]condrial?

O que é transferência horizontal e vertical de genes?

O que é bioinformática?

O que é imunoinformática?

O que é bioinformática

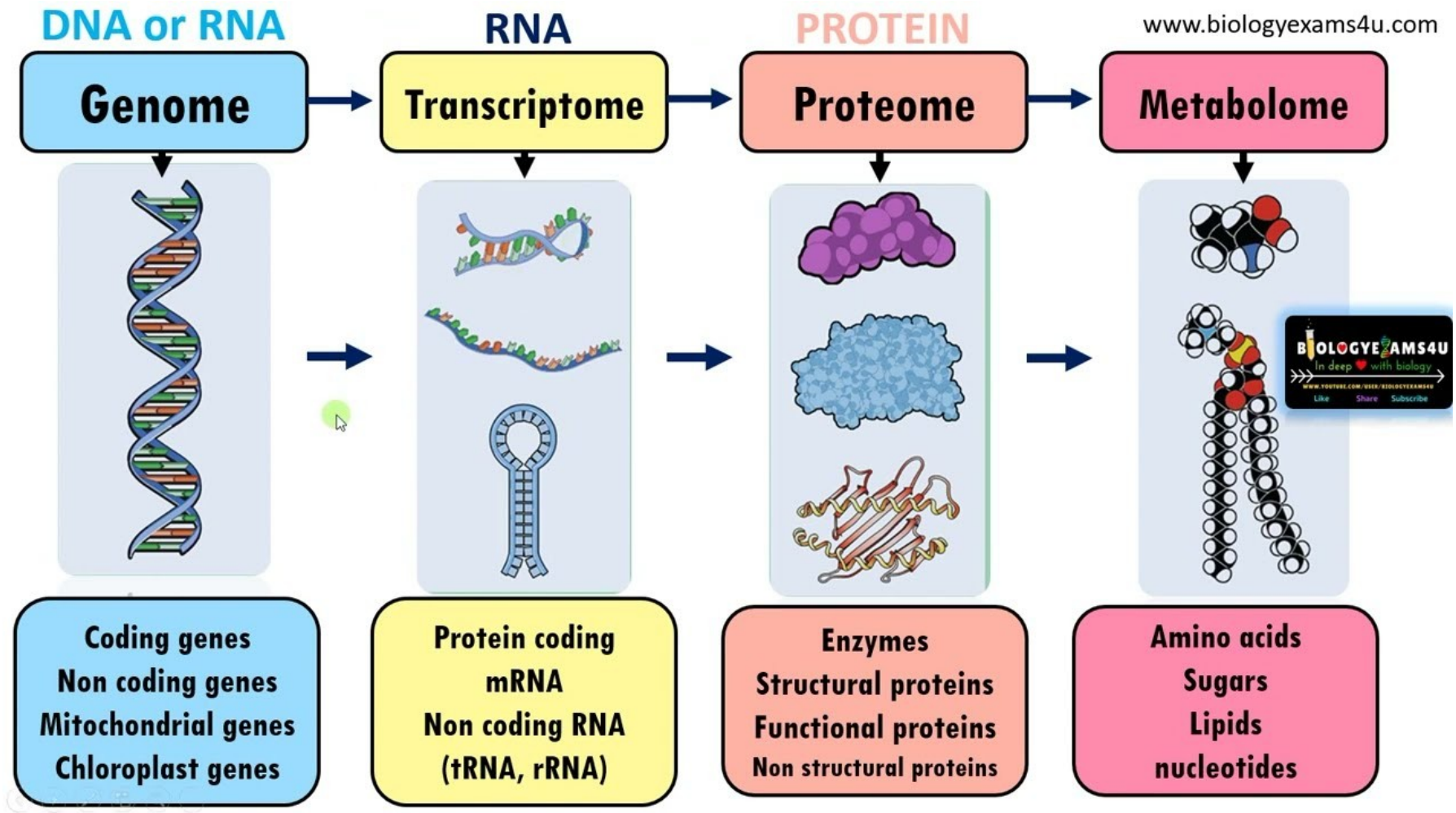
Bioinformática é o estudo de algoritmos matemáticos, aprendizado de máquina e bioestatística aplicados para melhor compreender sequências de nucleotídeos e proteínas e os resultados de experimentos das ômicas.

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

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

https://www.youtube.com/watch?v=FJ5iN-v0MFM

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



https://www.youtube.com/watch?v=FJ5iN-v0MFM

… 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”

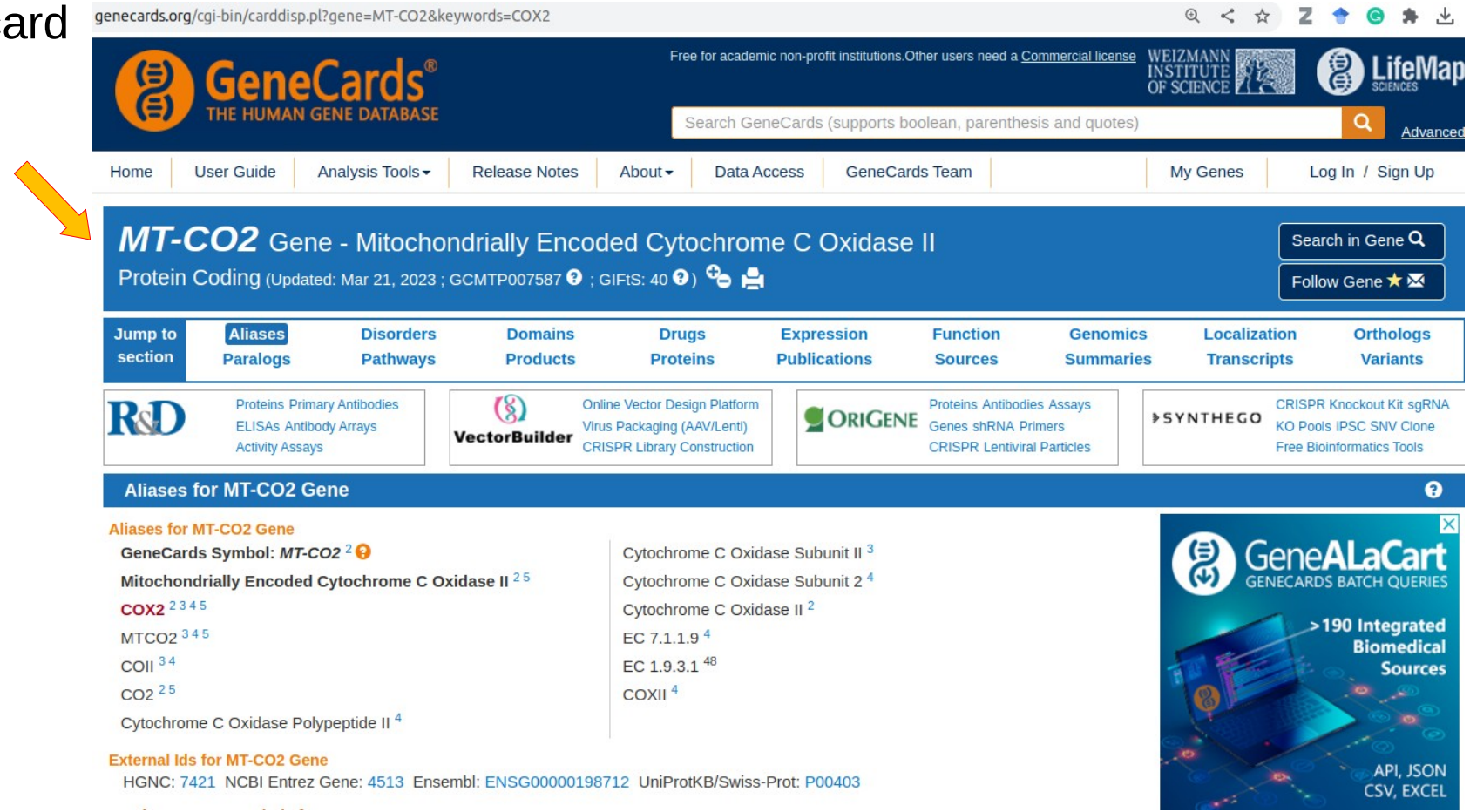
  https://www.genome.jp/dbget-bin/www_bget?dha:cox2

- Como posso localizá-lo?

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

COX2 na verdade é MT-CO2

Genecard



https://www.genecards.org/

COX2 ou MT-CO2 no KEGG

Debaryomyces hansenii: cox2

Help

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 BRITE hierarchy BRITE hierarchy
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 MIWTDVPTPWGMRFQDAATPNAEGMHELYDHMMYYLALMLGLVSYMLYVMMKDYKNNTFA YKYIKHGQTLEIMWTMFPAVMLLLMAFPSFMLLYLCDEVLTPAMTVKVVGLQWYWKYEYS DFVSETGETVEYESYVMPEDMLEEGQLRLLDTDTSMVVPVDTHVRFMVTANDVLHCFTMP SLGIKVDACPGRLNQVSALMQRTGVYYGQCSELCGVNHGLMPIKTECVPIGDFVEWLGEQ ENVYVA
NT seq	741 nt NT seq +upstream 0 nt +downstream 0 nt atgatttgactagatgtaccaacaccttgaggaatacgtttccaagatgccgcaacacct

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

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

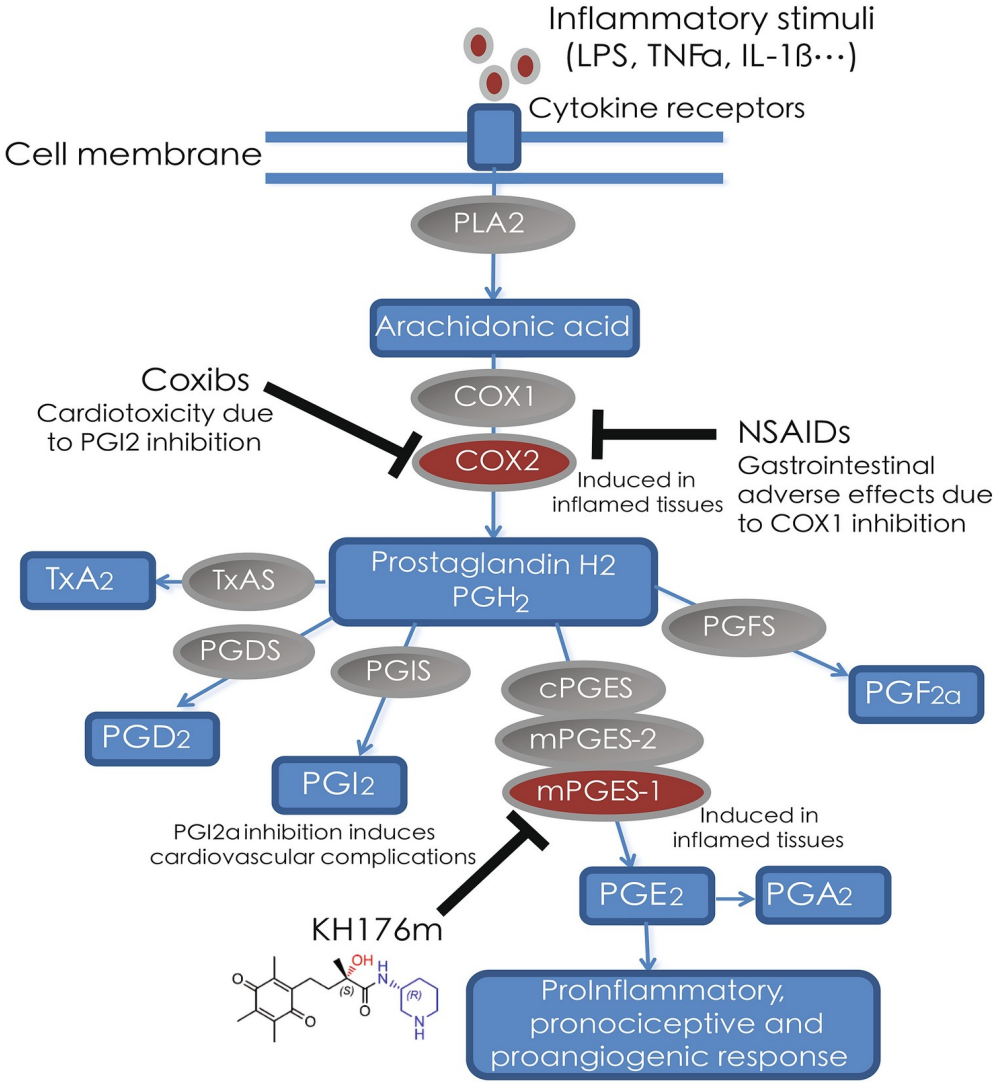
KEGG Ortology

ORTHOLOGY: K11987

Help

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

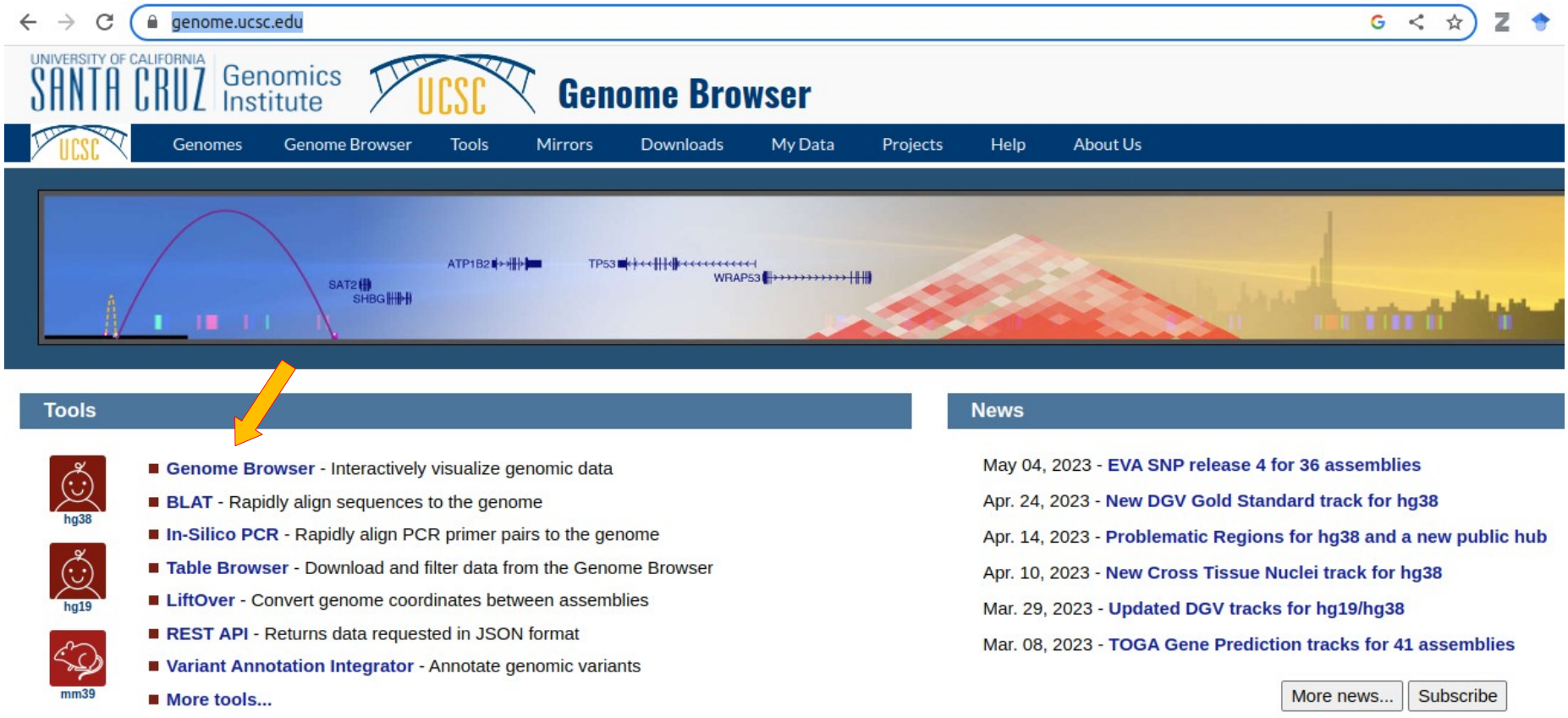
Exemplo de um pahtway contendo COX2



Jiang, X., Renkema, H., Pennings, B.Â et al.Â Mechanism of action and potential applications of selective inhibition of microsomal prostaglandin E synthase-1-mediated PGE2Â biosynthesis by sonlicromanolâ€™s metabolite KH176m.Â Sci RepÂ 11, 880 (2021). https://doi.org/10.1038/s41598-020-79466-w

Vamos achar genes? e regiões intergênicas...

UCSC genome browser



https://genome.ucsc.edu/

Vamos achar genes? e regiões intergênicas...

UCSC genome browser

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

Browse/Select Species

POPULAR SPECIES

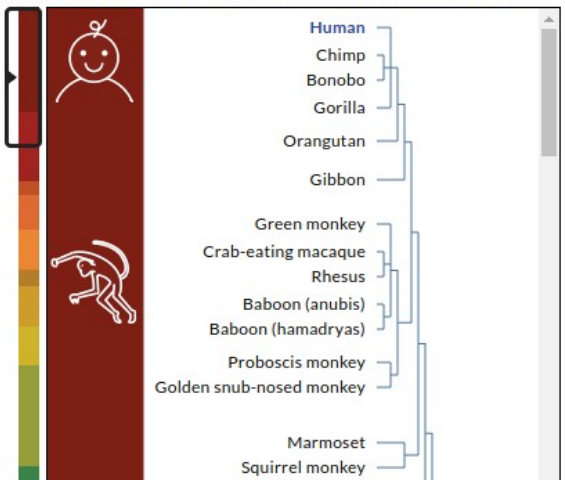
Human | Mouse | Rat | Zebrafish | Fruitfly | Worm | Yeast

Search through thousands of genome browsers

Enter species, common name or assembly ID

Unable to find a genome? Send us a request.

UCSC SPECIES TREE AND CONNECTED ASSEMBLY HUBS



Find Position

Human Assembly: Dec. 2013 (GRCh38/hg38)

Position/Search Term: Enter position, gene symbol or search terms

Current position: chr2:25,160,[illegible]25,168,903

Digite MT-CO2

Human Genome Browser - hg38 [illegible] view sequences

UCSC Genome Browser assembly ID: hg38
Sequencing/Assembly provider ID: Genome Refere[illegible]1405.29)
Assembly date: Dec. 2013 initial release; June 2022 pa[illegible]
Assembly accession: GCA_000001405.29
NCBI Genome ID: 51 (Homo sapiens (human))
NCBI Assembly ID: GCF_000001405.40 (GRCh38.p14, GCA_000001405.29)
BioProject ID: PRJNA31257

Homo sapiens
(Graphic courtesy of CBSE)

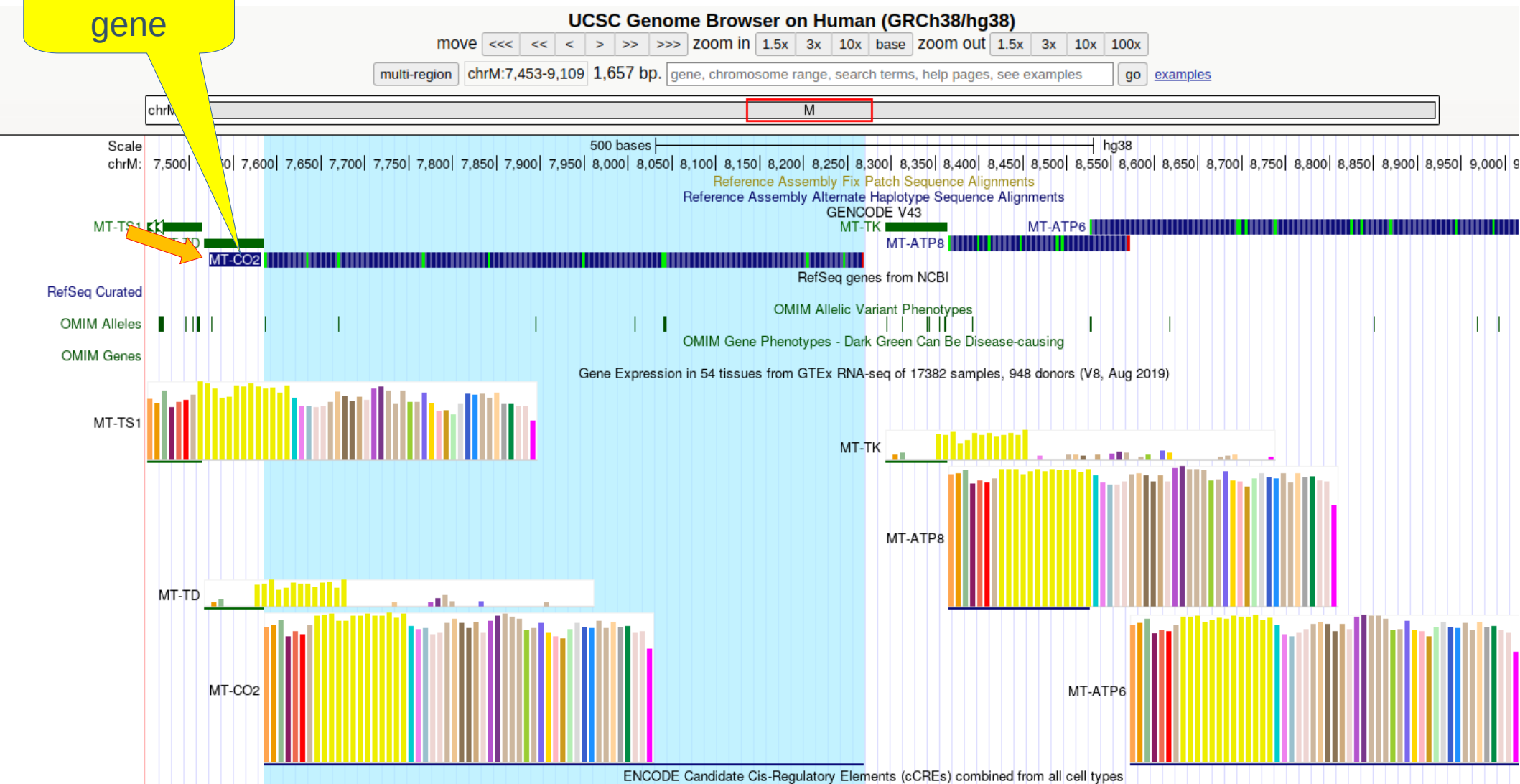
Search the assembly:

- **By position or search term:** Use the "position or search term" box to find areas of the genome associated with many different attributes, such as a specific chromosomal coordinate range; mRNA, EST, or STS marker names; or keywords from the GenBank description of an mRNA. More information, including sample queries.
- **By gene name:** Type a gene name into the "search term" box, choose your gene from the drop-down list, then press "submit" to go directly to the assembly location associated with that gene. More information.
- **By track type:** Click the "track search" button to find Genome Browser tracks that match specific selection criteria. More information.

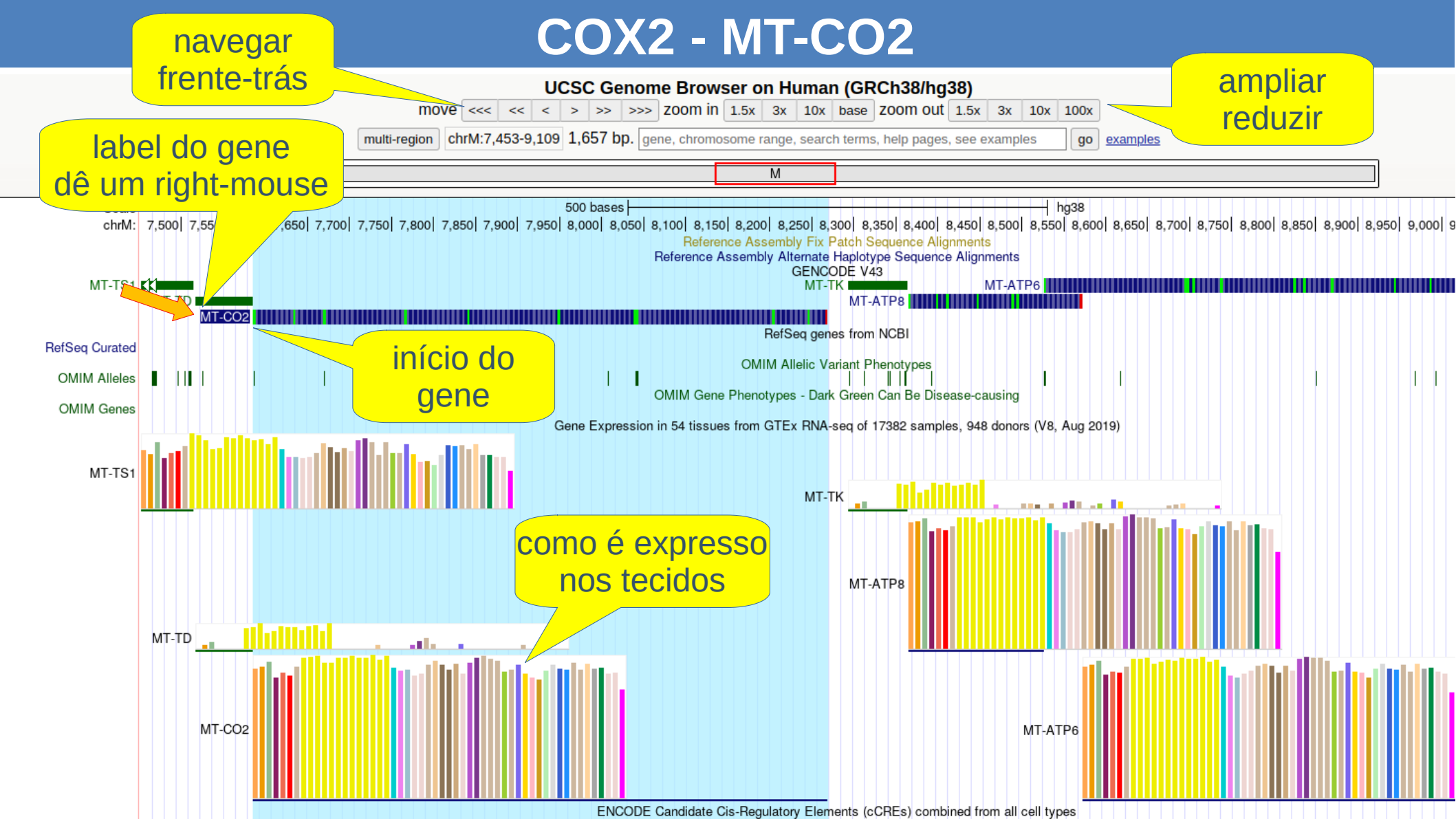
https://genome.ucsc.edu/cgi-bin/hgGateway

COX2 - MT-CO2

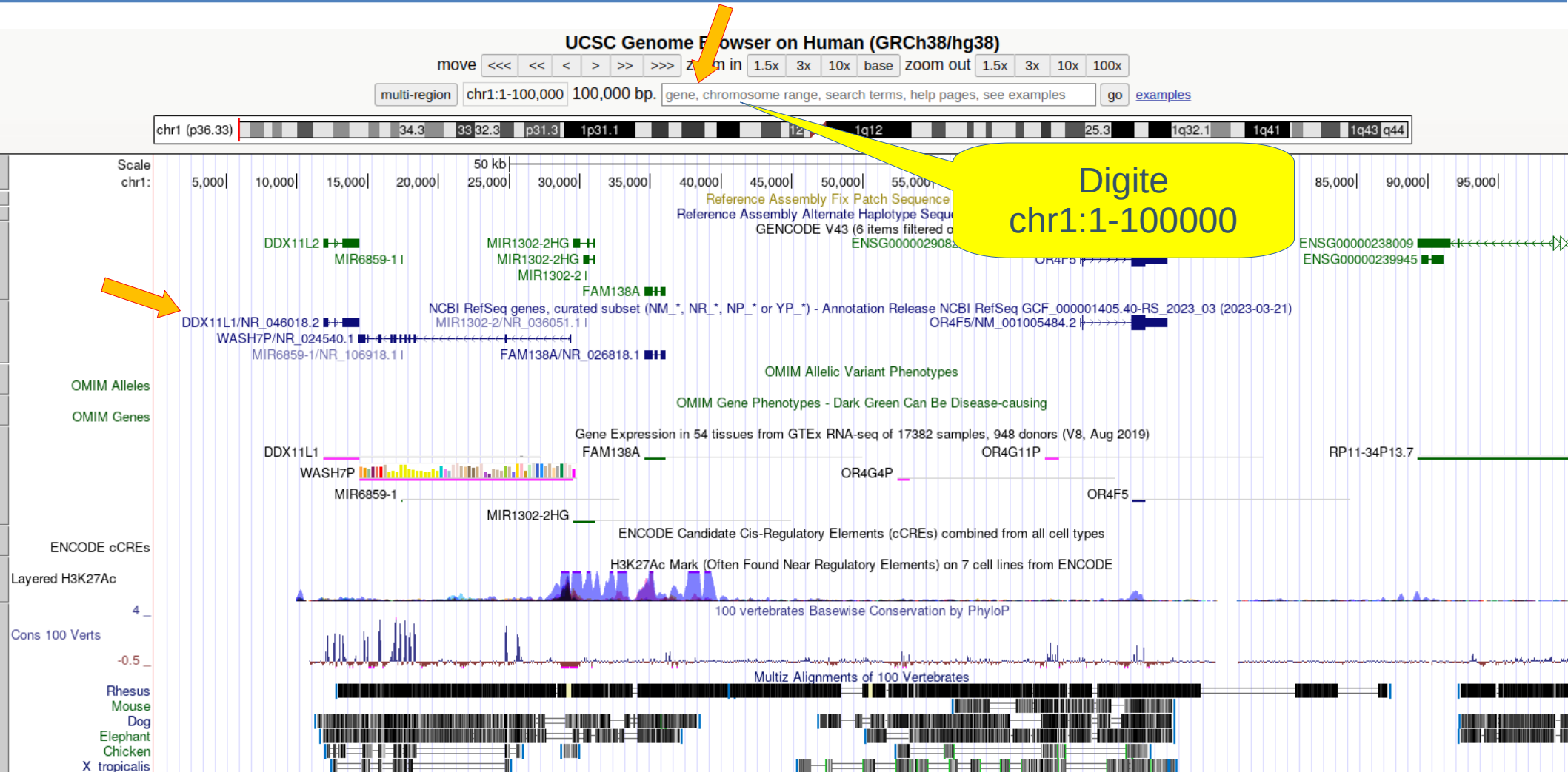
achei o gene



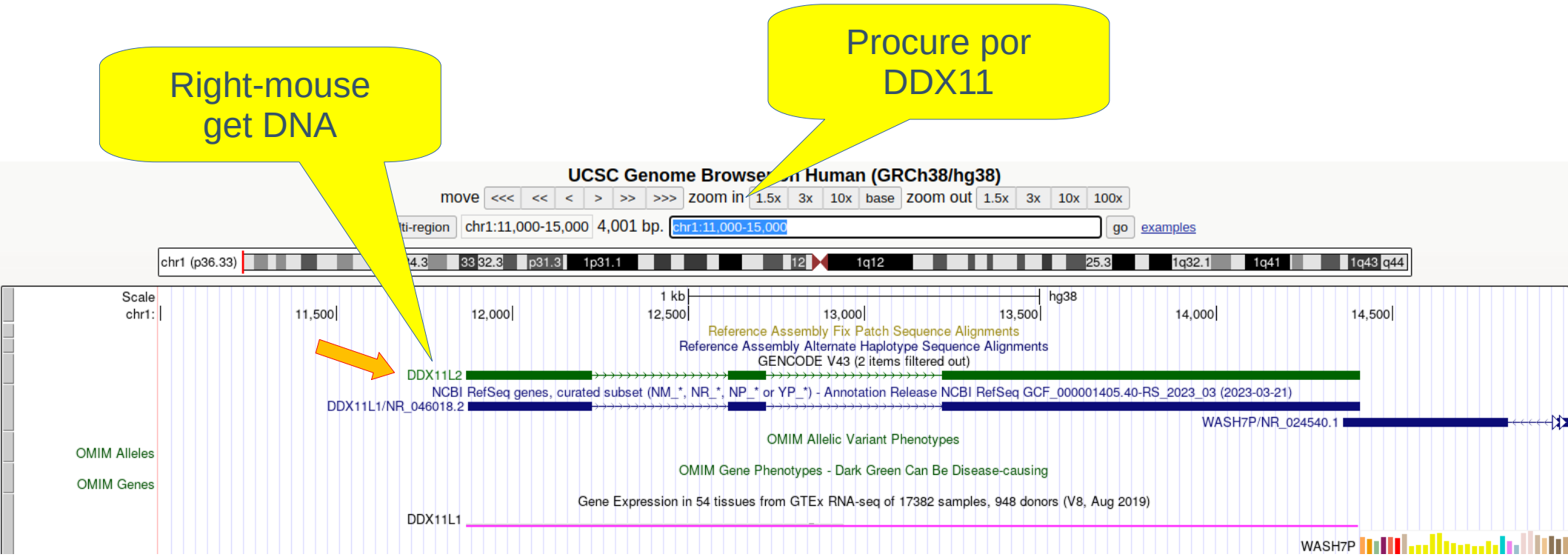
COX2 - MT-CO2


navegar frente-trás
ampliar reduzir
label do gene dê um right-mouse
início do gene
como é expresso nos tecidos
UCSC Genome Browser on Human (GRCh38/hg38)
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
M
500 bases hg38
chrM: 7,500 7,550 7,600 7,650 7,700 7,750 7,800 7,850 7,900 7,950 8,000 8,050 8,100 8,150 8,200 8,250 8,300 8,350 8,400 8,450 8,500 8,550 8,600 8,650 8,700 8,750 8,800 8,850 8,900 8,950 9,000
Reference Assembly Fix Patch Sequence Alignments
Reference Assembly Alternate Haplotype Sequence Alignments
GENCODE V43
MT-TS1
MT-TD
MT-CO2
MT-TK
MT-ATP6
MT-ATP8
RefSeq Curated
RefSeq genes from NCBI
OMIM Alleles
OMIM Allelic Variant Phenotypes
OMIM Genes
OMIM Gene Phenotypes - Dark Green Can Be Disease-causing
Gene Expression in 54 tissues from GTEx RNA-seq of 17382 samples, 948 donors (V8, Aug 2019)
MT-TS1
MT-TK
MT-ATP8
MT-TD
MT-CO2
MT-ATP6
ENCODE Candidate Cis-Regulatory Elements (cCREs) combined from all cell types

Cromossomo 1 de 1:100000



Zoom: chr1:11,000-15,000

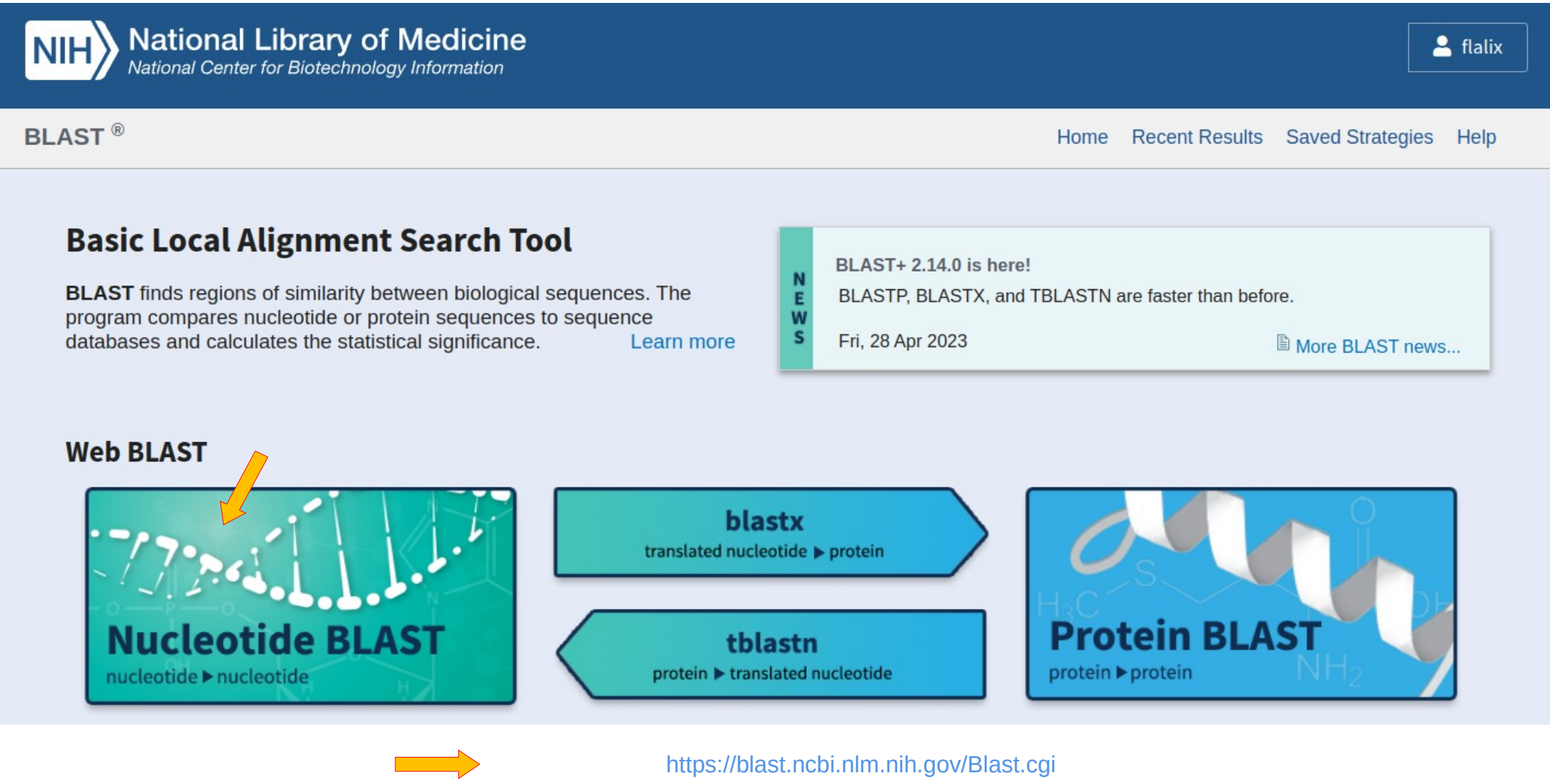


Vamos fazer um BLAST de nucleotídeos

Basic Local Alignment Search Tool

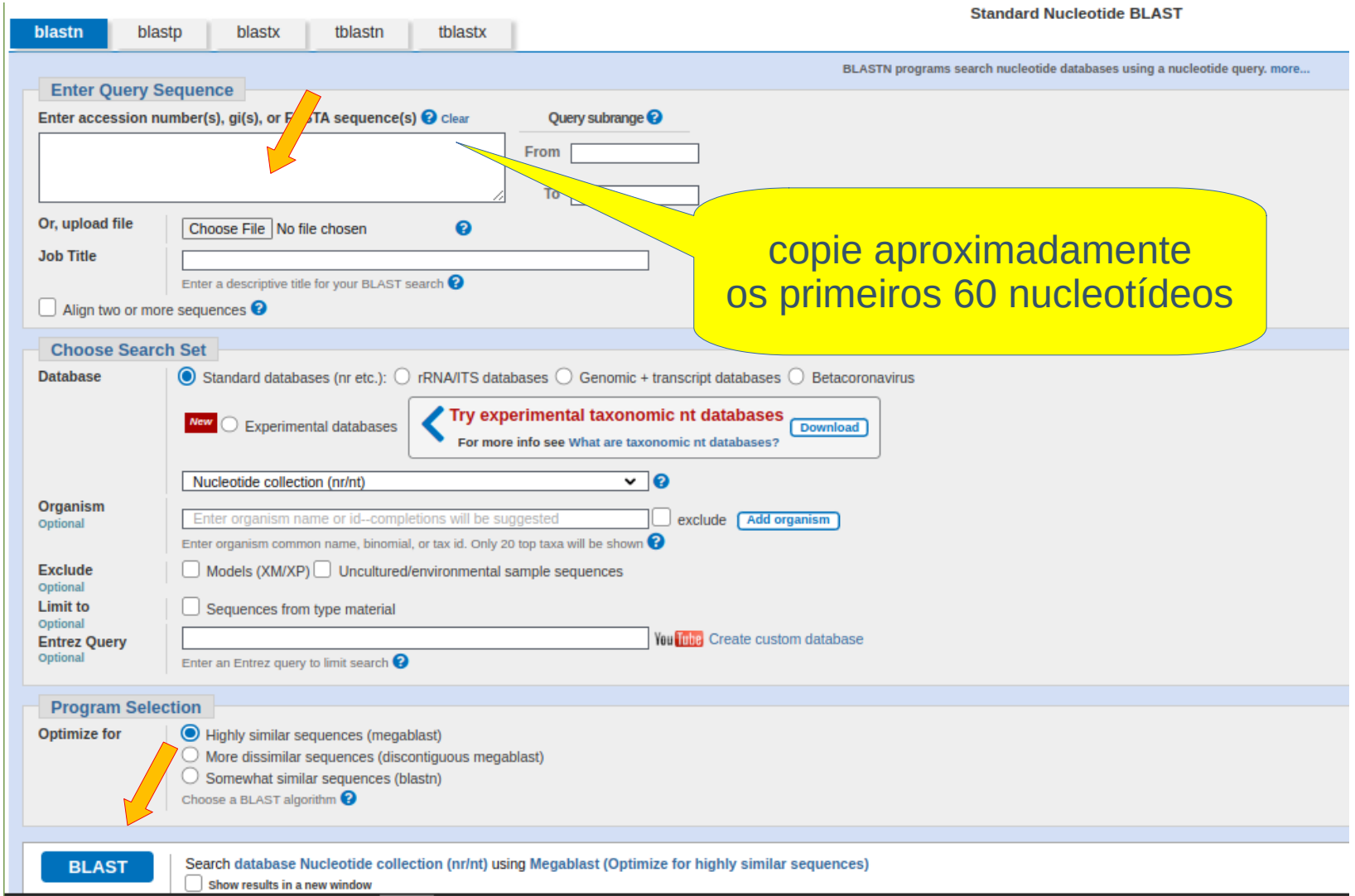
https://blast.ncbi.nlm.nih.gov/Blast.cgi

BLAST - nucleotideos



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

BLAST – copie a sequência e dê um blast



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

Type common name, binomial, taxid or group name

+ 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 100

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
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
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
Homo sapiens chromosome 12 clone CH17-321K24, complete sequence	Homo sapiens	2632	2632	100%	0.0	99.45%	210014	AC277891.1
Homo sapiens FOSMID clone ABC13-48569500C5 from chromosome 12, complete sequence	Homo sapiens	2503	2503	100%	0.0	97.80%	37211	AC215997.4
Homo sapiens BAC clone CH17-138H12 from chromosome 12, complete sequence	Homo sapiens	2497	2497	100%	0.0	97.73%	211866	AC254631.1
Homo sapiens 12p13 PAC RPCI11-726G1 (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 similar

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 ty	
Query Leng	
Other repo	ewer

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

to to to

Filter Reset

Descriptions | Graphic Summ | Alignments | Taxonomy

Sequences producing significant a ments Download Sel columns Show 100

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
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
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	8013.8
Homo sapiens chromosome 12 clone CH17-321K24, complete sequence	Homo sapiens	2632	2632	100%				
Homo sapiens FOSMID clone ABC13-48569500C5 from chromosome 12, complete sequence	Homo sapiens	2503	2503	100%				
Homo sapiens BAC clone CH17-138H12 from chromosome 12, complete sequence	Homo sapiens	2497	2497	100%	0.0	97.73%	211866	AC254631.1
Homo sapiens 12p13 PAC RPCI11-726G1 (Roswell Park Cancer Institute Human PAC Library) complete s...	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
GGCGTTTGCCCTGATTCCCGGGGCCTGGCTTTCAGCGTAGCAATTCTGCC
GGCGAAGAAGGTGAGCGCAGTGCTGTGTGGCAGCAGAGCTCCTTAGGACG
AGGAGCAGCGGGACGAGGAAGGGCAGACTGGTGAAATCGCAAACTGGGCG
TCTGTTCCGGCGCCGGACCCCTATTTGCAAAGGTGGGTGGCCCGTCCGGA
GCGGGAAAACATTCCGGAAGTGGAGGGCCGGGCCAGCGTGATTGACAAGC
GGGAACCCCTGTGTGGGGACGGGTAGGCCTAGGAAGGTTGTGCCTGCGGT
GGAACTGGGCGGTGCGCAGAAGTGGGCATTAACAGCAGCCGCGTGTCTGG
GTCTTAGATTTGGCCCAGCTGTGTTGAGCTTTTCATGGATTATCTTACGT
AGATAAGACACTGCAACAGTGAGTGAGCGCTTGTAACCCACCTGTCTCTT
ACGGAAACTGAGCCCCAGCGATGCTAACTTTAGCAAGGATACAGCTGGGA
TCCTAAACTTGGCAATCAGAGCCCAGAGCCGATAAAGTTAGCTGCATGAG
TCTAGCTTCCCCCAGGGCGGGAATCGAGGCGGAGCAGGGTACAGTACGGA
GGCCAGGAATGAGTGCACTTGACCAGATTGTTGACGGAAGTGTCATAAAA
ATGGACTTAAATGCTGATAAGCAGCTGTTTGGGTTCCACACAGGGTGCGA
GCTCAGGATGCACGTTGGAGGGGACACAGGGCCAGAGCAAGGTGGGAATG
CGGGTATTATGGGCCAGGCCATCCTCCAGCTGGTGGAGCACCGCAGTACT
GCAGTGTGTGGCCCTGGCTTAACAGCAGTGCGGAAAAGCTTTTTTCTTGG
GGCTGTGGTGCTTTCCAGGTGTGTGAAGTTAAAACATTTTAGGGCCATAC
GGTAGATAGTACCTGCCACATAGTTGTCTTGGTCAGCCCCGGCTGCCATG
ACAAAATACCATAAACTGGGTGGCTTAGACACAGAAATGTGTTTTCTCAC
AGTCCCAGAGGCTTGGGATGTCCAAGATCAATGTGGGGAGGACACAACTG
AGCCCACAACAATAGTCTGTACTCAGTAATGAGTGATAATTGGGGACTGA
AGAAAATGAACACGTTAAGAACTAATATGTTCCTGAAGTGCTTTCACAAC
TCTAACCTCGTTTTATGAGCGTGAGCTTTGCTGTCCTGGTGTGTGCCTTG
GCACTGGGAGGTGATGGTTGTCCTCCACACAGCCAACCTGAAGAGGGCTG
AACAAGTCACTGCAAATGTTTTTAATAGGGCTTAGTGAATCCGTTATACT
CAGATTTATCTAAACCTCTATGATTTAGCCTGTGCTGCTTCTGGAATAAT
GAGATCCATAATTACCACTGATGGGGAAGTGAAATAATACTTACGTTTCT
TCTGTATTGTTTAGTGTTGTCATTCTAGAGTTTGGTAACCAAGTCTGTCT
TTTTTATCCCACTTATCCTGGTGGGAGGAAAAGTGAGGAGATAGAAAGTT
TCAGGTGGCTTGGGGGTCTGGCAGATGTGGTTCAAATCCTGAGTTCAAGC
ACTTGCTGAGTGACCTTGGGCAAGTCATATAAGTTTACTGAGTCTCAGTT
TCTTTCTCTATAAAATAGAGCTTATAAAAATACCACACAGGGTTTTTGTG
GGGTTATCTGAAAACGGTTTGAAACCATTAAAGAACTGGCCATTTAACTA
```

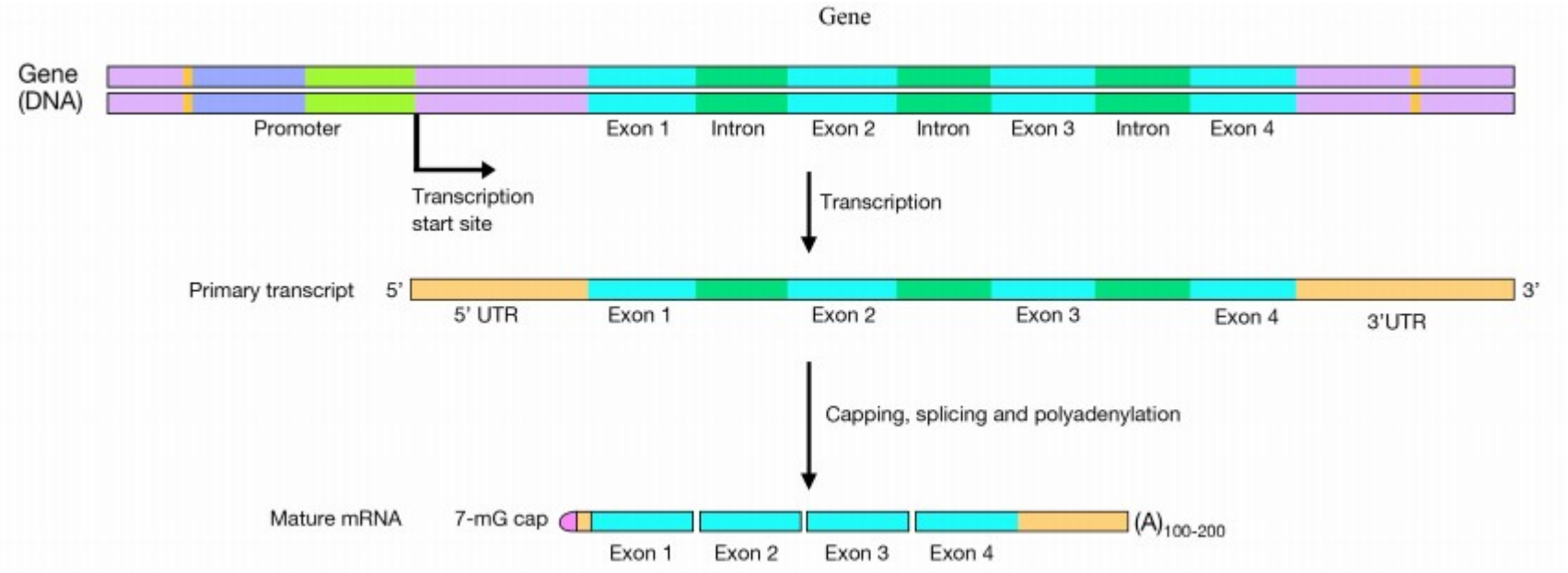
Aonde estão os Éxons?
Onde começa este gene? start-codon ATG

```
>hg38_dna range=chr12:31073860-31104799 5'pad=0 3'pad=0 strand=+ repeatMasking=none
GTTGTTCCGGCTGCCTTTCACTGAGGGGACCCGCCAGTTTCTAACTCAGT
GGCGTTTGCCCTGATTCCCGGGGCCTGGCTTTCAGCGTAGCAATTCTGCC
GGCGAAGAAGGTGAGCGCAGTGCTGTGTGGCAGCAGAGCTCCTTAGGACG
AGGAGCAGCGGGACGAGGAAGGGCAGACTGGTGAAATCGCAAACTGGGCG
TCTGTTCCGGCGCCGGACCCCTATTTGCAAAGGTGGGTGGCCCGTCCGGA
GCGGGAAAACATTCCGGAAGTGGAGGGCCGGGCCAGCGTGATTGACAAGC
GGGAACCCCTGTGTGGGGACGGGTAGGCCTAGGAAGGTTGTGCCTGCGGT
GGAACTGGGCGGTGCGCAGAAGTGGGCATTAACAGCAGCCGCGTGTCTGG
GTCTTAGATTTGGCCCAGCTGTGTTGAGCTTTTCATGGATTATCTTACGT
AGATAAGACACTGCAACAGTGAGTGAGCGCTTGTAACCCACCTGTCTCTT
ACGGAAACTGAGCCCCAGCGATGCTAACTTTAGCAAGGATACAGCTGGGA
TCCTAAACTTGGCAATCAGAGCCCAGAGCCGATAAAGTTAGCTGCATGAG
TCTAGCTTCCCCCAGGGCGGGAATCGAGGCGGAGCAGGGTACAGTACGGA
GGCCAGGAATGAGTGCACTTGACCAGATTGTTGACGGAAGTGTCATAAAA
ATGGACTTAAATGCTGATAAGCAGCTGTTTGGGTTCCACACAGGGTGCGA
GCTCAGGATGCACGTTGGAGGGGACACAGGGCCAGAGCAAGGTGGGAATG
CGGGTATTATGGGCCAGGCCATCCTCCAGCTGGTGGAGCACCGCAGTACT
GCAGTGTGTGGCCCTGGCTTAACAGCAGTGCGGAAAAGCTTTTTTCTTGG
GGCTGTGGTGCTTTCCAGGTGTGTGAAGTTAAAACATTTTAGGGCCATAC
GGTAGATAGTACCTGCCACATAGTTGTCTTGGTCAGCCCCGGCTGCCATG
ACAAAATACCATAAACTGGGTGGCTTAGACACAGAAATGTGTTTTCTCAC
AGTCCCAGAGGCTTGGGATGTCCAAGATCAATGTGGGGAGGACACAACTG
AGCCCACAACAATAGTCTGTACTCAGTAATGAGTGATAATTGGGGACTGA
AGAAAATGAACACGTTAAGAACTAATATGTTCCTGAAGTGCTTTCACAAC
TCTAACCTCGTTTTATGAGCGTGAGCTTTGCTGTCCTGGTGTGTGCCTTG
GCACTGGGAGGTGATGGTTGTCCTCCACACAGCCAACCTGAAGAGGGCTG
AACAAGTCACTGCAAATGTTTTTAATAGGGCTTAGTGAATCCGTTATACT
CAGATTTATCTAAACCTCTATGATTTAGCCTGTGCTGCTTCTGGAATAAT
GAGATCCATAATTACCACTGATGGGGAAGTGAAATAATACTTACGTTTCT
TCTGTATTGTTTAGTGTTGTCATTCTAGAGTTTGGTAACCAAGTCTGTCT
TTTTTATCCCACTTATCCTGGTGGGAGGAAAAGTGAGGAGATAGAAAGTT
TCAGGTGGCTTGGGGGTCTGGCAGATGTGGTTCAAATCCTGAGTTCAAGC
ACTTGCTGAGTGACCTTGGGCAAGTCATATAAGTTTACTGAGTCTCAGTT
TCTTTCTCTATAAAATAGAGCTTATAAAAATACCACACAGGGTTTTTGTG
GGGTTATCTGAAAACGGTTTGAAACCATTAAAGAACTGGCCATTTAACTA
```

Está lá para baixo.
Mas como foi achado?

```
ACACGGTGAAACCCCGTCTCTACCAAAAAAAAAAA                GCGTGGT
GGCGGGCACCTGTAGTCCCAGCTACTCGGGAGGCT                AGAATGG
TGTGAACCCGGGAGGCGGAGCGTGCAGTGAGCCGA                CACTGCA
CTCCAGCCTGGGCGACAGAGTGAGACTCTGTCTCGA                AAAAAAAA
AAGATAGGGACGTGTATGTTAACTTGCGCTATCCAA                AGCTGTGC
TTATGGTCCTCTGCCTGTGCGTCATGATTTTCCAGG                ACAACGGGA
TAAAGTGAAGTAGCTTCGGCTTGTGAATGTGCATTG                ACGTGGGAG
AAGAAAGCTGCAAAAGTCATTATAAGCAACACCCTTC               TTAGGGGTGC
TGGTCTGTGAGAAGAGAGCTCTAAGCCTTTTGTAGAG               TTAGAAAGGT
ATCGGAGCCACGGTGAAATGCAGGGGAGATTGGGTTT               GGGCTTTCCTG
GTCTGCATTCTGCTACAGCCGTTAAATGCCGCTAGAT               AGTGCGTGATT
CTGGTATGGCCTCACGTGGACCTGCTGCGAAGGATGGA              AGAACATGGTC
TCTGCTTCCCAGAAAAAAGGAGAAATTTGGTAATAAGT              GTGGAGACTGCT
CTTAAATAATGCTCCAGATTTCAAGCCACTTCTTCCTGGACCATGAGAGA
GCTCCCTAATGTTGTATTTATTTTTCCTAGGTCCATGGCTAATGAAACAC
AGAAGGTTGGTGCCATCCATTTTCCTTTTCCCTTCACACCCTATTCCATC
```

Aonde estão os Éxons?
Onde começa este gene?



https://bio.libretexts.org/Bookshelves/Biochemistry/Book%3A_Biochemistry_Free_and_Easy_%28Ahern_and_Rajagopal%29/05%3A_Flow_of_Genetic_Information/5.05%3A_RNA_Processing

Localizando os éxons no texto do Genebank

```
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 tggctaatga aacacagaag gttggtgcca tccatttcc ttttccctc
9601 acaccctatt ccatccagga agacttcatg gcagagctgt accgggtttt ggaggctggc
9661 aagattggga tatttgagag tccaactggc actgtgagta tgaacagtga gagatactga
9721 aaaggacaac ttaacagcag ccgtactagc ttttcctgtt tgcctatcca gagattttca
```

Parte do registro do Genebank refere-se a anotações

outras bases do NCBI

Protein
CDS
HGNC (Hugo)
MIM

```
/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="MANETQKVGAIHFPFPFTPYSIQEDFMAELYRVLEAGKIGIFES
PTGTGKSLSLICGALSWLRDFEQKKREEEARLLETGTGPLHDEKDESLCLSSSCEGAA
GTPRPAGEPAWVTQFVQKKEERDLVDRLKAEQARRKQREERLQQLQHRVQLKYAAKRL
RQEEEERENLLRLSREMLETGPEAERLEQLESGEEELVLAEYESDEEKKVASRVDEDE
DDLEEEHITKIYYCSRTHSQLAQFVHEVKKSPFGKDVRLVSLGSRQNLCVNEDVKSLG
SVQLINDRCVDMQRSRHEKKKGAEEEKPKRRRQEKQAACPFYNHEQMGLLRDEALAEV
KDMEQLLALGKEARACPYYGSRLAIPAAQLVVLPYQMLLHAATRQAAGIRLQDQVVII
DEAHNLIDTITGMHSVEVSGSQLCQAHSQLLQYVERYGKRLKAKNLMYLKQILYLLEK
FVAVLGGNIKQNPNTQSLSQTGTELKTINDFLFQSQIDNINLFKVQRYCEKSMISRKL
FGFTERYGAVFSSREQPKLAGFQQFLQSLQPRTTEALAAPADESQASTLRPASPLMHI
QGFLAALTTANQDGRVILSRQGSLSQSTLKFLLLNPAVHFAQVVKECRAVVIAGGTMQ
PVSDFRQQLLACAGVEAERVVEFSCGHVIPPDNILPLVICSGISNQPLEFTFQKRELP
QMMDEVGRILCNLCGVVPGGVVCFFPSYEYLRQVHAHWEKGGLLGRLAARKKIFQEPK
SAHQVEQVLLAYSRCIQACGQERGQVTGALLLSVVGGKMSEGINFSDNLGRCVVMVGM
PFPNIRSAELQEKMAYLDQTLPRAPGQAPPGKALVENLCMKAVNQSIGRAIRHQKDFA
SVVLLDQRYARPPVLAKLPAWIRARVEVKATFGPAIAAVQKFHREKSASS"
```

clique aqui para ir para as anotações da proteína

clique aqui também para ver o conding 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

Export: File Fasta

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-Pedrera C.
  TITLE     Expression of DDX11 and DNM1L 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 DNM1L at the 12p11 Locus Modulates
            Systemic Lupus Erythematosus Susceptibility.
```

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

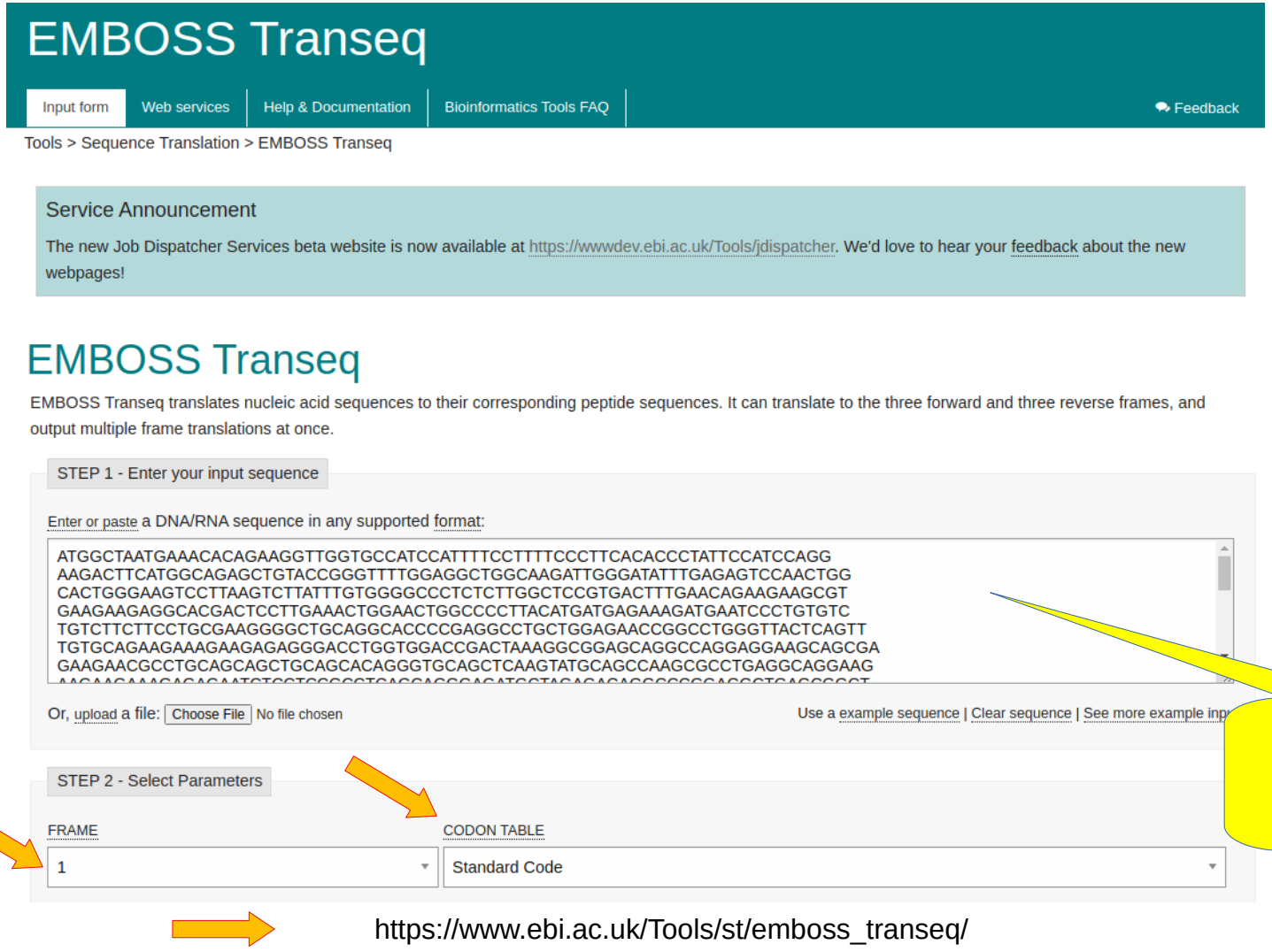
Copie as sequências CDS no conversor do EMBOSS

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

https://www.embl.org/

EMBL

The European Molecular Biology Laboratory



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-I20230518-184221-0857-82638710-p2m

Tool Output | Submission Details

Download | Show Colors

```
>EMBOSS_001_1
MANETQKVGAIHFPFPFTPYSIQEDFMAELYRVLEAGKIGIFESPTGTGKSLSLICGALS
WLRDFEQKKREEEARLLETGTGPLHDEKDESLCLSSSCEGAAGTPRPAGEPAWVTQFVQK
KEERDLVDRLKAEQARRKQREERLQQLQHRVQLKYAAKRLRQEEEERENLLRLSREMLET
GPEAERLEQLESGEEELVLAEYESDEEKKVASRVDEDEDDLEEEHITKIYYCSRTHSQLA
QFVHEVKKSPFGKDVRLVSLGSRQNLCVNEDVKSLGSVQLINDRCVDMQRSRHEKKKGAE
EEKPKRRRQEKQAACPFYNHEQMGLLRDEALAEVKDMEQLLALGKEARACPYYGSRLAIP
AAQLVVLPYQMLLHAATRQAAGIRLQDQVVIIDEAHNLIDTITGMHSVEVSGSQLCQAHS
QLLQYVERYGKRLKAKNLMYLKQILYLLEKFVAVLGGNIKQNPNTQSLSQTGTELKTIND
FLFQSQIDNINLFKVQRYCEKSMISRKLFGFTERYGAVFSSREQPKLAGFQQFLQSLQPR
TTEALAAPADESQASTLRPASPLMHIQGFLAALTTANQDGRVILSRQGSLSQSTLKFLLL
NPAVHFAQVVKECRAVVIAGGTMQPVSDFRQQLLACAGVEAERVVEFSCGHVIPPDNILP
LVICSGISNQPLEFTFQKRELPQMMDEVGRILCNLCGVVPGGVVCFFPSYEYLRQVHAHW
EKGGLLGRLAARKKIFQEPKSAHQVEQVLLAYSRCIQACGQERGQVTGALLLSVVGGKMS
EGINFSDNLGRCVVMVGMPFPNIRSAELQEKMAYLDQTLPRAPGQAPPGKALVENLCMKA
VNQSIGRAIRHQKDFASVVLLDQRYARPPVLAKLPAWIRARVEVKATFGPAIAAVQKFHR
EKSASS*
```

exercício

Exercício fazer um BLAST de amino ácidos

Pegue os 30 primeiros aa de DDX11 e faça um BLAST

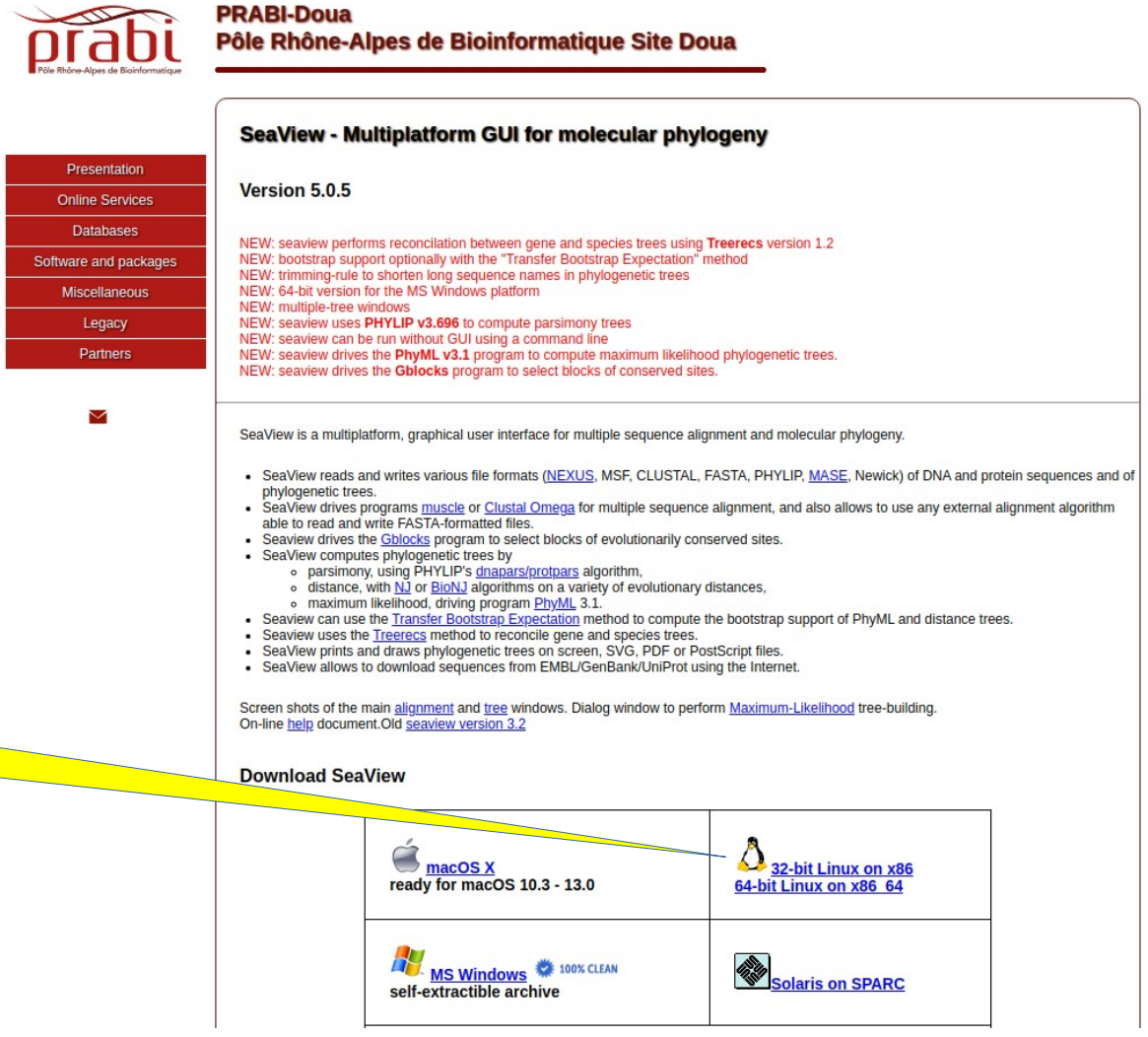
https://blast.ncbi.nlm.nih.gov/Blast.cgi

Mas, na verdade eu gosto do Seaview

Download

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

Use Linux,
por favor



INSTITUTO BUTANTAN

Escola Superior Instituto Butantan

Bioinformática

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

PhD Flavio Lichtenstein

centre of excellence in new target discovery

Bioinformatics, Systems Biology, and Biostatistics

dúvidas: flavio.lichtenstein@butantan.gov.br

abril/2025 e maio/2023