

Functional Annotation

Ana Conesa, Professor

Institute for Integrative Systems Biology , CSIC, Spain

ana.conesa@csic.es @anaconesa @conesa_lab

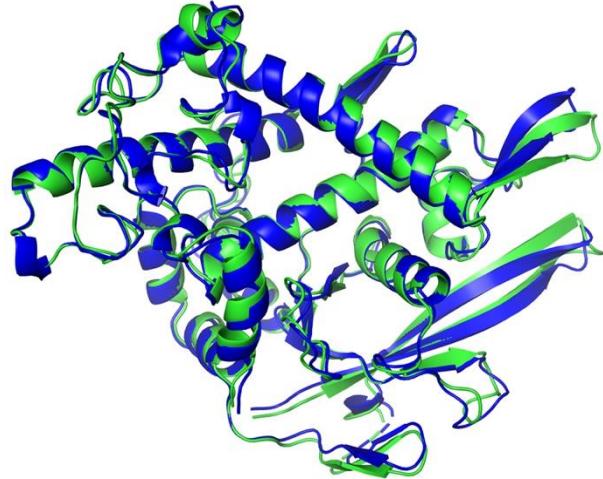


My cats, older

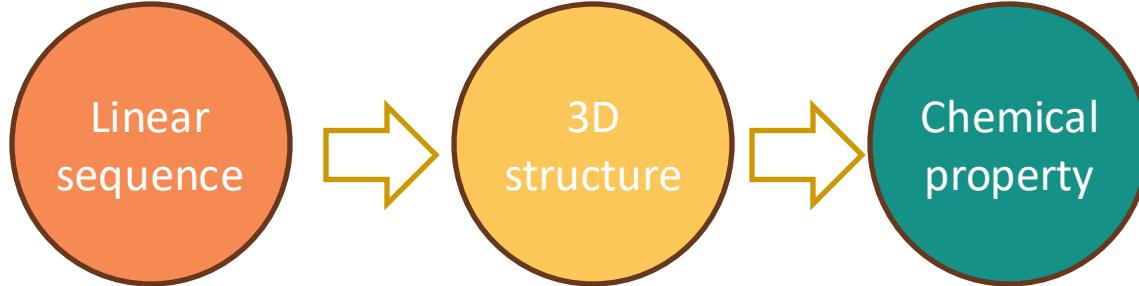
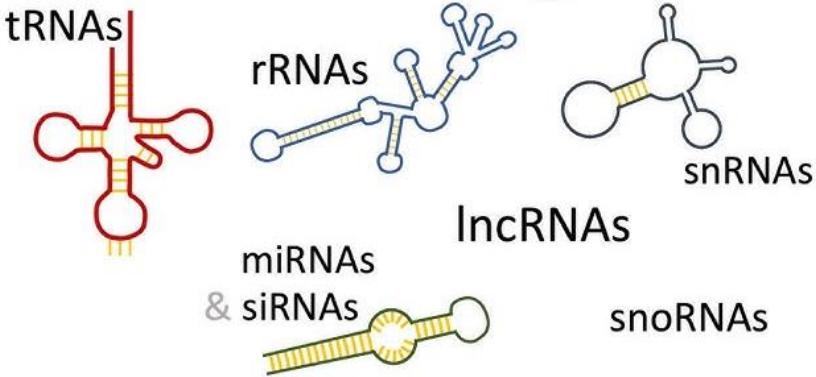


Functional annotation of biomolecules

Proteins



Non-coding RNAs



Why is this important to infer function?



Catalysis: Accelerate biochemical reactions as enzymes



Muscle contraction: Enable movement through interaction with muscle fibers



Structural support: Maintain cell and tissue integrity



Immune defense: Recognize and neutralize pathogens



Signaling: Transmit cellular messages and regulate activities

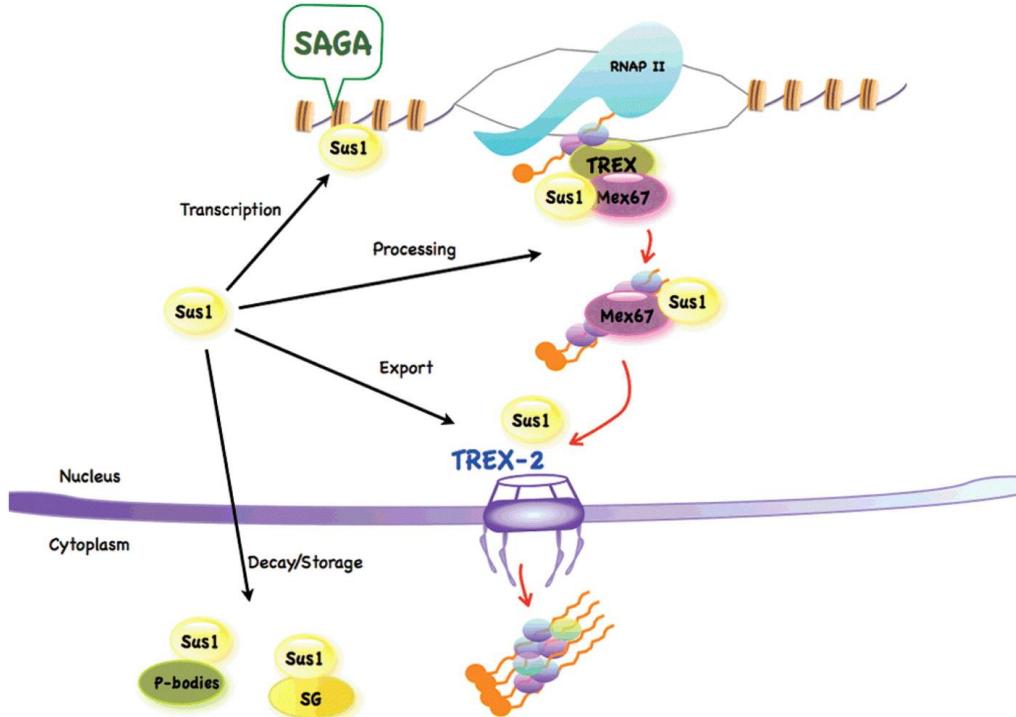


Storage: Store essential molecules for cell use

Why is this relevant?

-  Depends of the tissue
-  Depends of cellular location
-  Depends of interaction partners
-  Depends of the organism

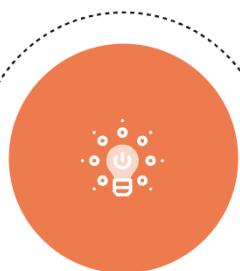
Sus1/ENY2: a multitasking protein in eukaryotic gene expression



What is this complicated?

Gene

A single gene contains multiple exons coding for protein sequences.



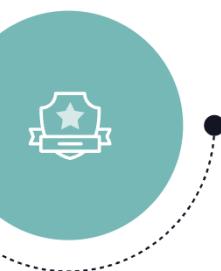
Protein Isoforms

Each mRNA isoform translates into distinct protein isoforms with varied sequences.



Protein Variants

Combinations of isoforms and over 400 PTM types generate numerous protein forms sharing the same canonical sequence but differing in function, complicating prediction.



mRNA Splicing

Exons are combined in different ways via alternative splicing, producing multiple mRNA isoforms.

Post-Translational Modifications (PTMs)

Chemical modifications like phosphorylation alter amino acids, expanding the protein's functional diversity.

Essential protein function databases

Database	Focus Area	URL
UniProtKB	Protein sequences and annotations	https://www.uniprot.org
Pfam	Protein sequence families	http://pfam.xfam.org
PDB	Protein 3D structures	https://www.rcsb.org
ModBase	Protein structure models	https://modbase.compbio.ucsf.edu
I2D	Protein-protein interactions	http://ophid.utoronto.ca/i2d
GEO	Gene expression data	https://www.ncbi.nlm.nih.gov/geo
PRIDE	Mass spectrometry proteomics data	https://www.ebi.ac.uk/pride

The need of computational annotation

Summary: Functional annotation of biomolecules

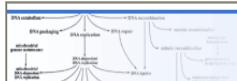
Gene Function

- Refers to the molecular function of a gene or a protein:
Tyrosine kinase
- One gene/protein may have more than one

Functional annotation

- Any aspect of functionality : Stress, nucleus
- Process of assigning function computationally
- Use of controlled vocabularies
- Multiple annotations per protein or gene

Functional Vocabularies



the Gene Ontology

Molecular Function
Biological Process
Cellular Component

Functional motifs



Example proteins

P25024 High affinity interleukin-8 receptor A (IL-8R A) (IL-8 rec)

More proteins

IPR000174 Interleukin-8 receptor



IPR000276 Rhodopsin-like GPCR superfamily



IPR001277 C-X-C chemokine receptor, type 4



IPR001355 Interleukin 8A receptor



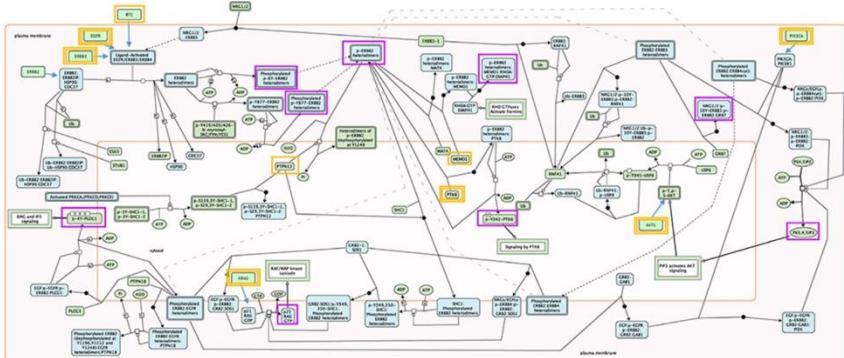
ModBase



PDB Chain



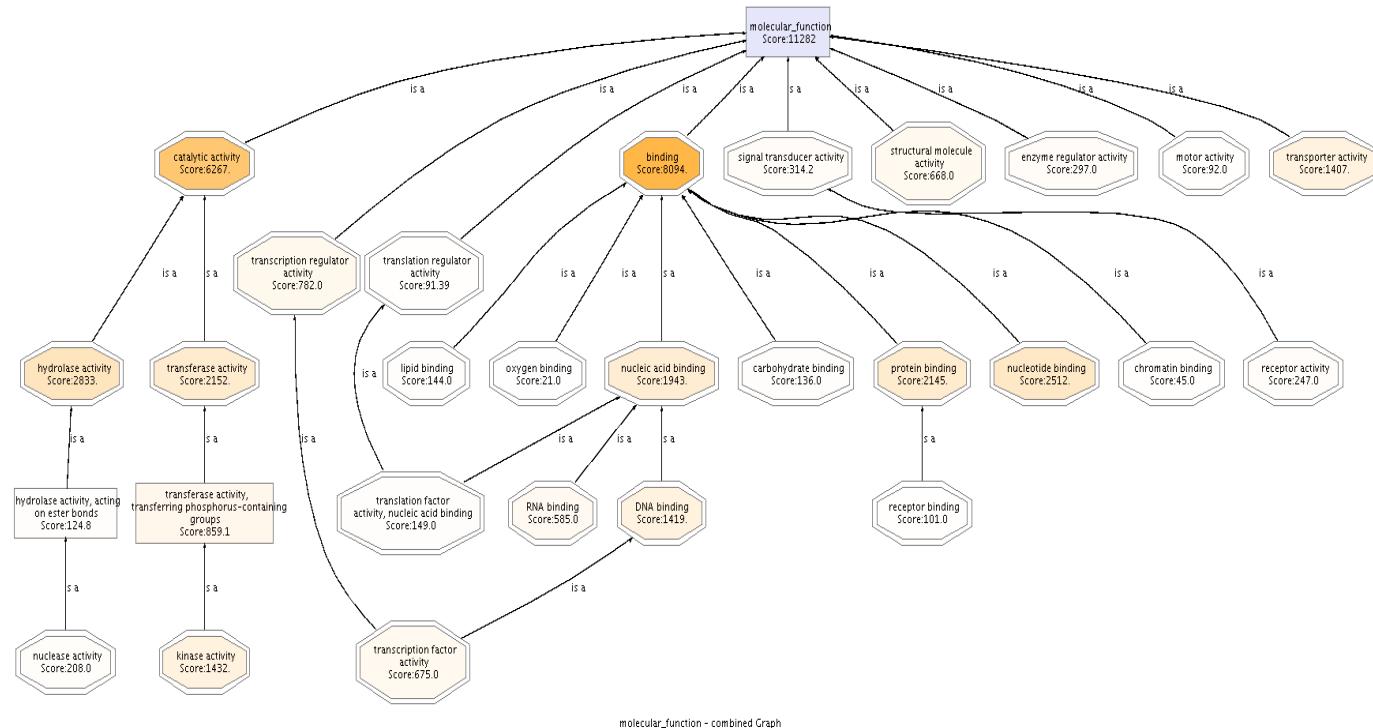
Pathways



The Gene Ontology

- Project developed by the **Gene Ontology Consortium**
- Provides a **controlled vocabulary** to describe gene and gene product attributes in **any organism**
- Includes both the development of the **Ontology** and the maintenance of a **Database** of annotations

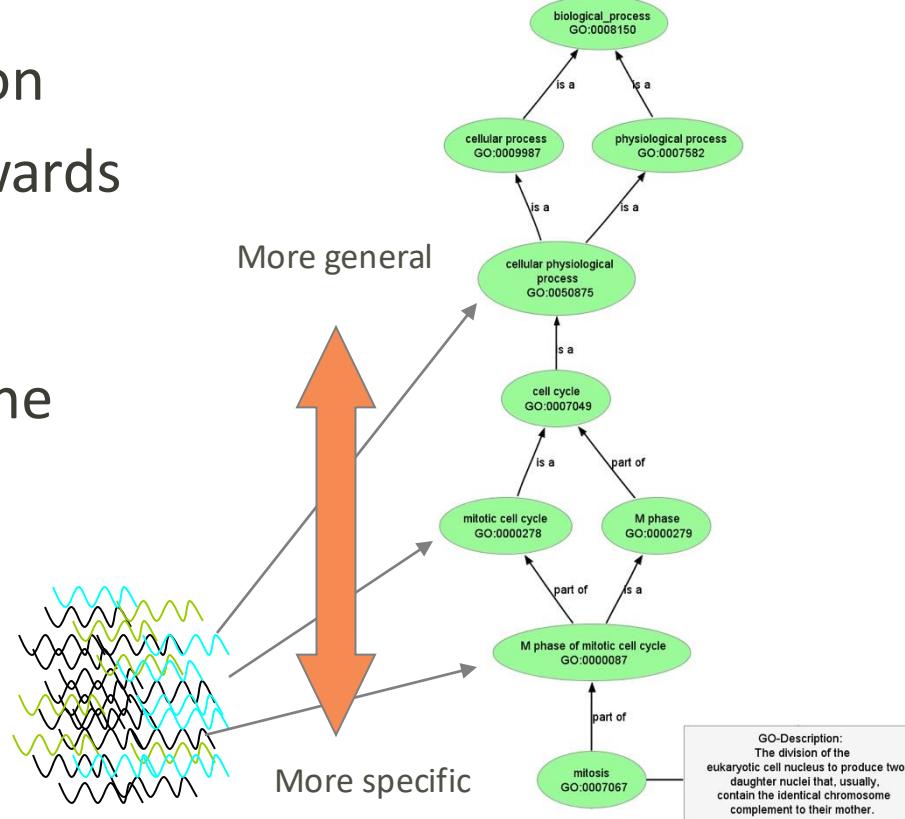
The GO has a DAG structure



molecular_function - combined Graph

What are the benefits of an ontology?

- **True path rule:** annotation are consistently true upwards in the gene ontology
- Annotation happens to the lowest possible level

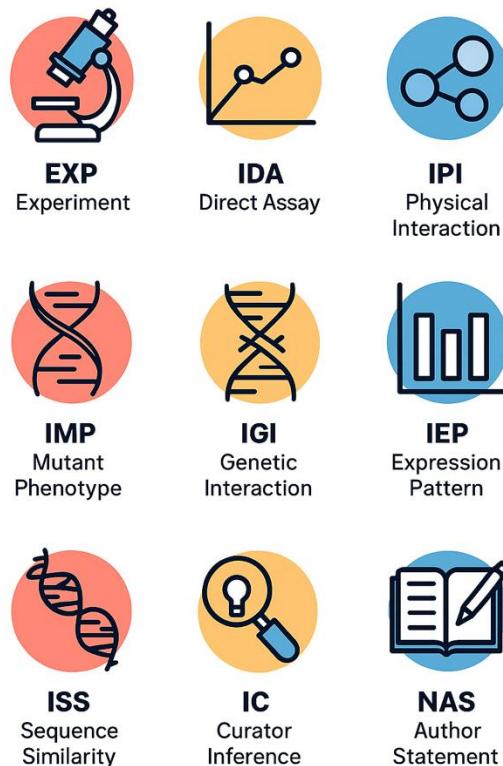


Gene Ontology branches

- **Biological Process (BP):** The series of events or molecular activities carried out by one or more gene products to achieve a specific biological goal:
 - Mitosis, stress response, cell signalling
- **Molecular Function (MF):** The elemental activities of a gene product at the molecular level, such as binding or catalysis:
 - Protein kinase, transcription factor, ABC transporter
- **Cellular Component (CC):** The specific location within a cell or its extracellular environment where a gene product is active
 - Nucleus, cytoplasm, membrane

Gene Ontology evidence codes

- Each annotation has an **evidence code**, indicating the source of the annotation
- Most of the GOA annotations come from **UniProt**
- Most of the annotations are **electronic annotations**
- There is a **collaborating institution** per organism to provide annotations



InterPro indicate functional motifs

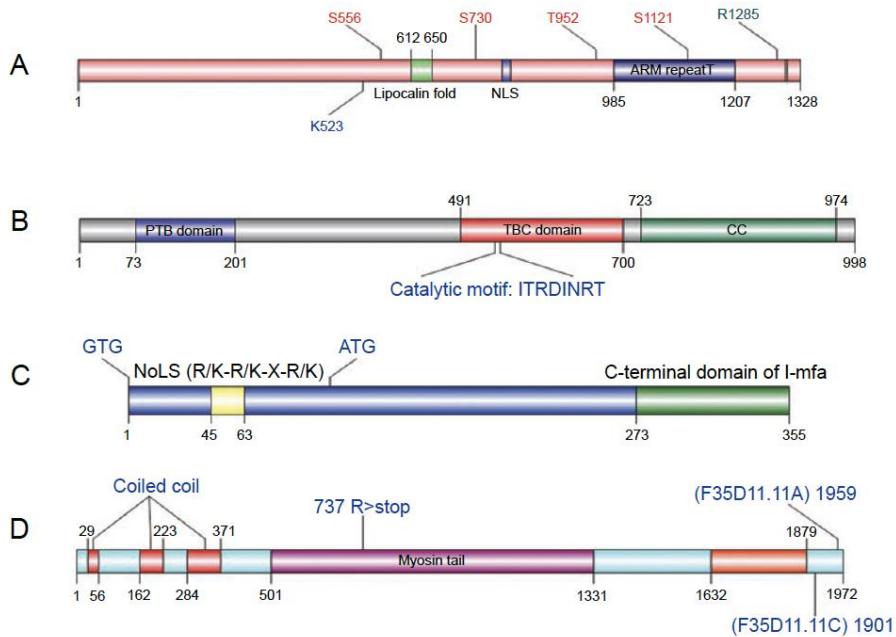
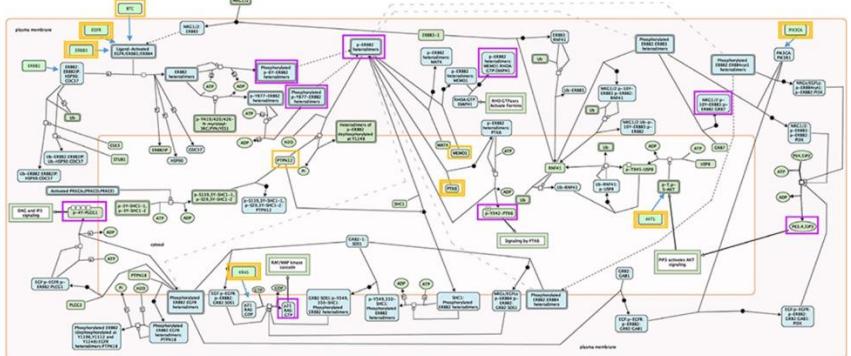
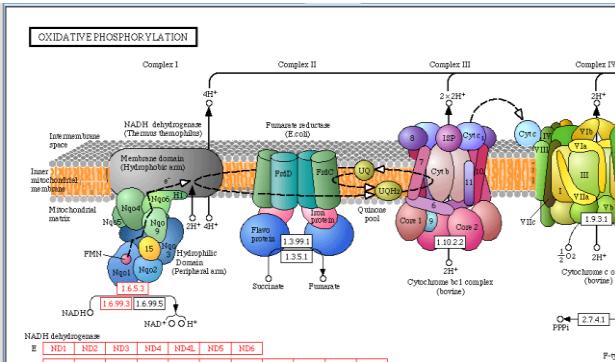


Figure 1 Domain structures for four proteins. **(A)** A monoubiquitinated and phosphorylated protein of FANCI [5]. **(B)** A GT-Pase activating protein (GAP) of TBC1D11/GAPCenA [6]. **(C)** A novel Hand1 interacting protein of HICp40 [7]. **(D)** A novel ciliary protein of F35D11.11B in *Caenorhabditis elegans* [8].

Pathways

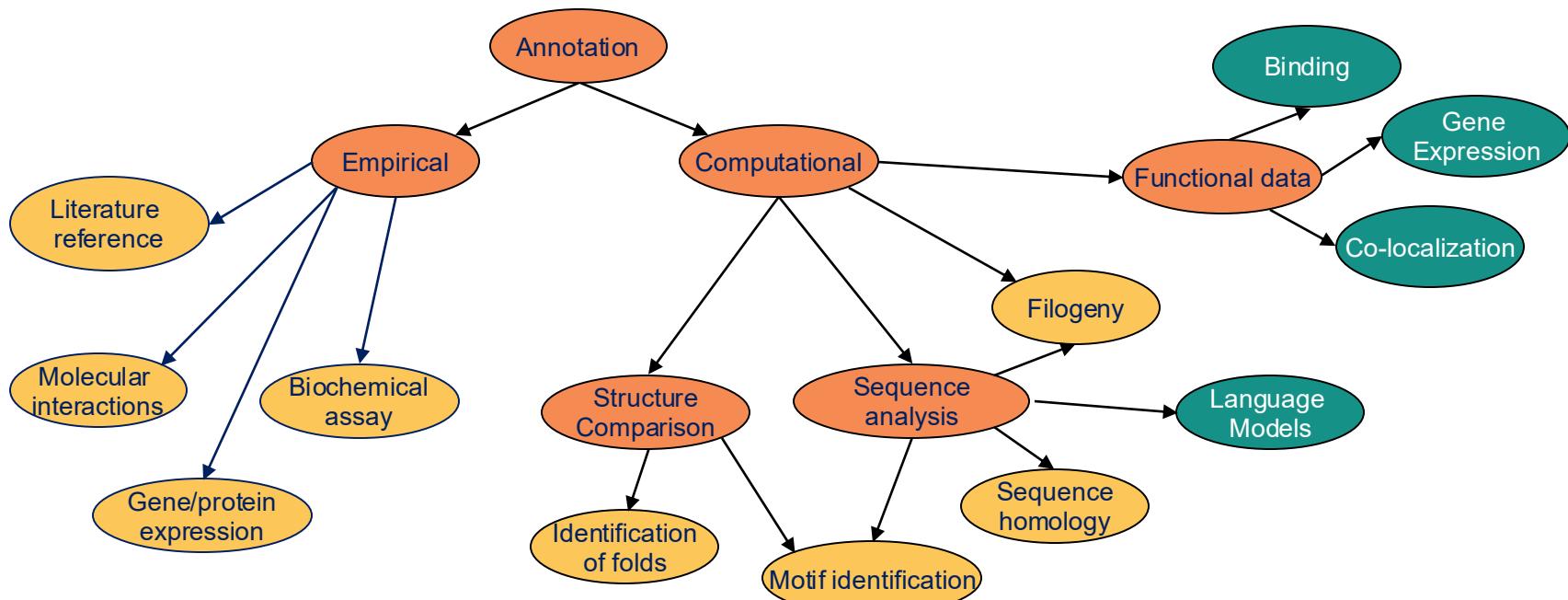


- Active development
- Complex
- Free
- Mostly human and mouse



- Standard
- Not completely free
- For all organisms

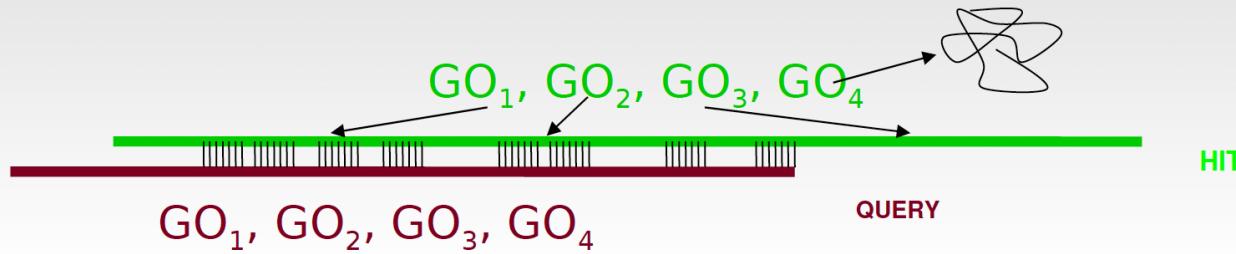
Functional assignment



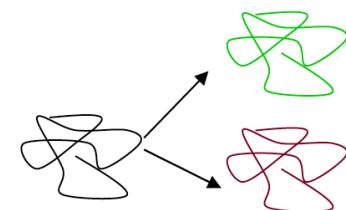
Automatic annotation

- GO annotations can be created by **comparision** to annotated sequences
- To achieve enough coverage, high-throughput, **automatic** annotation is required
- The most effective (also error prone) automatic annotation method is transfer from **sequence similarity**

Concerns in sequence similarity annotation



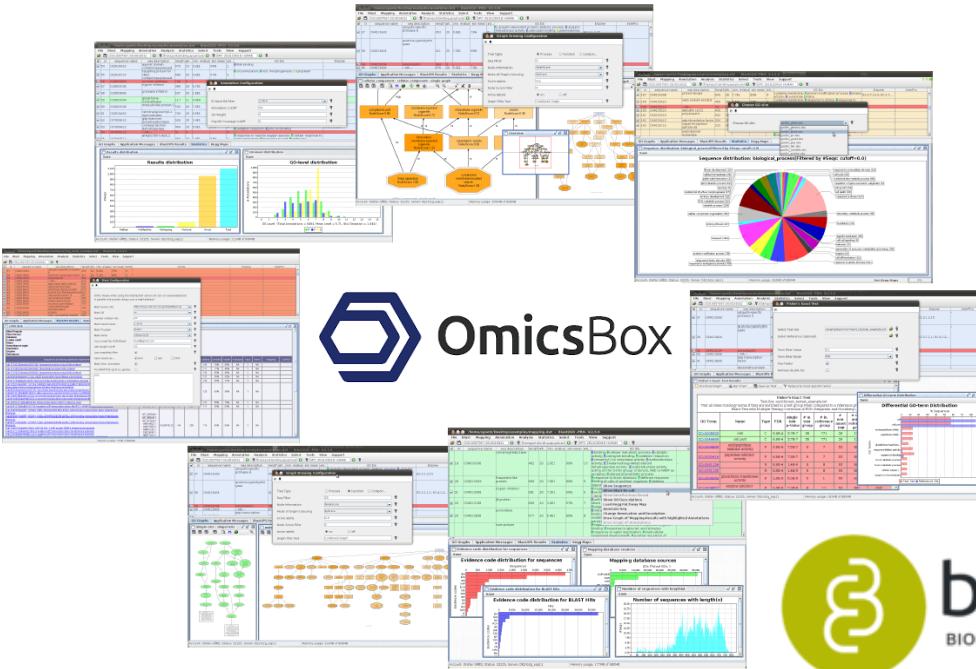
- ✓ Level of **homology** (~ from 40-60% is possible)
- ✓ The **overlap** query and hit sequences
- ✓ The domain or structure function association
- ✓ The **paralog** problem: genes with similar sequences might have different functional specifications
- ✓ The **evidence** for the original annotation
- ✓ **Balance** between quality and quantity: depends on the use





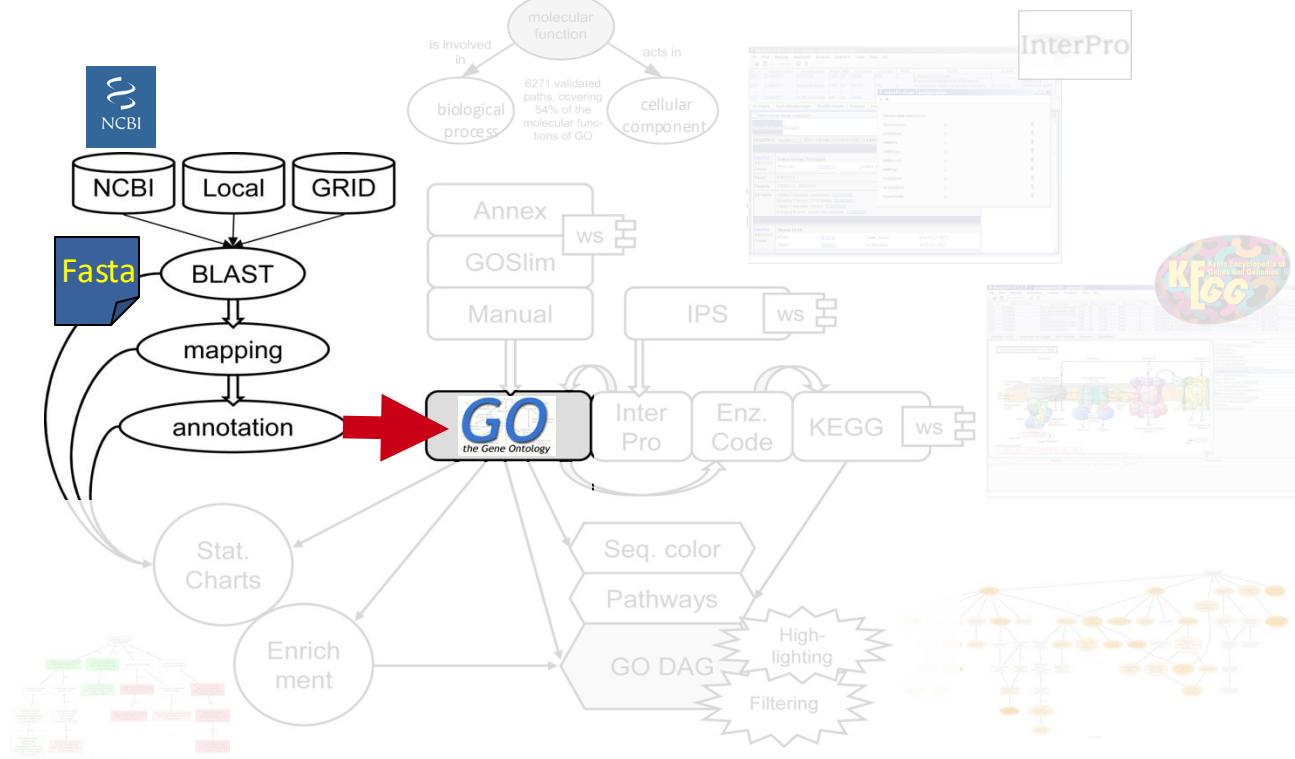
blast2go

de novo functional profiling



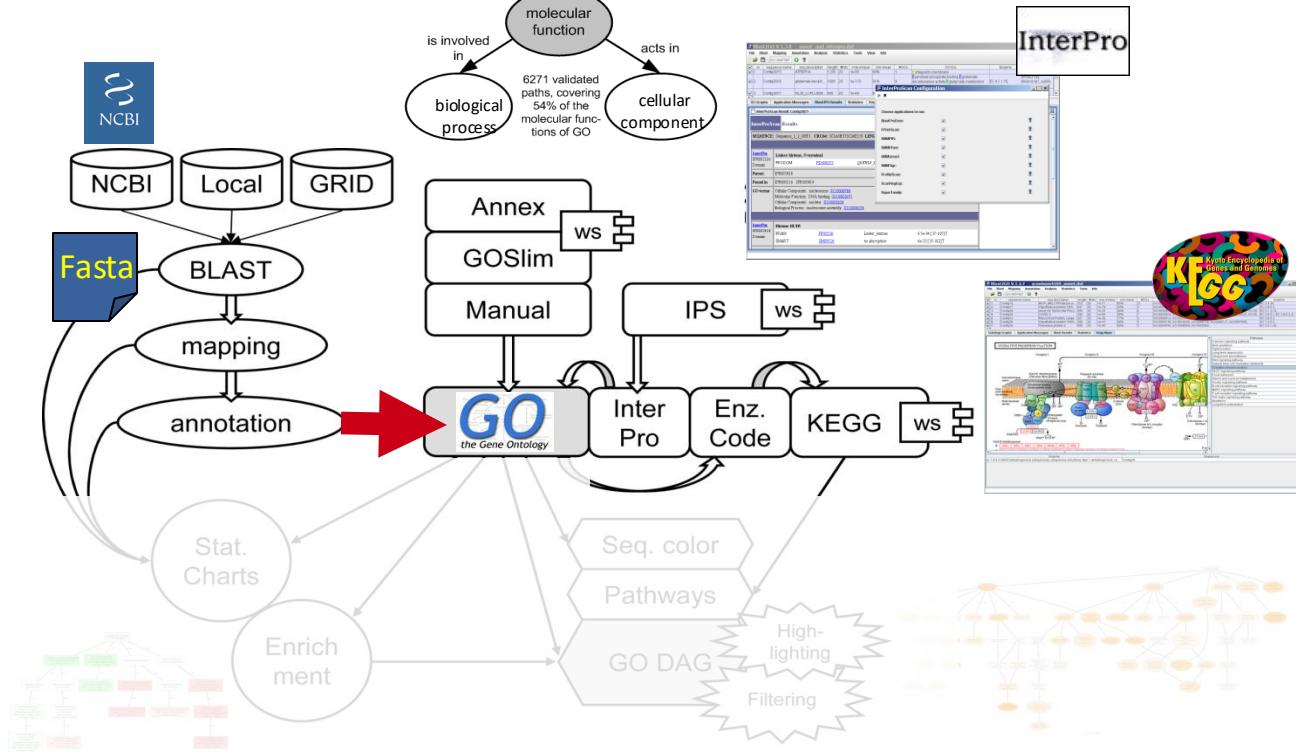


OmicsBox for de novo functional profiling



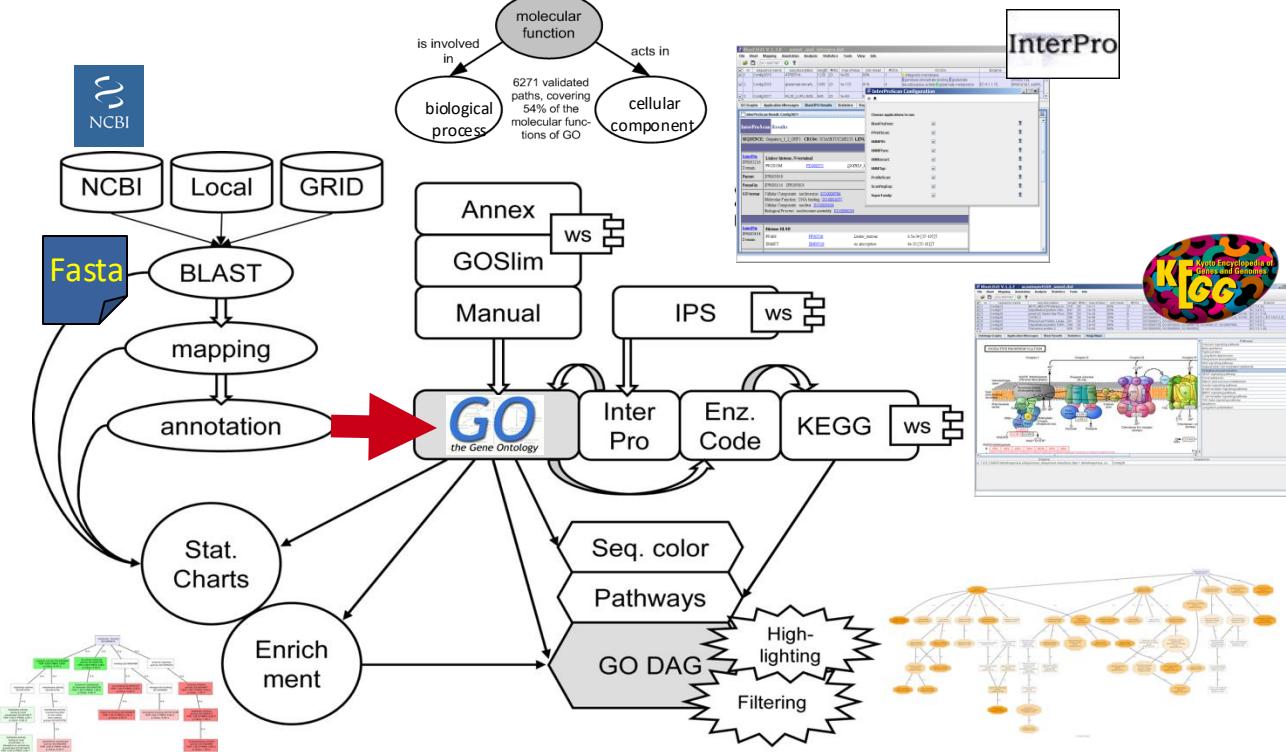


OmicsBox for de novo functional profiling





OmicsBox for de novo functional profiling



Goetz et al, NAR, 2008

Input

Result

Blast2GO V.1.2.7 - acaninum4169_annot.dat

File Blast Mapping Annotation Analysis Statistics Tools Info

GO:0007067

nr	sequence name	seq description	length	#hits	max eValue	sim mean	#GOs	GO IDs	Enzyme
✓ 1	Contig10	MCP1_MELCPProtease in...	318	20	1e-11	60%	21	GO:0005886, GO:0005576, GO:0042803, GO:0031012, GO:0019862, GO:00...	EC:3.4.24,
✓ 2	Contig17	Hypothetical protein CB...	647	20	1e-79	83%	3	GO:0006412, GO:0003735, GO:0005840,	EC:3.6.5.3,
✓ 3	Contig29	yeast Gic Seven-like Phos...	595	20	1e-24	94%	6	GO:0005977, GO:0051301, GO:0007049, GO:0005506, GO:0004721, GO:00...	EC:3.1.3.16,
✓ 4	Contig30	C47B2.3	396	20	1e-56	95%	7	GO:0005874, GO:0005198, GO:0003924, GO:0051258, GO:0005525, GO:00...	EC:3.6.5.1, EC:3.6.5.2, E...
✓ 5	Contig33	Ribosomal Protein, Large...	281	20	1e-26	77%	3	GO:0006412, GO:0003735, GO:0005840,	EC:3.6.5.3,
✓ 6	Contig36	Hypothetical protein Y94H...	399	20	1e-43	68%	5	GO:0005739, GO:0016020, GO:0006118, GO:0008137, GO:0007585,	EC:1.6.5.3,
✓ 7	Contig38	Trehalase protein 2	856	20	1e-40	66%	3	GO:0004555, GO:0046658, GO:0005993,	EC:3.2.1.28,

Ontology Graphs Application Messages Blast Results Statistics Kegg Maps

OXIDATIVE PHOSPHORYLATION

Pathways

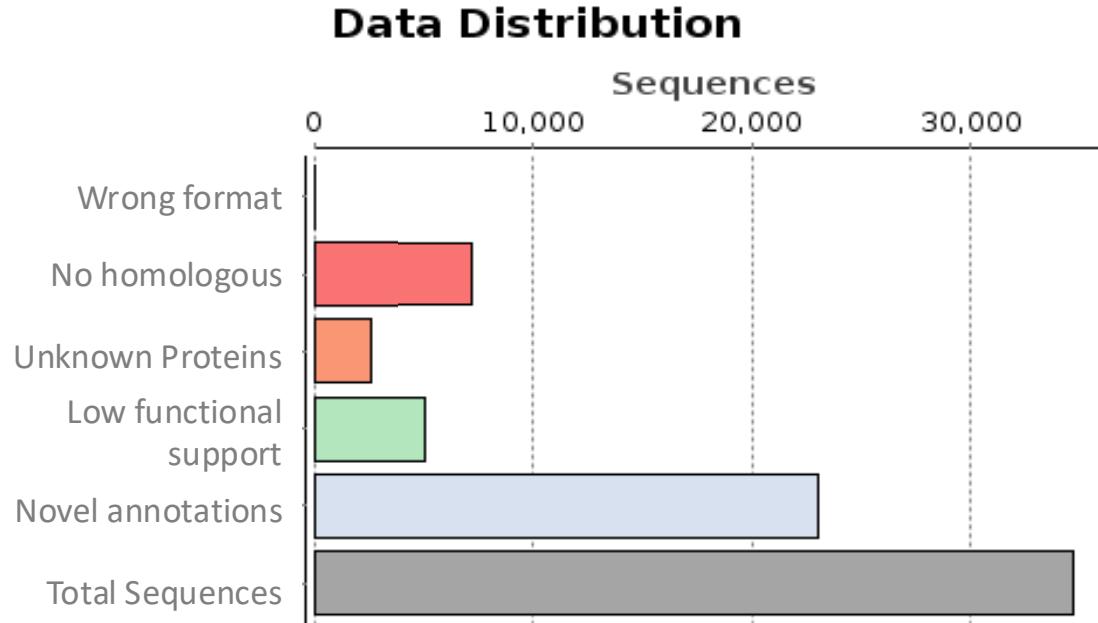
- Calcium signaling pathway
- Avon guidance
- Tight junction
- Long-term depression
- Ubiquinone biosynthesis
- Wnt signaling pathway
- Natural killer cell mediated cytotoxicity
- Oxidative phosphorylation
- VEGF signaling pathway
- Focal adhesion
- Starch and sucrose metabolism
- Insulin signaling pathway
- B cell receptor signaling pathway
- MAPK signaling pathway
- T cell receptor signaling pathway
- TOF-beta signaling pathway
- Apoptosis
- Long-term potentiation

Enzyme

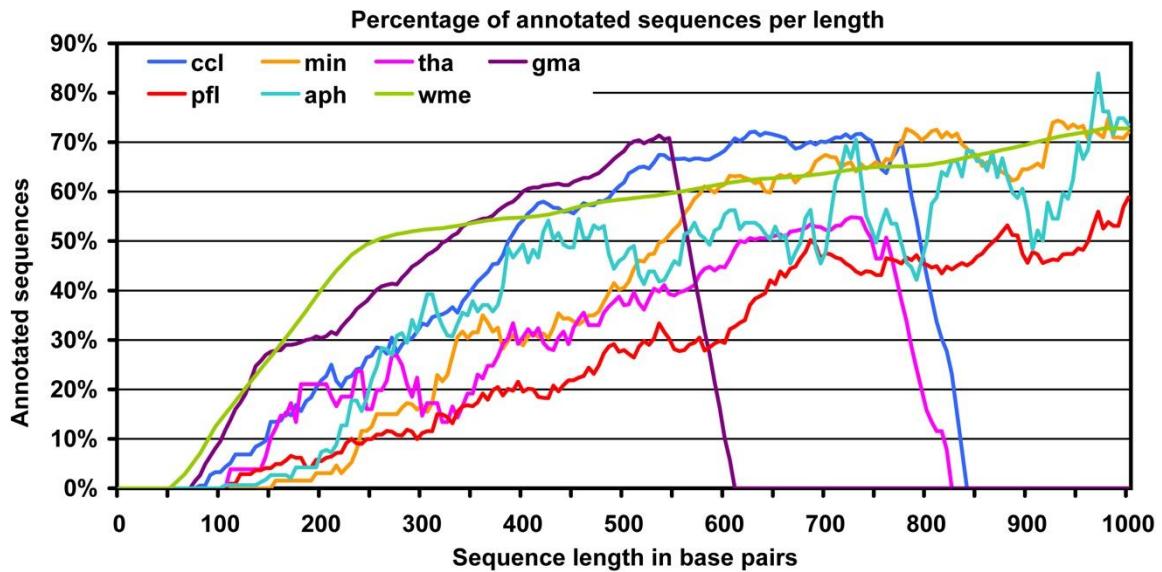
ec:1.6.5.3 NADH dehydrogenase (ubiquinone); ubiquinone reductase; type 1 dehydrogenase; co... Contig36

Sequences

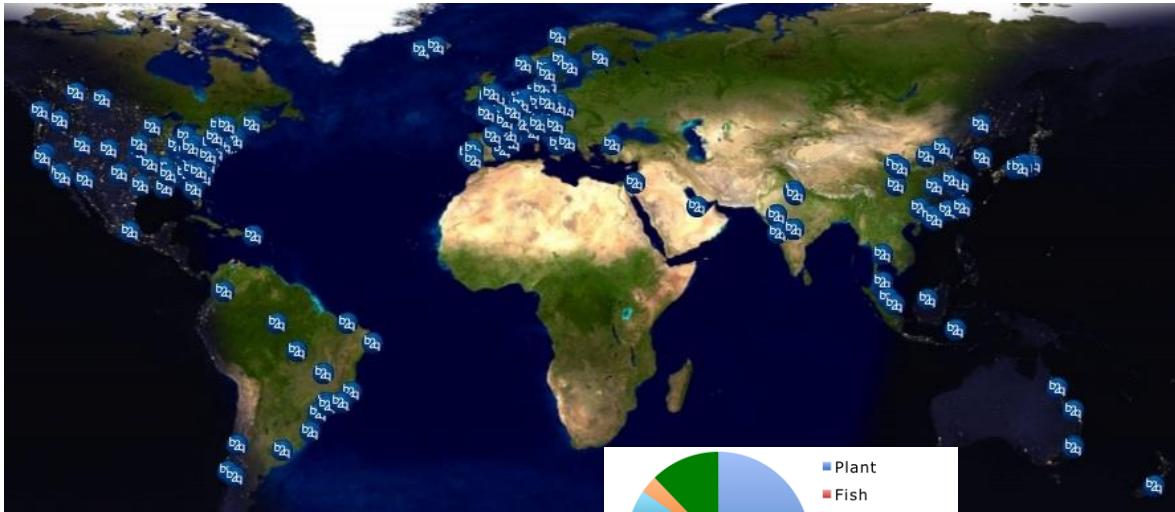
Typical Blast2GO annotation result



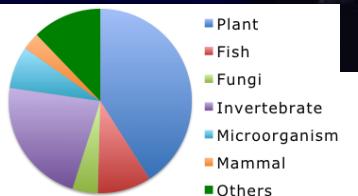
Annotation success depends on transcript sequence length

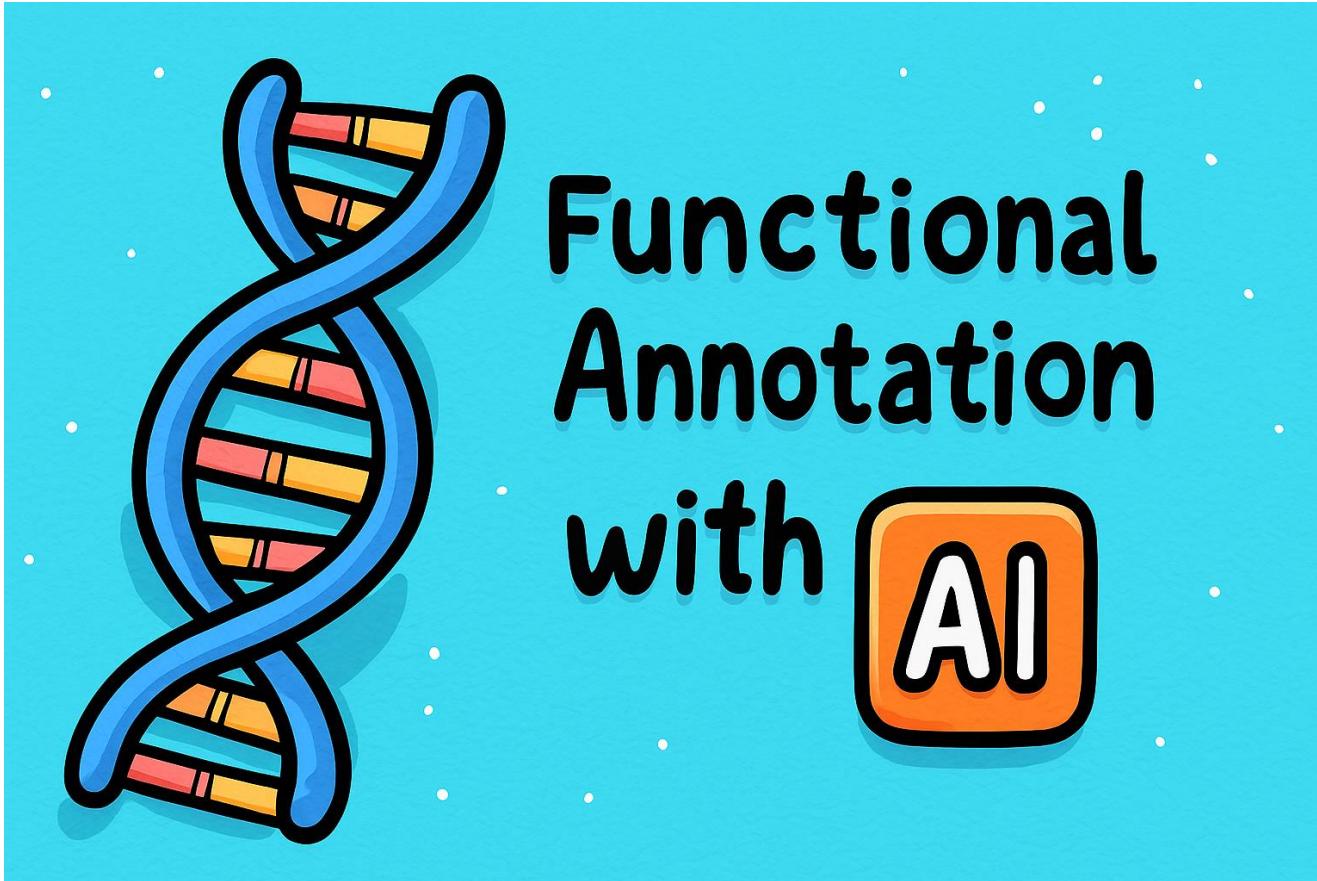


Blast2GO Impact



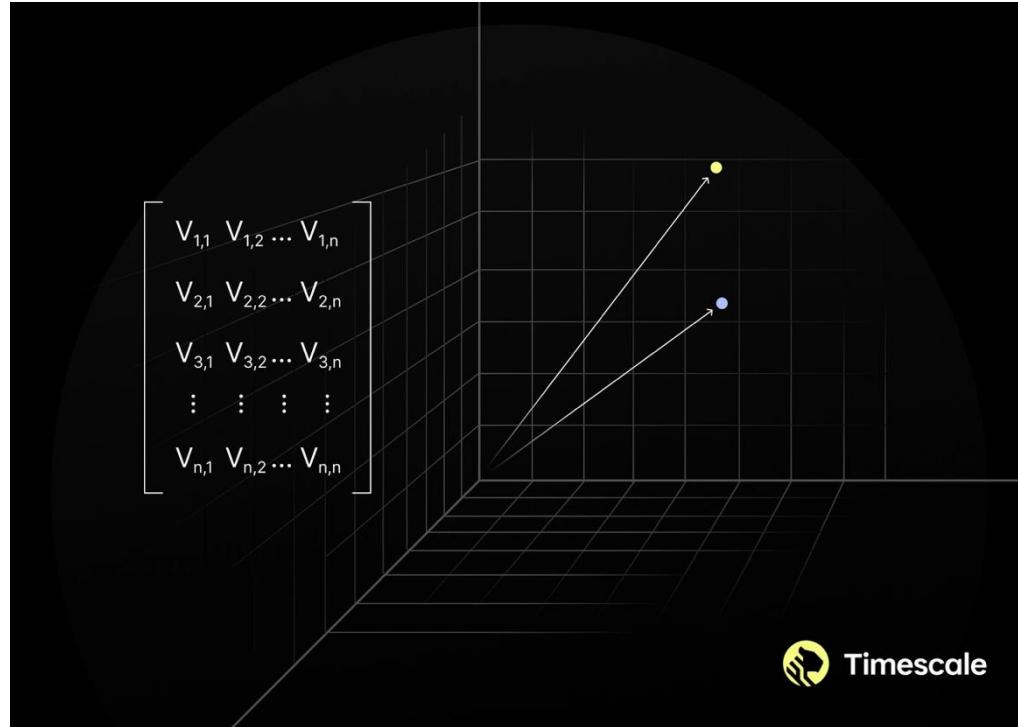
10,000 different users/month
>15,000 scientific citations





Functional annotation using Large Language models

Embeddings



What is a protein embedding?

generate feature-rich numerical representations of protein sequences.

Id [PK] integer	sequence character varying
1	1 MNAPLGGIWLWLPLLLTWTPEVNSSWWYMRATGGSSRVMCDNVPGVLSSQRQLCHRHPDVMAISQGVAEWTAEQCQHQFRQHRWCNTLDRDHSLFGRVLLRSSRESAFVYAISSAGVVFAITRCSQGEVKSCSDPKKMGSAKDSKGIFDWGGCSNDIDYGKIFARAFVDAKERKGKD
2	2 MKKSIGILSPVGALGMAGSAMSKFFLVALIAIFFSFAQVVIYEANSWWSLGMNNPVQMSEYYIAGAQPLCSQLAGLSOGKKLCHLYQDHMQYIGEAKTGKECQYOFRRHWNCSVNTSFGVRVMQGSRETAFTYAVSAAGVVNAMSRACREGELSTCGCSRAARPDKLPRDWLGCG
3	3 MAWEARREPPGRRAAFTREVMLLCLGVPTRGRPYNNVDTESALLYQGPHTFLFGYSVVLLSHSGNWLWVGAAPTANWLNASVINPQNGRIGKQGOTCEPCGKTCLEERDNOVGLVTLSRQPGENGSTVCGHWRKNIFYIKENKLPTGCGYGVPPDLRTELSRIAPC
4	4 MQMSPALTCLVGLALVFGEGSAVHPSYVAHLASFQDFGRVFRFQVQAQSKDRVNPSYGVASVAMLQMLTTGETQKQIAAMGFKIDDKGMAPRALHLYKELMGWPWNKEDISTDAIFVLDKLKLQVGFMPHFRFLRSTVQDFSEVERARFIINDWVKTHTKGMISNLLGKGAQDQL
5	5 MAEVPELASEMAYMSGNEDDLLFADGPQKMKCSQFDLQLPDLQGGIQLRISDHVSKFQRAQASVAMDKLRLMVLPCPQTQFQNEDLSTFFFPIFEDTWDNEAVYHDAPVRSNLTRLSDSQSKSLVMSGPYELKALHQQGDMEQQVFMSMFVQGEESNDIKPVALGLKEKNI
6	6 MELRPWLWVAAATGTVLALLAADDQKVFTNTWAVRPGGPAVANSVARKHGFLNLUQIFGDYVWHFHWRGVTKRSLSLPHRPHRSRLQRFQVQWLEQVAKRRTKRDVYQEPDPKFPQWVYLSGVQFQDNLNAKAQWAGQYGHGIVVSIIDGIEKHNDLAGNYDPGASFDVNDQPF
7	7 MTSTGQDSTTQRSSRNQNPQSPPQDSSVTSKRNIKKGAVPRSPINLAEVKKKGKMKLKGQAMEEDLIVGLQGMDLNLEAEALAGTGLVILDEQNEFLCWLDSSPFGEPERLHAIKEQLIQEGGLDRCSVFSQARFAEKEELMVLHVSLEIDLMTTQYMNEGELRVLADTYDTSVYLPNSYSCA
8	8 MTELAKGPRAPHVAGGPPSVEPGSPCLLCPRAAGPFFGQSQTSLPEVSAIPISLDLGLLFRPCQCGQDPSDEKTQDQQSLSDVEGAYSRAEATRGAGGSSSSPPEKDSGLDSDLTLLAPSGPQGSQSPSPACEVTSSWCLFGPelpDPPAATQVRVSLSPMSRGCKVGDSSGTAHHK
9	9 MMSASRLAQTJIPAMCSPCRVPRGPNPQCNMAMADELSEKQVYDQVDAHTKEIDVLNRDPKHNNDVPLRQKIDFEDVIAEPEGTHSFQDGIPALMIALIWFQDLSLHFVWVPCIKSFLIEIQCSIRVSYVHVTCDPLFEAVKGFSNVRINLQKEI
10	10 MSGKQYVVDHLYTVPIREQNLYKPKNAMADELSEKQVYDQVDAHTKEIDVLNRDPKHNNDVPLRQKIDFEDVIAEPEGTHSFQDGIPALMIALIWFQDLSLHFVWVPCIKSFLIEIQCSIRVSYVHVTCDPLFEAVKGFSNVRINLQKEI
11	11 MPSPGLRLPLLPLLWLWLTGPRAAGLSTCKTDIMELVLRKRKMLAIRGQSLRASPSSQGEVPPGPELPEAVLWLNSTRDVEGAEEPEPEADYDAKEVTRVLMVETHNEIDFKQSTHTVSYIMFTNLSELEAVEPVLLSRAELRLRLLKLVQEHVLYQKYSNNNSWRYSLNRLAI
12	12 MAFFPRRRLRPLPGLLPLLSSGLLPLCRAFNLDVDSPEAECSPGEYSYFGFAVDFVFPSSAQRMFLVYQVAPKANTTOPGIVEQGVLCWDWSSTRRCQPIFEDATGNDRYAKDQDPLFKEKSHQWFGASVRSKQDQKLLACPLYHWRTEMKQEREVYGTCLQDTKTVYEAPCRSDQDIDADQGF
13	13 MDVFMKGLSKEAKGVAAAEEKTQKGVAEAAKGTKEVLVYGSKTKTEGVVHGATVAEKTKEQVNTVNGGAVVTVGTVATAQVKTVEGASIAAATGVKKDQLKGNEEAPQGEILEDMVDPDNEAYEMPSEEQYQDYDEPEA
14	14 MNLQPIFWIGLSSVCVFAQTDENRCLKANAKSCGEICQAGPNCGWCNTSTFLQEGMPTSAACDDLEALKKGCPDIDNPRGSKDIIKKNNKVNRSKGTAEKLKPEDITIQOPQQLVRLRSGEPOTFTLKFKAEDYIDYLMLDSYSMKDDLENVKSGLGTDLMNEMRRTSDFRIGFGS
15	15 MRAPRPRPLRWTAVSLALGAGVQHGPNCNCTTRGVSSCQCLAVSPACWCSACDPLGSPRCLQENLKNCAPSEFVSEARPELRLPDRLSGDSGGSDFQDNLSKDFNSKIPQVNRQEDVYDYLWMSYSMKDDLWQSLQNLGTLKATQMRKLTSNLRFGAFVDI
16	16 MEFSSP3REECPKPLSRVSIIMAGSLTGLLQWASGARCPKSPFGYSSVVCVNQNTYCDSDPFTPFLGTFSYESTRSGRMELSPQGKMPQIYAHNTQGKQFQVKGFGGAMTDAALNLLSPANLQNLKSYSEEGYNYIIRVPMASCDFSIRTYTADTPDFQLHNF
17	17 MMKTFLLFVGLLWTSQGQDQVTSNLEQEMGSQKVYKEIQAQNQVQKQFQVKTQKLTNEERKTLTNSLNLEAKDNEALNTERSETKQHLLPEQKQFQVKGFGGAMTDAALNLLSPANLQNLKSYSEEGYNYIIRVPMASCDFSIRTYTADTPDFQLHNF
18	18 MEEFQSDPSVEPLSPOETSDWLKLPENNVLPSPQGQAMDLMLSPDDEIOWPFTEDPGPDEAPRMPAEVAPAPAAPTAAAPAPSPWPLSSSPVSPQKTYQSGYFLRGLFHSGTAKSVTCTYSPLANKMFCQALKTQPCVQLWLVDSTPPGTRVRAIMYKQSQHMTEVRRCPHHEI
19	19 MSLWQPLVLVLLVLCGCCAAPPQRQSTVLPGDRLNTLDRQLAEEYLYRGRYGRVRAEMGESKSLGPALLLQQQLSLSPETGEDLSATLKMARTPRCQVGPDLRGFQTFEGDLKWHHHNITYWIQNSEDLPRAVDDAARAFALWSAVTPLTFRVSYRSDADIVIQFGVAEHGDGVYPFDGKDG
20	20 MQYLNKEDCNAAMFAKMRSSKKTEVNLEAPEPQGEVIFYLSDREPPLRLGSGEYTAELCIRAAQACRISPLCHNLFALYDENTKLWYAPNRTITVDDKMSLRLHYMRMRFYTNWHGTNDNEQSVRHSPPKKQNGYKEKKIYPDATLDDASSLEYFAQGQYDLVKCLAPRDPKTEQOGHDI
21	21 MSARGEAGQPTSAQGQPAAPAPKGRGRGRPKQQQEPTEPSPKPRPGRPKGSKNSKSPCAAQKKAEEQKPRPGRPKWPPQVVKKPAQEETEETSSQESAED
22	22 MPKTSIVRTTMDAELAEIOPQNTTQKQLDFQVKTIGLREVWFFLQLQYDQTKRQFSTWLNLKVKVTAQDVRKESPLLFLKFRAKFQYPEDVSEELIOPQNTDQRLFQVKEGILNDDYCPPEAVTLLASAVYQSKYQDFNKEVHSGYLAGDKLPPRVLQHKLNKDQWEERIYQVWEEHRGMLRED
23	23 MAAIVGRVSGFSSPPGLSSTYTPGGLNEIAASGNAAAQDDEDQGDNQWLSCILSEEVTRSRSKLPLAGKVNLLQEDGAGKTSRQKIQGIEEKKGRGLYLNVNHEDRDDQTRCNWILDGDLYHKGLLFLSDAVSKLTDLVMVLDMSKPWTALDSLQKWWASVREHV/DKLKIPPEEMKI
24	24 MAAIGRGRSLSKNVLVRGRNDSGEENVPDLTRPDSNRLREILONVARLQGVSNMRKLGHLLNNTFLKLCQDIGHSEEKLGFHYEDIIICRLALLNEAKERAQGLRALRLYQDSSILQVKLKLVDFLQYIARCIQDIOQSNEVERTQARLVRKMTVITASLFLSSVTSNLIAVGNQDQLERDMRVMRACIAI
25	25 MSTDPPLGLGTPRGPSPGPSPGAMLGPSPGPSPGSAHSMGGPSPGPPSAGHPIPTQGPQGPGYQDNMHQMKPMESMHKGMSDPPRNQYQKMGGMGRSSGHAGMGPSPMDQHQSQGPSPGLGSEHASSPVASPGSSGPQMSGGPAGPLDGPQALQGQNRGPTFN
26	26 MEGAGGANDKKKISSERRKEKSDDAARSRSRKESEFVYELAHQLPLPHNVSSHLDKASVMLRITISYLRVRLKLLDQGALDQDIEDDMQAKMCNCFVLLKALDGFVMLTDDGDMYIISDNVNKYMLGQTFELGHSVDFTHPCDHEEMRMLTHRNGLKKGEQNTQRSFFLRMKCTLSRGRTTM
27	27 MTKSLAVALLAALFUSALCEGAVPLRSPAKELRQCICKTYSKPFHKFKEKLVRVIESGPHCANTIEVLLKLSGRELCPDKPENVRVQVKEFKLRAENS
28	28 MRQSLFLTSVVPVLFVAPRPPDPDFGPHQRLREKLQDLSLSDYDILSLSNQQHSVRLKQDTSRHTVETLTSFALKRHFQYLTSSTERFSQNFQVWVVGKNESEYTVKWDQDFFTGHVYEPDSRVLAHIRDDVIIINTDGAEYNEIPWLRFVNDTLDKRMVLVYKSEDIKNSVRLQSPKVCGLY
29	29 MSTVHEILKLSLEGDHSTPPSAYGSVKAYNTFDAERDALNIEATAKTKVGDDEVITVNLNTSRNSAQDIAFAYQRTKKEKALASALSGHLETVLGLLTPAQYDASELKASMKGKLTDEDSLIEIICSRNTQELQEIQRVYKEMYKTDLEKDIISDTSQDFRKLMLVALAKGRRaedGsvIDYELII

What is a transformer?

"The cat sat on the mat because **it** was tired."

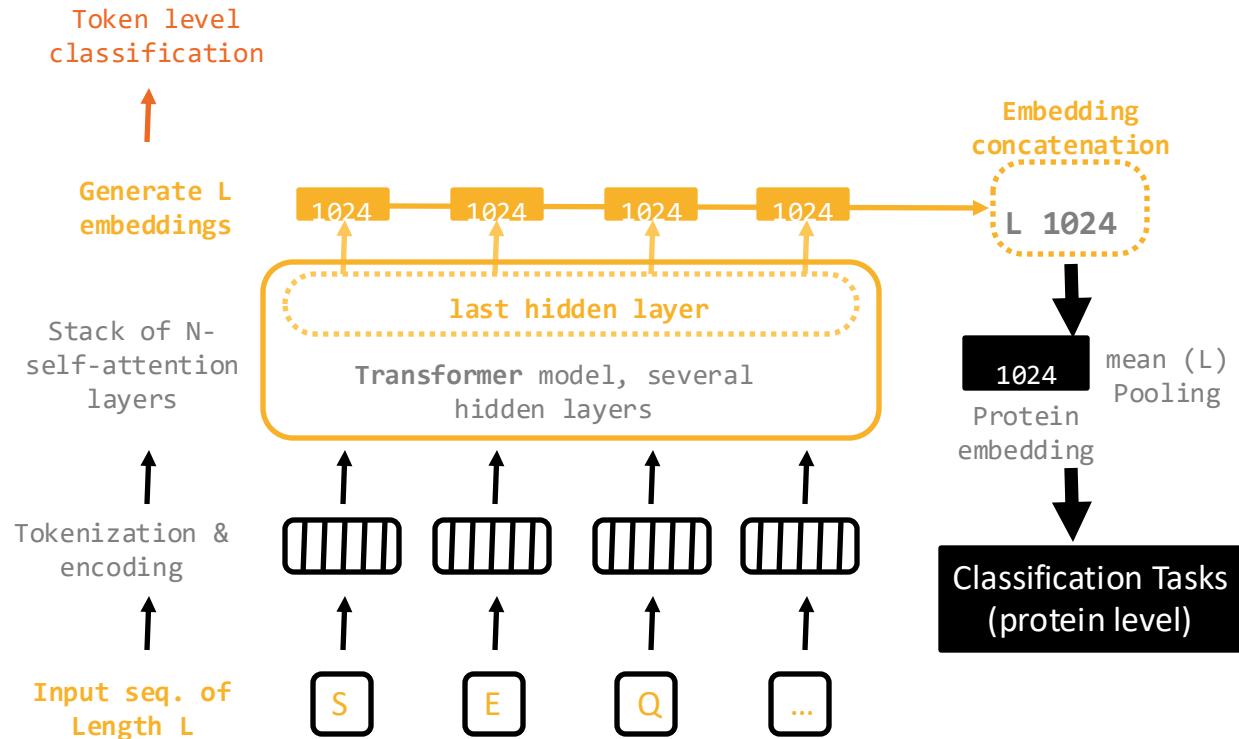
WHAT IS "**IT**"?

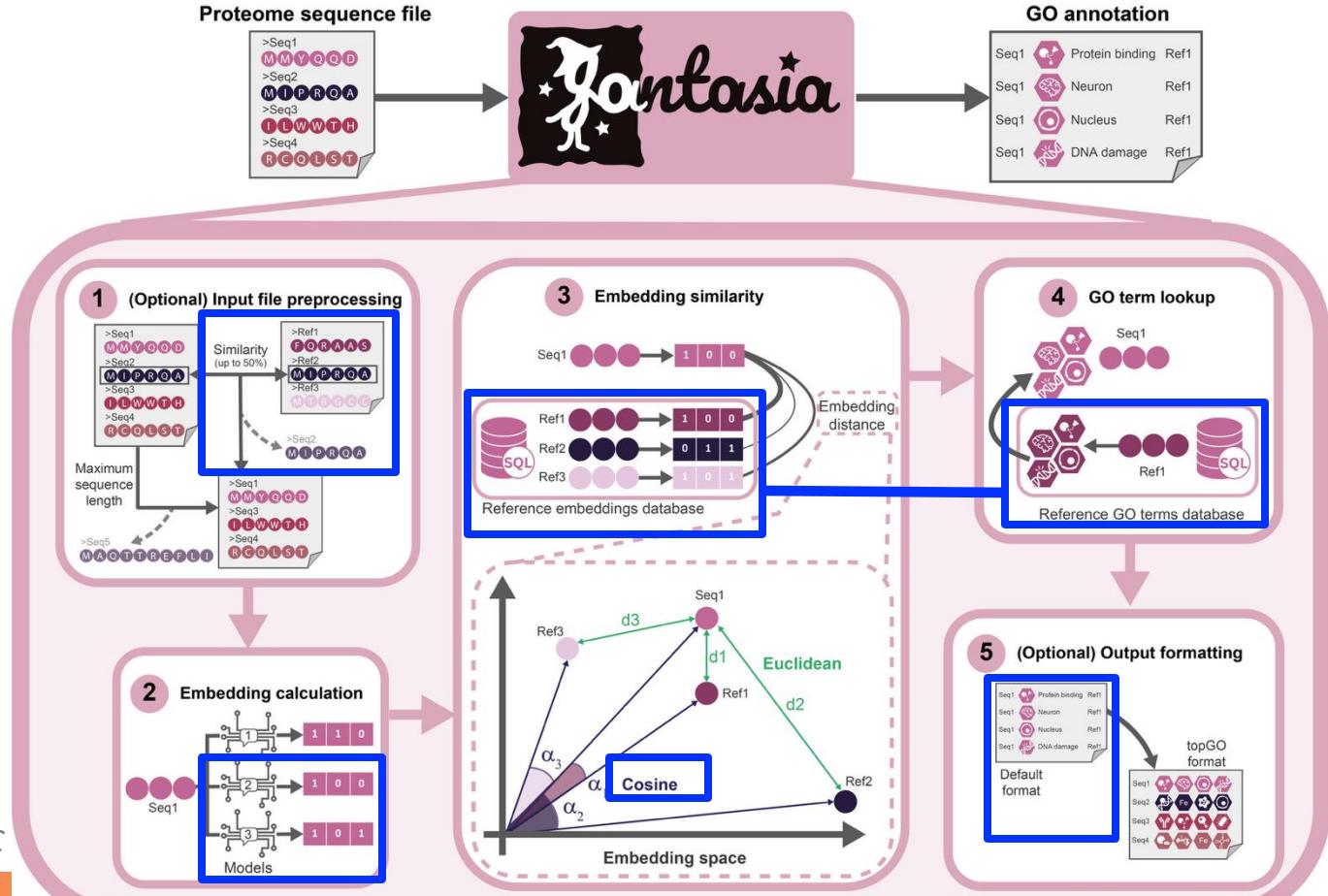
Transformers are AI models designed to understand language using context using a special mechanism called attention.

Attention allows:

- look at all the words at once (not just one at a time)
- decide which words are more important for understanding

The ProtT5-XL





Evaluating function prediction: The CAFA challenge



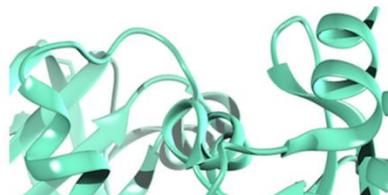
CRITICAL ASSESSMENT OF FUNCTIONAL ANNOTATION · RESEARCH CODE COMPETITION · 2 YEARS AGO

Late Submission

...

CAFA 5 Protein Function Prediction

Predict the biological function of a protein



Consortium *Genome Biology* (2019) 20:244
<https://doi.org/10.1186/s13059-019-1835-8>

Genome Biology

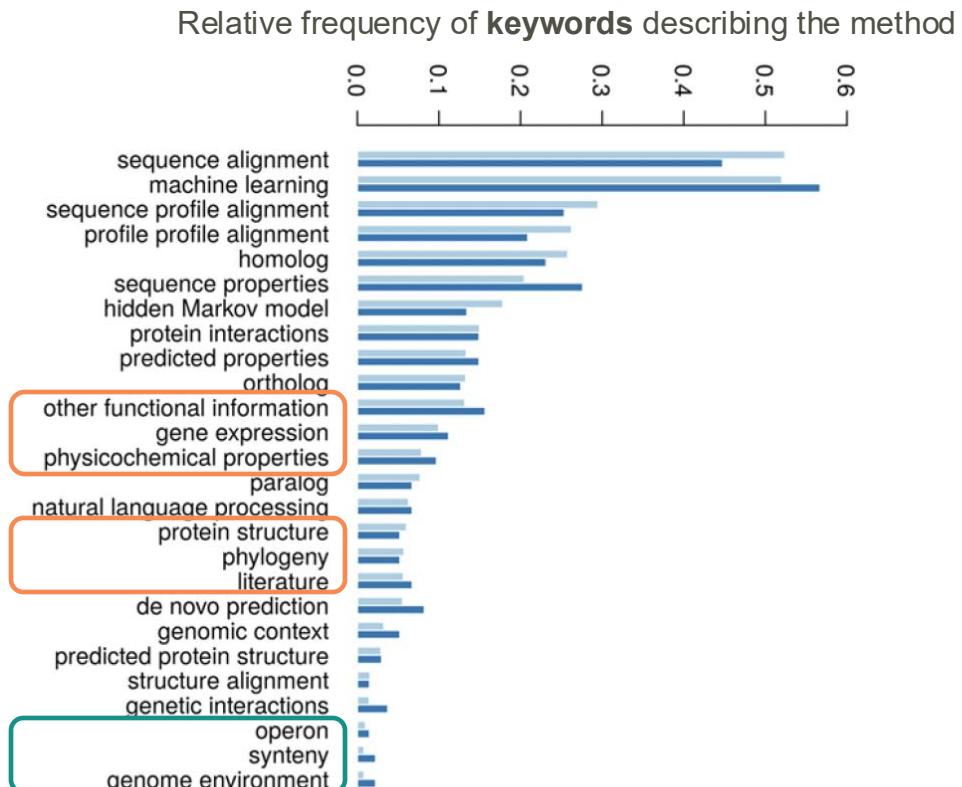
RESEARCH

Open Access

The CAFA challenge reports improved protein function prediction and new functional annotations for hundreds of genes through experimental screens



Type of information used by function prediction methods



CAFA challenge main results

Major improvement from CAFA1 → CAFA2; smaller but consistent gains in CAFA3.

Top performer CAFA3 : GOLabeler – ensemble classifier excelling in Molecular Function Ontology.

Integration of co-expression and network data (NetGO) enhanced performance.

Large-language models are the new best performers



Non-coding RNA Research

Available online 29 September 2025

In Press, Journal Pre-proof [?](#) [What's this?](#)



F
N
...

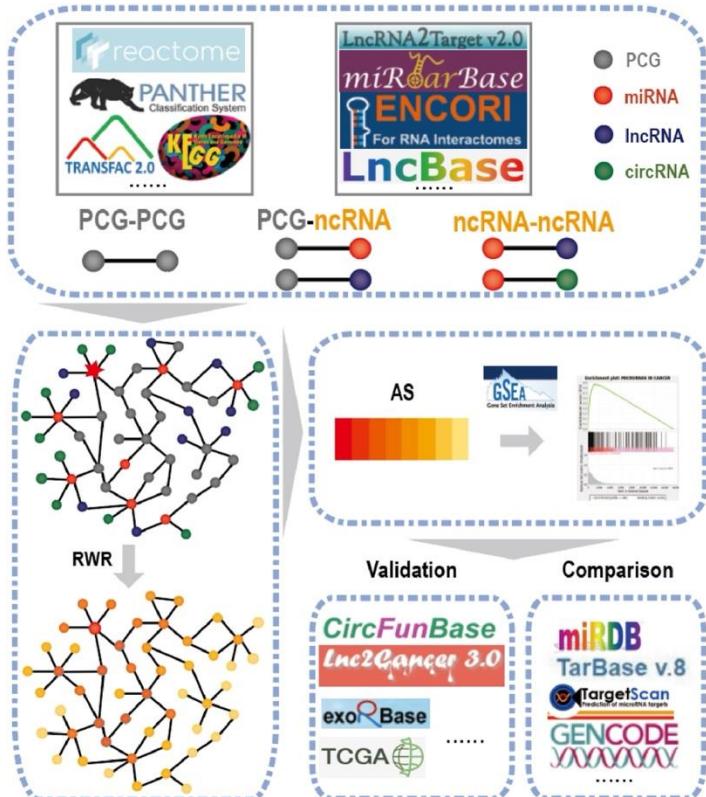
ncFN: a comprehensive non-coding RNA function annotation framework based on a global and heterogeneous biomolecular network

Min Long^{1†}, Haizhou Liu^{2 3†}, Mengqin Yuan^{1†}, Xu Zhou¹, Tao Zhang¹, Quan Wang¹ ,
Wei Jiang^{2 3 1}

Show more

Annotation of non coding RNAs

- Databases for ncRNAs

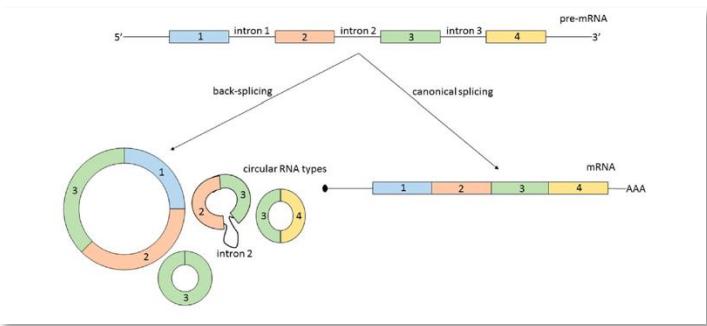


- Annotation by co-expression

Annotation of circular RNAs

Circular RNAs in Cancer – Lessons Learned From microRNAs

Mihnea Dragomir^{1,2,3} and George A. Calin^{1,4*}

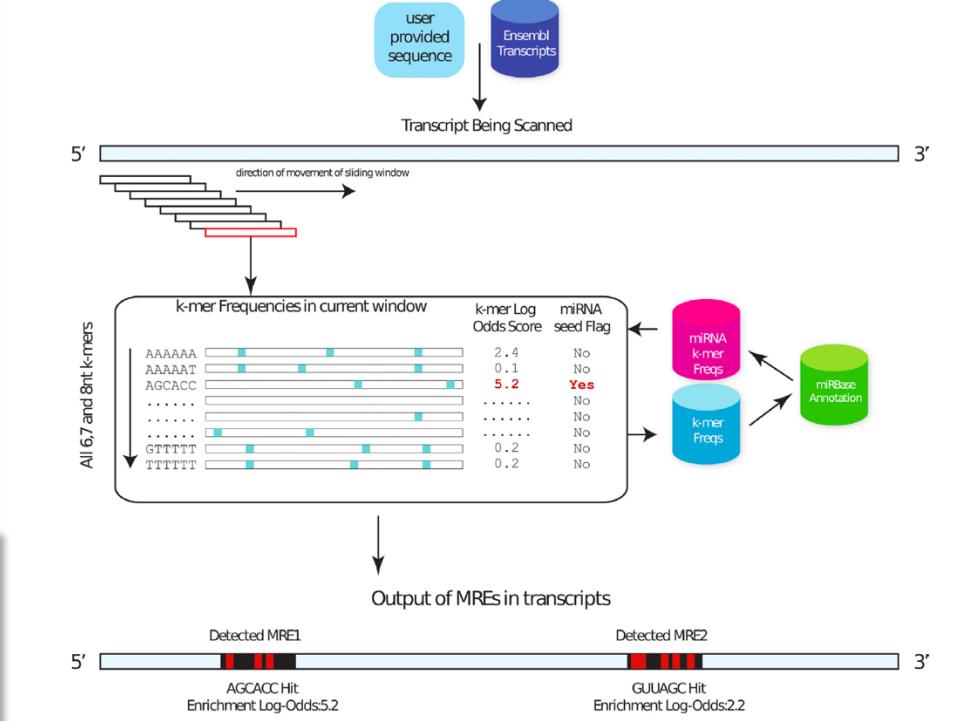


Nucleic Acids Research Advance Access published May 19, 2016

Nucleic Acids Research, 2016 **I**
doi: 10.1093/nar/gkw443

spongeScan: A web for detecting microRNA binding elements in lncRNA sequences

Pedro Furió-Tarí¹, Sonia Tarazona^{1,2}, Toni Gabaldón^{3,4,5}, Anton J. Enright^{6,*} and Ana Conesa^{1,7}

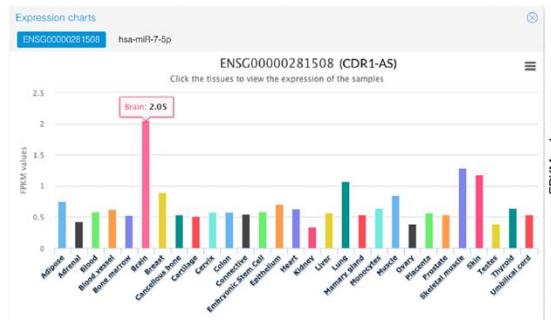


SpongeScan application

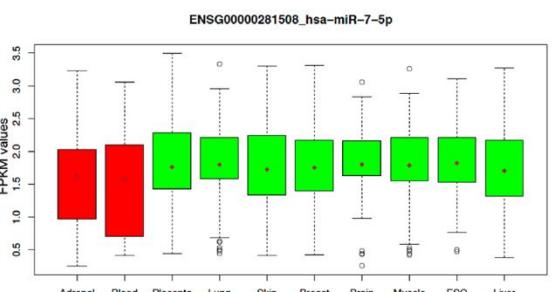
A

Gene name	Transcript name	Kmer	miRNAs	LOD Score	Complexity...	Standar...	Graph	Select filter fields...	Enter text	Add Filter
CDR1-AS	CDR1-AS-001	TCTTC	hsa-miR-7-5p	2.590	3.5	1.395				
Binding sites:										
Exons (+):										
FLJ16779	FLJ16779-001	ACAGTG	hsa-miR-1...	2.736	4.5	2.626				
MUC2	MUC2-001	GACCCC	hsa-miR-6...	2.448	2.5	2.931				
CTA-414D7.1	CTA-414D7.1-...	GTACT	hsa-miR-802	1.736	4	2.978				
CTA-414D7.1	CTA-414D7.1-...	TGTAC	hsa-miR-1...	1.728	4	2.978				
LINC01043	LINC01043-001	TCAAGA	hsa-miR-1...	2.383	4.5	3.167				
GA56-AS1	GA56-AS1-001	TCCCTC	hsa-miR-765	2.903	2.5	3.443				
MUC19	MUC19-001	ACAGGG	hsa-miR-3...	2.223	3.5	3.578				
FLJ16779	FLJ16779-001	AGTGT	hsa-miR-3...	3.038	4.5	3.768				
RP11-326	RP11-326C3.1...	ATCCCC	hsa-miR-4...	1.471	3.5	3.899				
RP11-326...	RP11-326C3.1...	GCCCCG	hsa-miR-610	1.398	3.5	3.899				
RP11-326...	RP11-326C3.1...	TGGTG	hsa-miR-7...	1.529	3.5	3.899				
RP11-102...	RP11-10214...	CACCTG	hsa-miR-1...	1.853	4	3.934				
RP11-102...	RP11-10214...	TTTTC	hsa-miR-5...	1.932	2.5	3.934				
RP11-102...	RP11-10214...	ATTCC	hsa-miR-5...	1.937	3.5	3.934				
RP11-102...	RP11-10214...	CCCCAC	hsa-miR-5...	1.848	2	3.934				
FAM230B	FAM230B-002	CGGCCA	hsa-miR-1...	2.489	3	3.952				
HP09025	HP09025-001	CTGGCA	hsa-miR-1...	2.964	4.5	4.044				
LINC01043	LINC01043-001	GCCAGG	hsa-miR-2...	2.297	4	4.191				
MUC19	MUC19-001	CAGGGG	hsa-miR-6...	2.344	3.5	4.207				

B



C



Check your knowledge

- **What is functional annotation?**
- **Why is this difficult?**
- **Which are the newest trends?**
- **How can you best annotate ncRNAs?**
- **What is**
 - **CAFA**
 - **Blast2GO**
 - **Fantasia**
 - **SpongScan**