

**UNIVERSIDADE FEDERAL DE VIÇOSA
CENTRO DE CIÊNCIAS AGRÁRIAS
DEPARTAMENTO DE ZOOTECNIA**

TUTORIAL

Análises de ontologia gênica (Gene Ontology - GO)

Renata de Fátima Bretanha Rocha

Pamela Itajara Otto

Arielly Oliveira Garcia

Mateus Guimarães dos Santos

Marcos Vinicius Gualberto Barbosa da Silva

Marta Fonseca Martins

Marco Antonio Machado

João Cláudio do Carmo Panetto

Simone Eliza Facioni Guimarães

**Viçosa
2023**

TUTORIAL

Análises de ontologia gênica (Gene Ontology - GO)

Renata de Fátima Bretanha Rocha¹, Pamela Itajara Otto², Arielly Oliveira Garcia¹, Mateus Guimarães dos Santos¹, Marcos Vinicius Gualberto Barbosa da Silva³, Marta Fonseca Martins³, Marco Antonio Machado³, João Cláudio do Carmo Panetto³, Simone Eliza Facioni Guimarães¹

¹Departamento de Zootecnia, Universidade Federal de Viçosa, Viçosa, MG 36570-900, Brasil.

²Departamento de Zootecnia, Universidade Federal de Santa Maria, Santa Maria, RS 97105-900, Brasil.

³EMBRAPA – Gado de Leite, Juiz de Fora, MG 36038-330, Brasil

ISBN: 978-65-5668-141-2

DOI: <http://dx.doi.org/10.26626/9786556681412.2023B0001>

Agradecimentos

Agradecemos às fazendas e à Empresa Brasileira de Pesquisa Agropecuária (Embrapa) – Gado de Leite, Juiz de Fora – MG, que forneceram os dados para este estudo.

Financiamento

Este estudo recebeu o apoio financeiro do Conselho Nacional de Desenvolvimento Científico e Tecnológico - Brasil (CNPQ) - Processos 402935/2021-7, 142600/2019-9 e 200147/2022-6, da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES)/PROEX 88887.844747/2023-00 e do Instituto Nacional de Ciência e Tecnologia de Ciência Animal (INCT-CA).

Tutorial

Análises GO no Cytoscape

Mais detalhes do uso do software podem ser consultados no Manual do Cytoscape

(Shannon et al. 2003; DOI: [10.1101/gr.1239303](https://doi.org/10.1101/gr.1239303))

Java version: 1.8.0_162



Para instalar o Cytoscape é preciso ter o Java instalado

Java: <https://www.java.com/pt-BR/download/manual.jsp>












Cytoscape 3.7.0:

<https://github.com/cytoscape/cytoscape/releases/3.7.0/>





INSTALANDO CYTOSCAPE 3.7.0

1º - No link do slide anterior, baixar a versão do Cytoscape condizente com seu pc

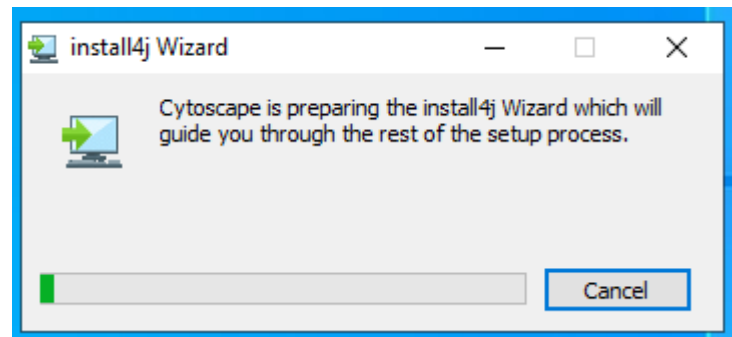
▼ Assets 11

 cytoscape-3.7.0.tar.gz
 cytoscape-3.7.0.zip
 Cytoscape_3_7_0_macos.dmg
 Cytoscape_3_7_0_unix.sh
 Cytoscape_3_7_0_windows_32bit.exe
 Cytoscape_3_7_0_windows_64bit.exe
 md5sums
 output.txt
 updates.xml
 Source code (zip)
 Source code (tar.gz)

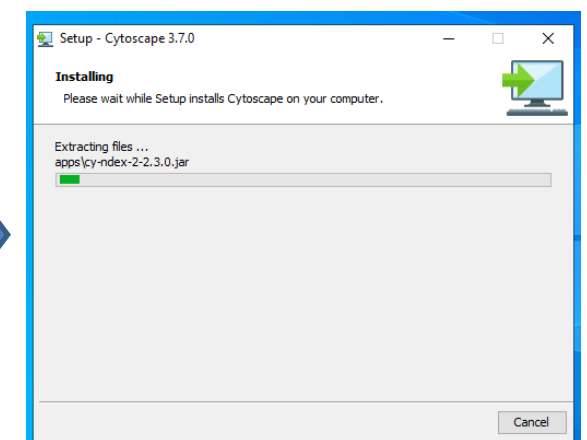
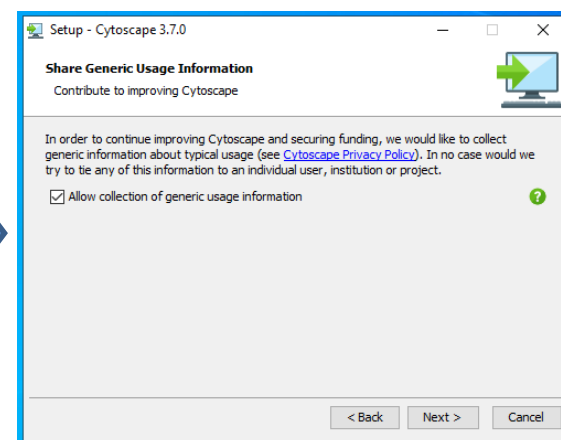
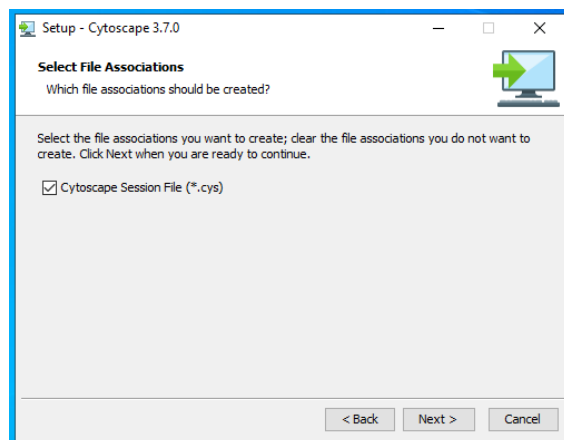
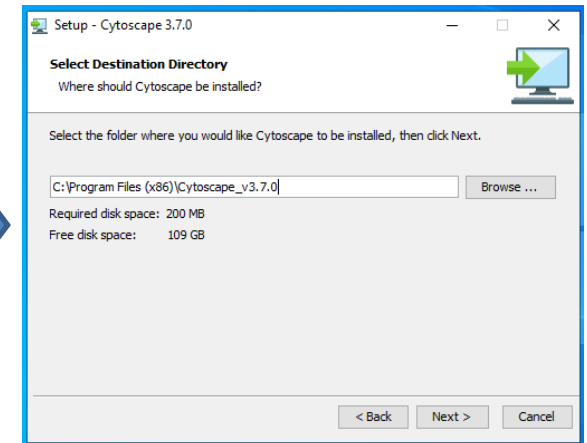
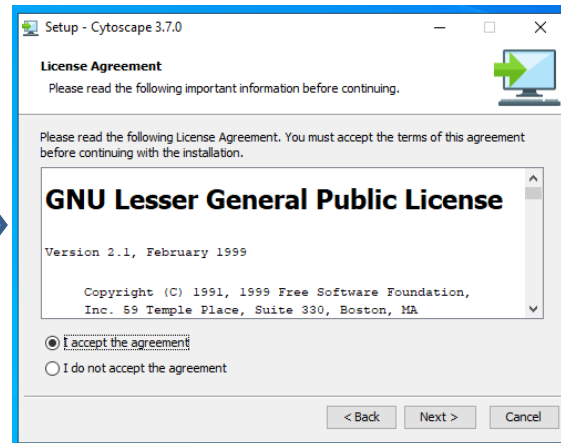
2º - Clicar duas vezes do ícone baixado

 FT	28/
 Bos_taurus.ARS-UCD1.2.104.gtf	27/
 Cytoscape_3_7_0_windows_32bit	31/
 Tutorial	04/

3º - Se não tiver o Java, vai aparecer uma tela perguntando se pode instalar o Java antes de instalar o Cytoscape
- Ou pode ir no site do Java, baixar e instalar
- Se já tiver o Java no seu PC, vai aparecer a seguinte tela, é só aguardar

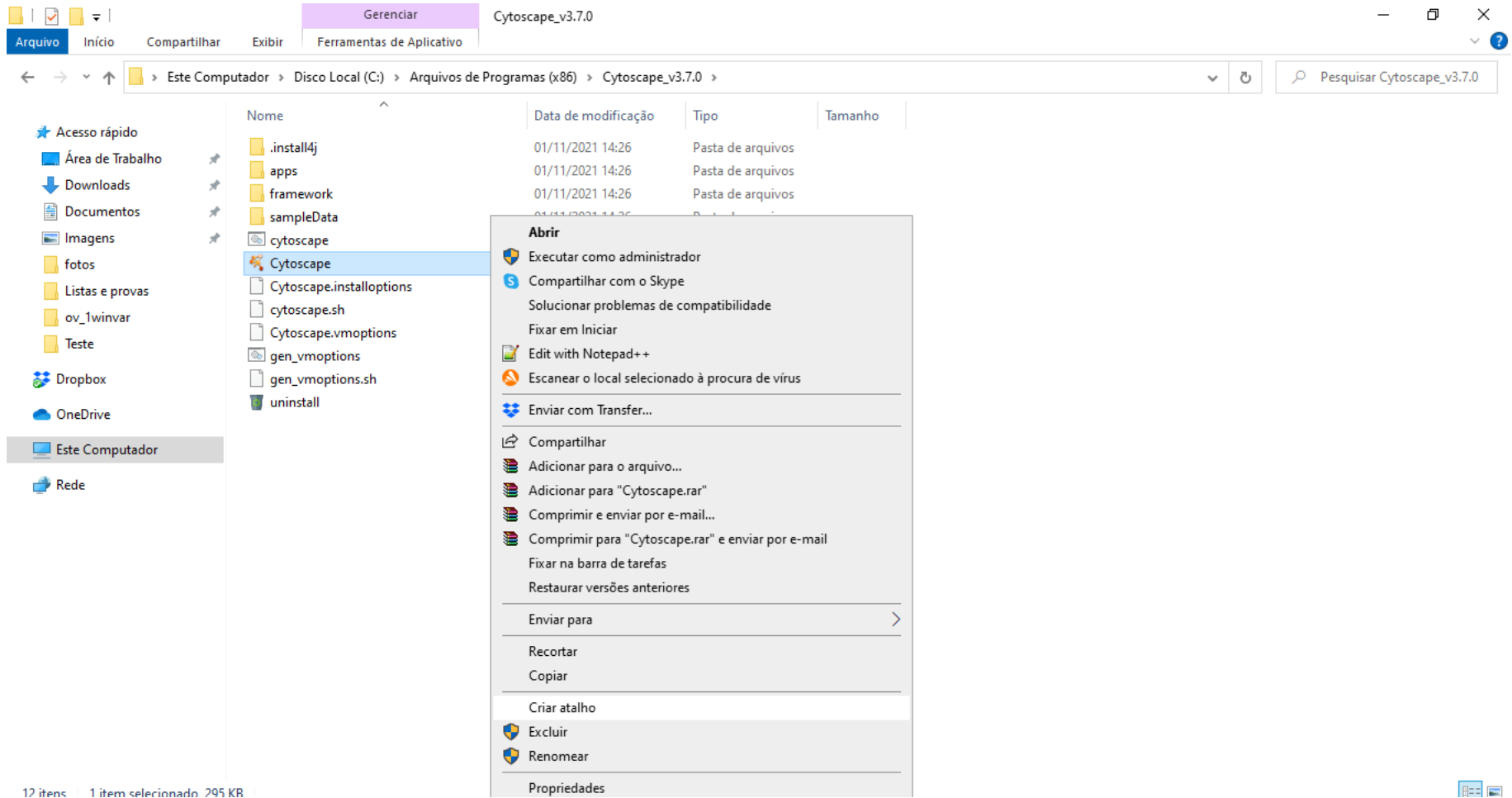


Não alterar nenhum padrão durante a instalação, apenas clicar em Next



- Clicar em *Finish*
- O atalho para o Cytoscape ainda não vai aparecer na área de trabalho
- Para verificar se o programa instalou corretamente e para colocar um atalho do Cytoscape na área de trabalho, verificar o slide seguinte





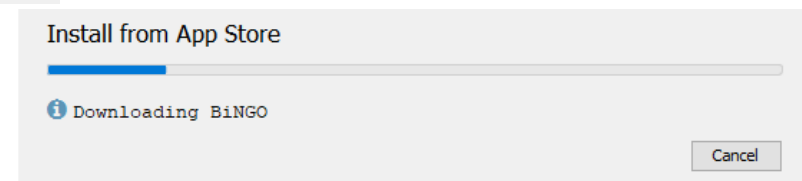
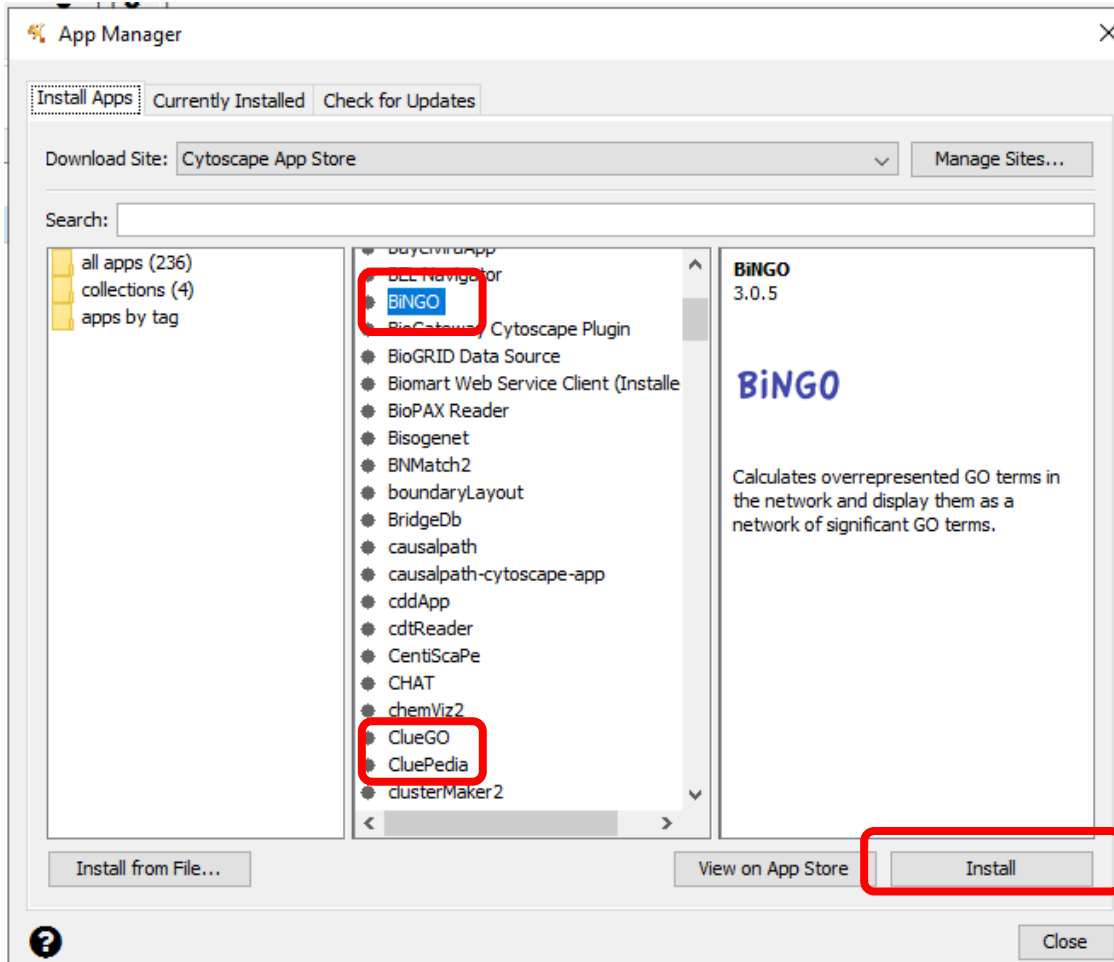
- Verificar se está na pasta direcionada na instalação
- Para verificar se foi instalado, clicar duas vezes no ícone Cytoscape com o desenho
- Se preferir, clicar com o botão direito sobre o mesmo ícone e clicar em 'criar atalho'
- Vai aparecer uma caixa de mensagens dizendo que o atalho não pode ser criado dentro desta pasta e perguntando se quer criar esse atalho na área de trabalho
- Dê OK!

Instalando os Aplicativos BiNGO, ClueGO e CluePedia

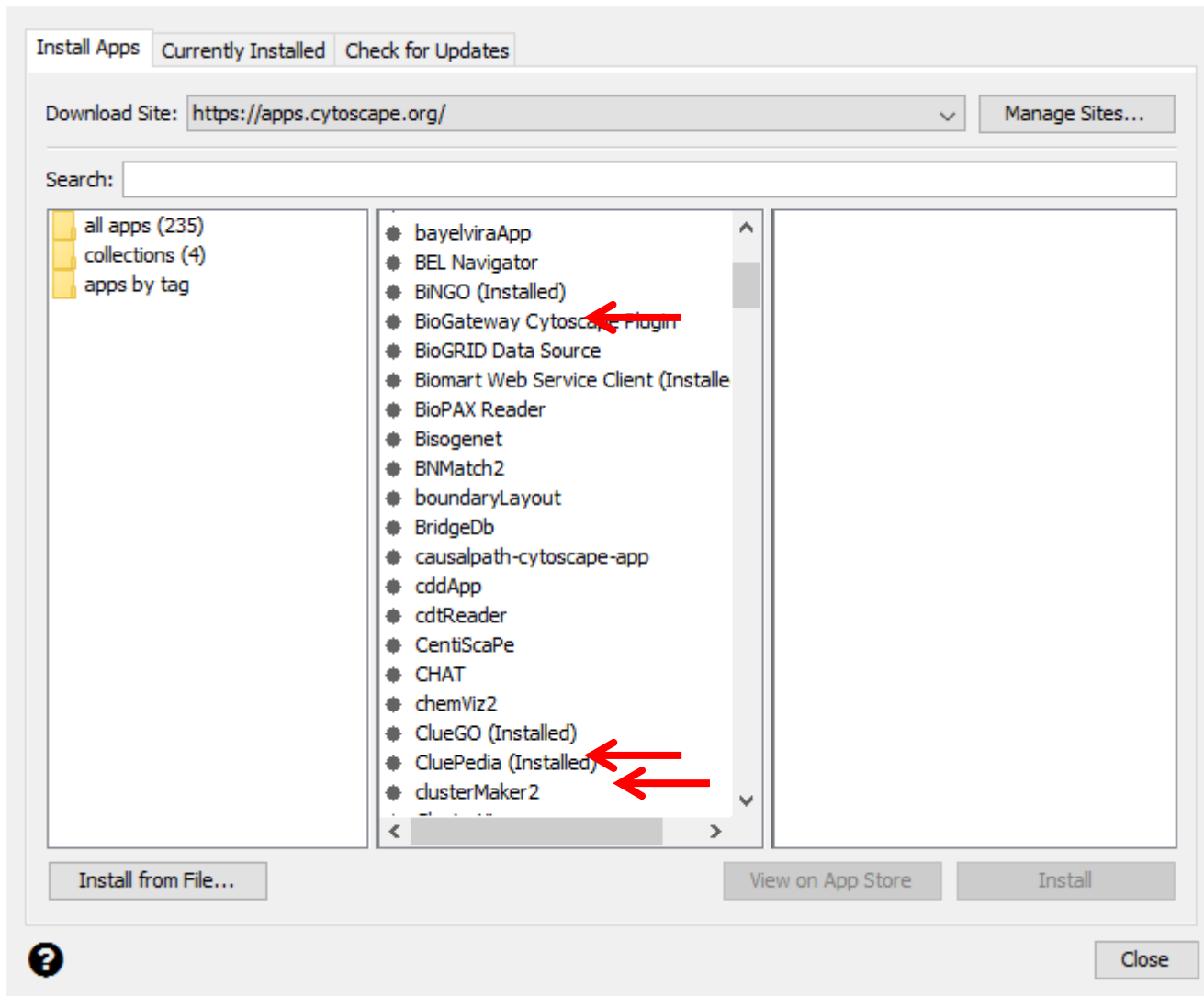
- Clicar em Apps → App Manager...

The image shows a screenshot of the Cytoscape software interface. At the top, the menu bar includes 'File', 'Edit', 'View', 'Select', 'Layout', 'Apps', 'Tools', and 'Help'. A red arrow points to the 'Apps' menu item. Below the menu bar, the 'App Manager...' option is highlighted in the toolbar. The main window displays a network diagram with nodes and edges. A dialog box titled 'App Manager' is open, showing the 'Install Apps' tab. The 'Download Site' is set to 'Cytoscape App Store'. A search bar is present, and the app list is currently empty. A progress window titled 'Cytoscape: Getting available apps' is overlaid on the App Manager dialog, showing a progress bar and the message 'Obtaining apps from: https://apps.cytoscape.org/'. The progress window has a 'Cancel' button. The background network diagram shows a central node labeled 'FOX03' connected to other nodes.

- Vai aparecer a tela com as opções dos aplicativos
- Clicar em BiNGO e clicar em Install
- Depois que terminar, fazer o mesmo com ClueGO e CluePedia, um de cada vez



Depois que está instalado aparece assim:



Mais detalhes dos aplicativos podem ser conferidos na documentação original do BiNGO (Maere et al., 2015; DOI: 10.1093/bioinformatics/bti551) e do ClueGO (Bindea et al. 2009; DOI: 10.1093/bioinformatics/btp101)

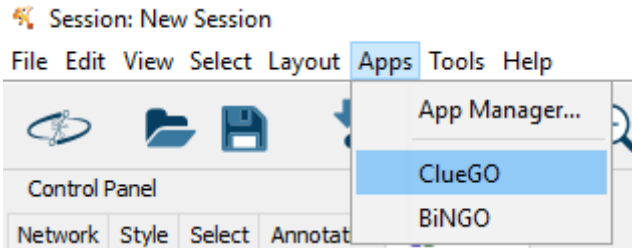
Análises Gene Ontology (GO)

- Os seguintes genes serão usados de exemplo neste tutorial da análise de Gene Ontology:
 - ABL2
 - TMEM245
 - LOC101907107
 - PEAK1

Obtendo os processos biológicos associados aos genes

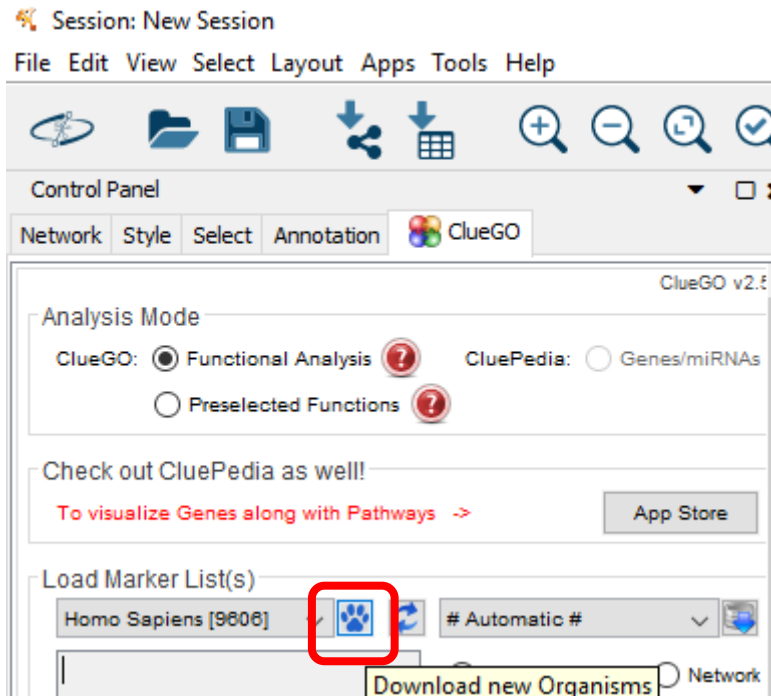
CLUEGO

ClueGO



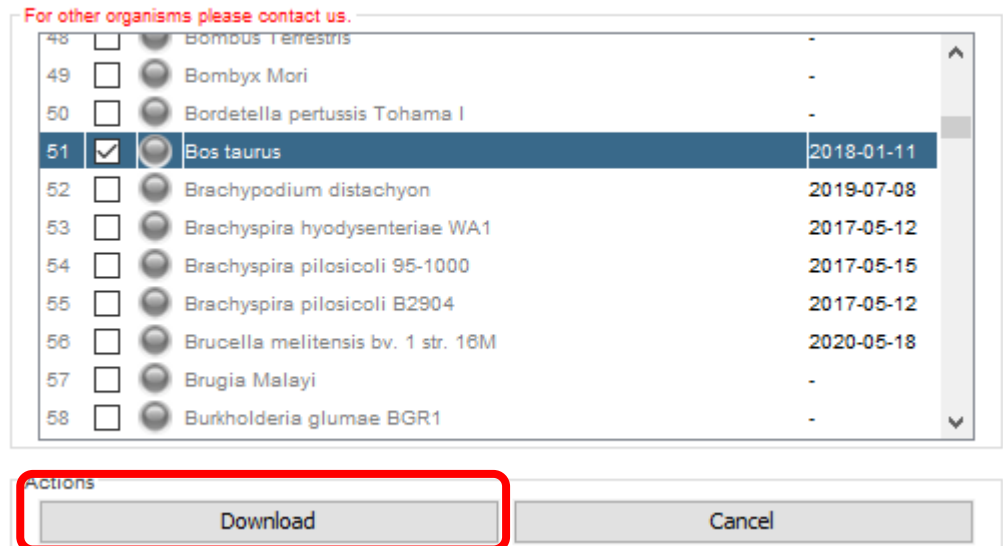
*Para usar o ClueGO é preciso ter uma licença. Para obter uma, usar o e-mail institucional, porque eles não enviam licença para e-mail pessoal. Essa licença pode ser usada por até 5 ou 6 pessoas

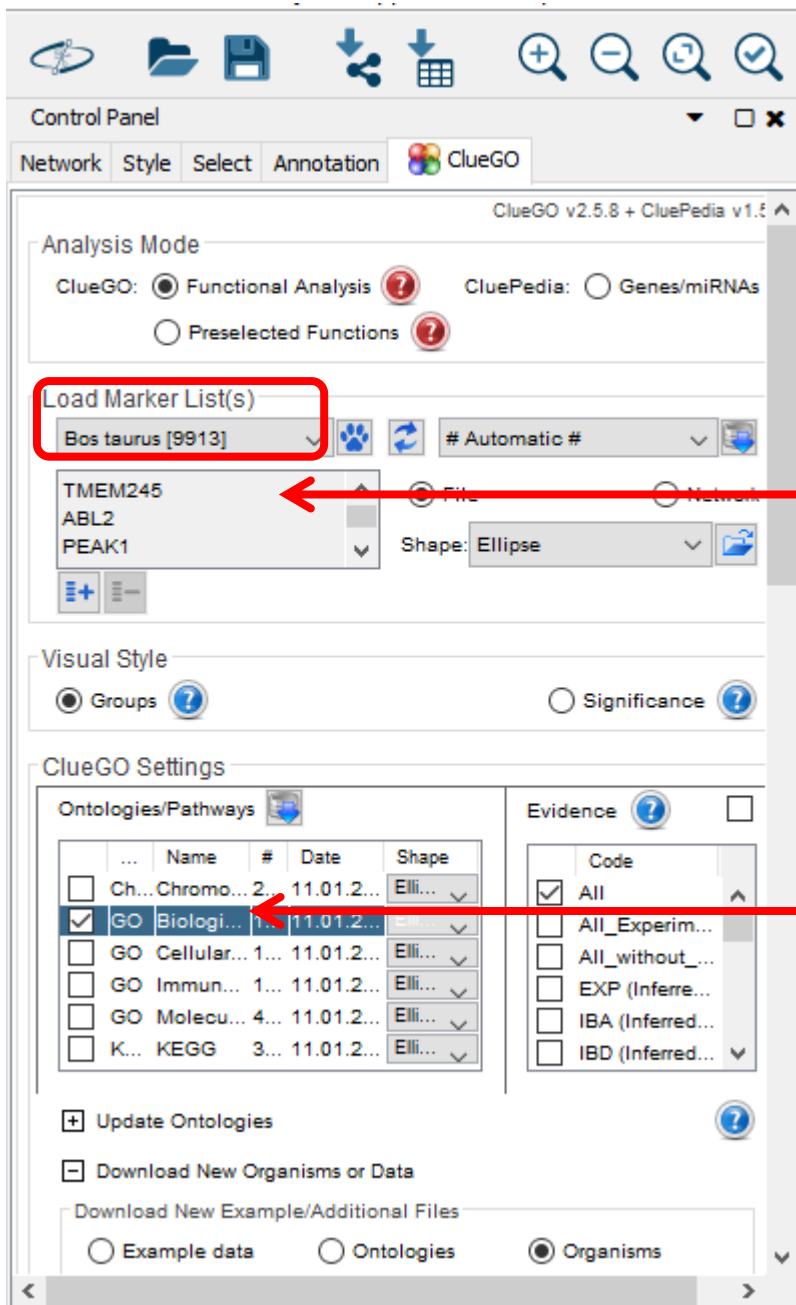
1º - Clicar na patinha



2º - Selecionar *Bos taurus* e fazer download

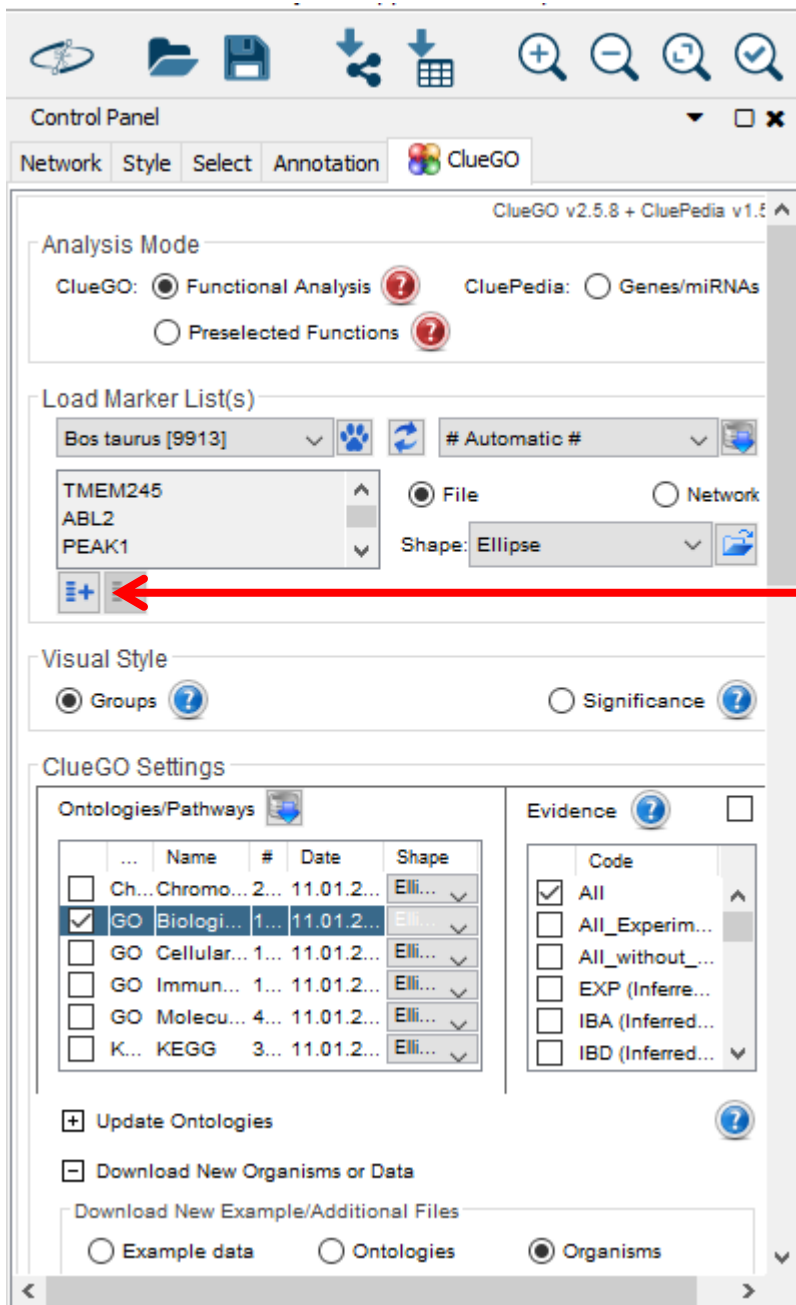
Se já tiver a espécie de interesse instalada pode apenas selecioná-la na setinha pra baixo



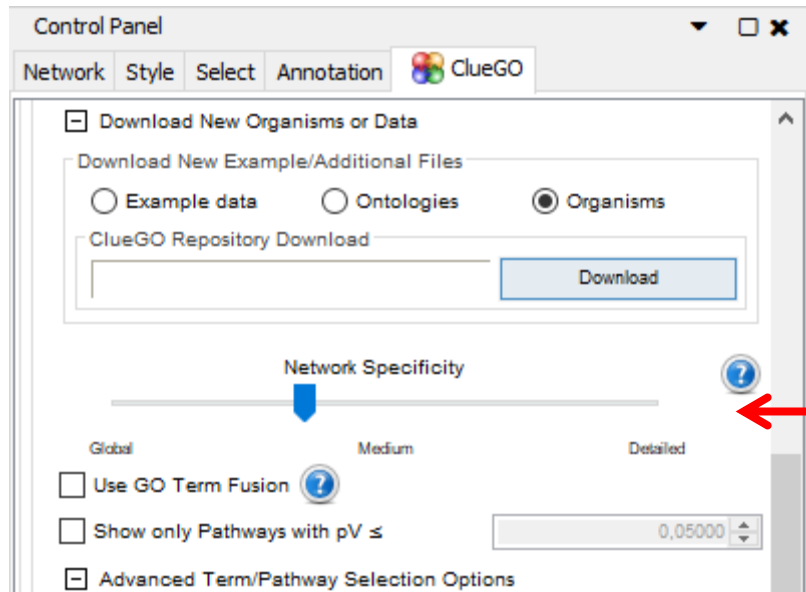


- Digitar o nome dos **genes** (Se tiver muitos genes, copiar de um arquivo pronto ou importar)

- Selecionar a opção de processos biológicos

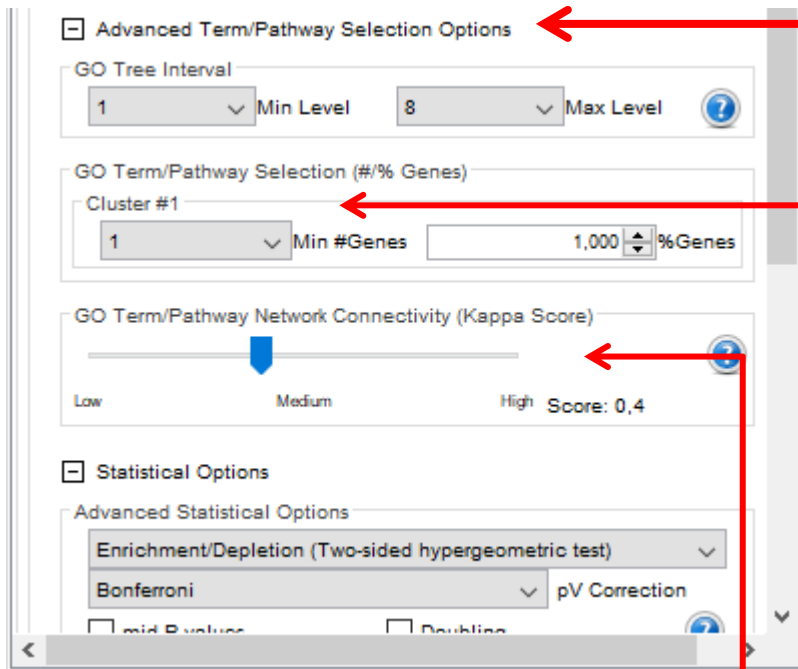


- Nesta parte, se tiver diferentes grupos de genes, por exemplo, para características diferentes, clicar no + e digitar os nomes dos genes de cada grupo em cada caixa.
- Se não tiver, pode ignorar essa parte



- Em **Network Specificity**
 - Quando se tem muitos genes, podemos arrastar a setinha para deixar uma análise mais detalhada / específica
 - Quando temos poucos genes, podemos fazer uma análise mais geral

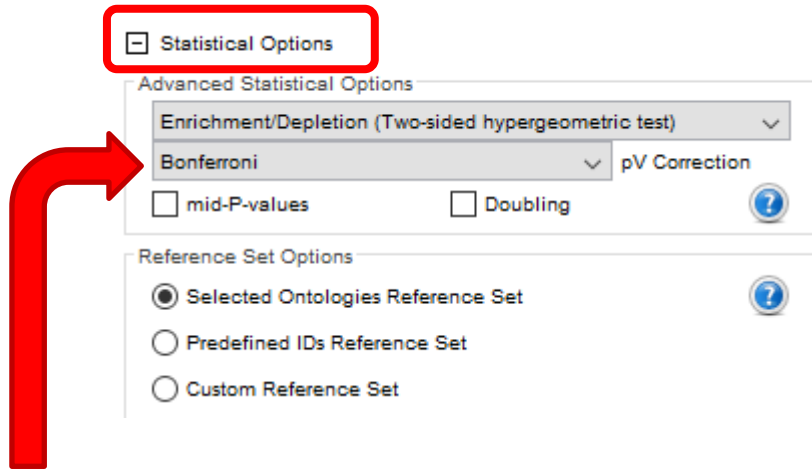
Ele não gera um valor para essa Network, mas é bom lembrarmos como esse parâmetro foi definido (se mais global, mais detalhado ou médio, ou outra opção nesse intervalo)



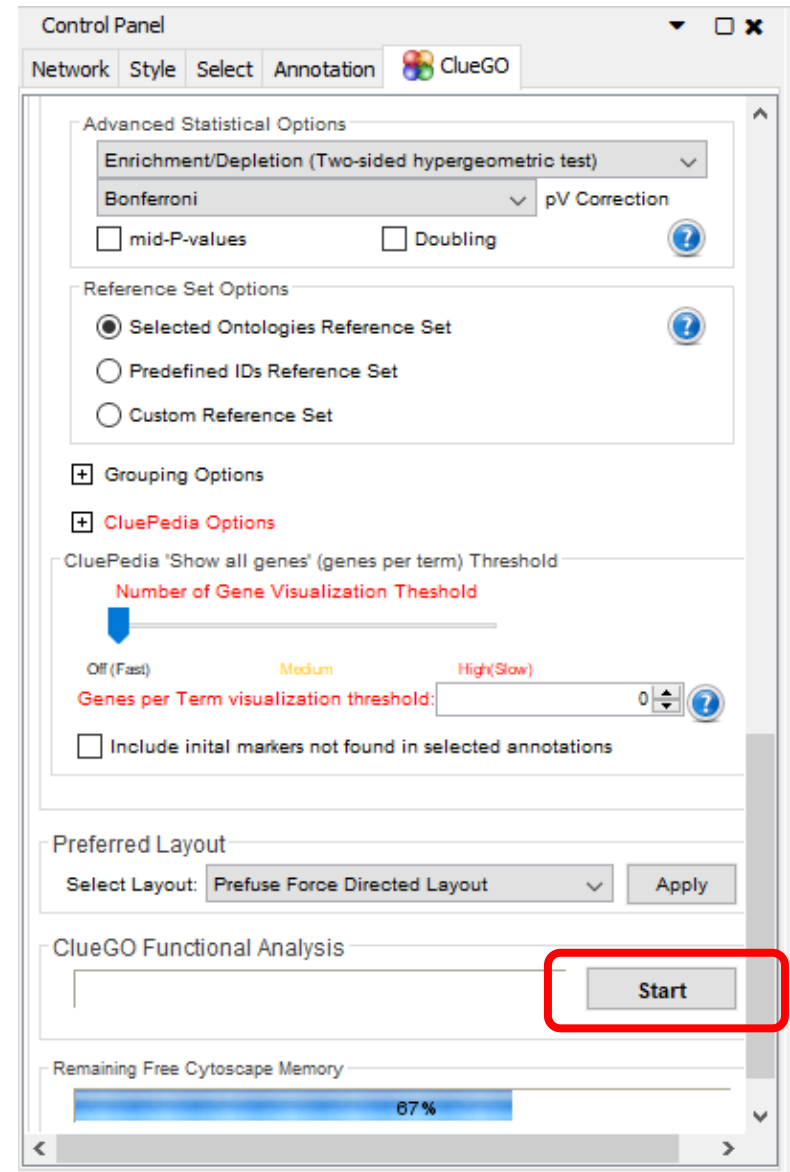
- Em **Advanced Term/Pathway Selection Options**
- GO Tree Interval
 - Mín = 1 e Máx = 8
- GO Term/Pathway Selection (#% Genes)
 - Mín = 1
- Não alterar o Kappa Score, pois o mínimo padrão é 0,4 e isso já fica na análise

Estes parâmetros, assim como a Network Specificity (slide anterior), podem ser alterados para encontrar os resultados mais ajustados para o nosso objetivo, por isso essa parte pode ser realizada mais de uma vez para irmos comparando qual resultado fica melhor

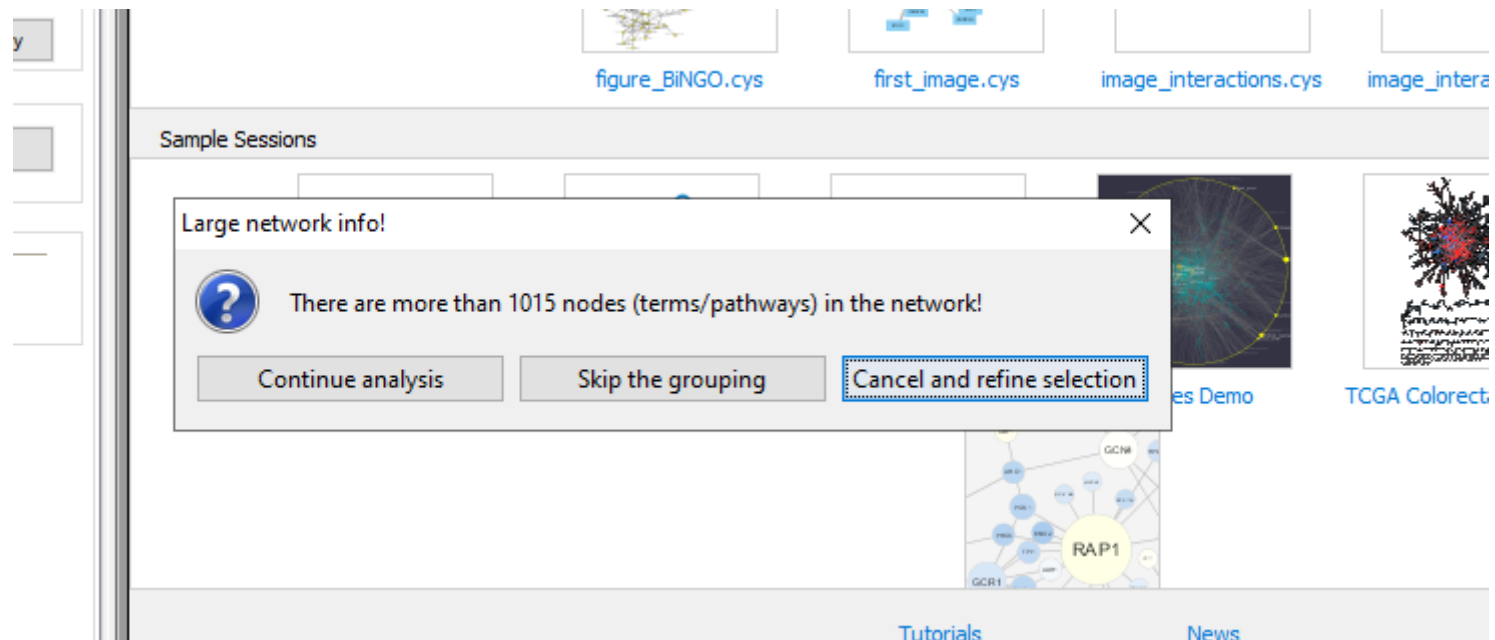
Clicar em Statistical Options



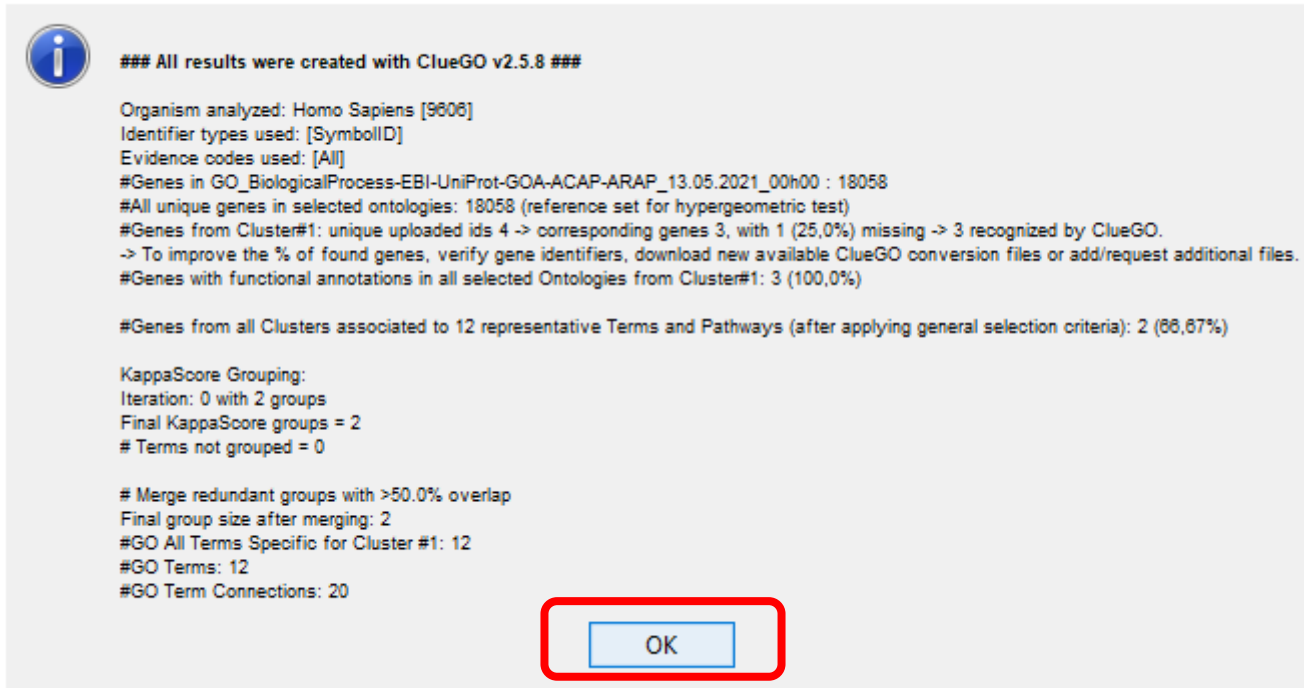
- Podemos escolher o teste estatístico mais apropriado
 - Neste caso, a escolha foi o teste de Bonferroni
 - Não alterar mais nenhum parâmetro
- Clicar em Start



Como isto é apenas um exemplo temos poucos genes, mas na pesquisa real pode ser que tenha muitos genes. Neste caso, o Cytoscape pode dar esse aviso:



Se a análise prosseguir normalmente, vai mostrar a descrição dos resultados



All results were created with ClueGO v2.5.8

Organism analyzed: Homo Sapiens [9606]
Identifier types used: [SymbolID]
Evidence codes used: [All]
#Genes in GO_BiologicalProcess-EBI-UniProt-GOA-ACAP-ARAP_13.05.2021_00h00 : 18058
#All unique genes in selected ontologies: 18058 (reference set for hypergeometric test)
#Genes from Cluster#1: unique uploaded ids 4 -> corresponding genes 3, with 1 (25,0%) missing -> 3 recognized by ClueGO.
-> To improve the % of found genes, verify gene identifiers, download new available ClueGO conversion files or add/request additional files.
#Genes with functional annotations in all selected Ontologies from Cluster#1: 3 (100,0%)

#Genes from all Clusters associated to 12 representative Terms and Pathways (after applying general selection criteria): 2 (66,67%)

KappaScore Grouping:
Iteration: 0 with 2 groups
Final KappaScore groups = 2
Terms not grouped = 0

Merge redundant groups with >50.0% overlap
Final group size after merging: 2
#GO All Terms Specific for Cluster #1: 12
#GO Terms: 12
#GO Term Connections: 20

OK

Nesta etapa o Cytoscape também indica quantos dos genes usados foram reconhecidos.

Geralmente microRNAs, RNAs e outras regiões não codificadoras como U2, U6, LOCs, etc não são identificadas pelo programa.

Na parte inferior:

- Clicar em CluePedia
- Clicar na 2ª opção de quadrinho: 'Show all genes from all Pathways/Terms'

The screenshot displays the ClueGO v2.5.8 + CluePedia v1.1.5 interface. The main window shows a network diagram with green nodes and grey edges. The nodes are labeled with GO terms such as "regulation of Rho protein signal transduction", "negative regulation of Ras protein", "signal transduction", "regulation of actin cytoskeleton reorganization", "positive regulation of small GTPase mediated signal transduction", "exploration behavior", "cellular response to retinoic acid", "positive regulation of oxidoreductase activity", "focal adhesion assembly", and "regulation of cell-substrate organization". A red box highlights the "CluePedia" button in the bottom right corner of the interface. Another red box highlights the second icon in the bottom left corner of the interface, which is the "Show all genes from all Pathways/Terms" option. The interface also includes a Control Panel on the left with tabs for Network, Style, Select, and Annotation. The bottom panel shows a Table Panel with columns for ClueGO P-terms (Cluster #1), Cluster #1, ClueGO Lo, and CluePedia. The CluePedia column is highlighted with a red box. The bottom right corner shows a Memory indicator.

Clicar na setinha (Detach View) na parte inferior para expandir a tela das interação

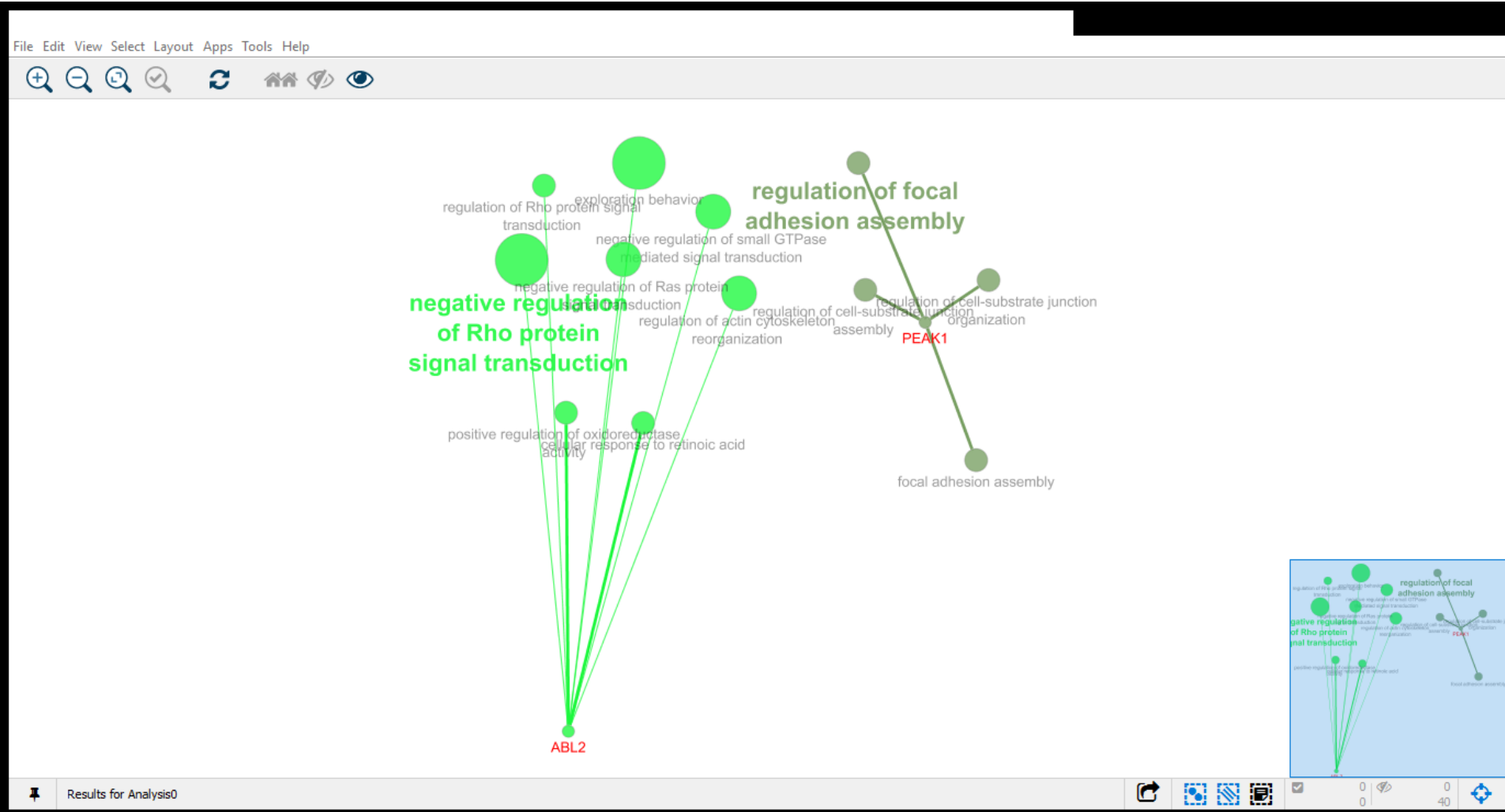
The screenshot displays the ClueGO v2.5.8 + CluePedia v1.4 interface. The main window shows a network diagram with nodes and edges. The nodes are labeled with biological processes, including "negative regulation of Rho protein signal transduction", "regulation of focal adhesion assembly", "regulation of Ras protein", "regulation of actin cytoskeleton reorganization", "regulation of cell-substrate junction assembly", "focal adhesion assembly", "positive regulation of oxidoreductase activity", and "response to amino acid". The network is visualized with green nodes and edges, and a red node is visible at the bottom.

The interface includes a Control Panel on the left with the following sections:

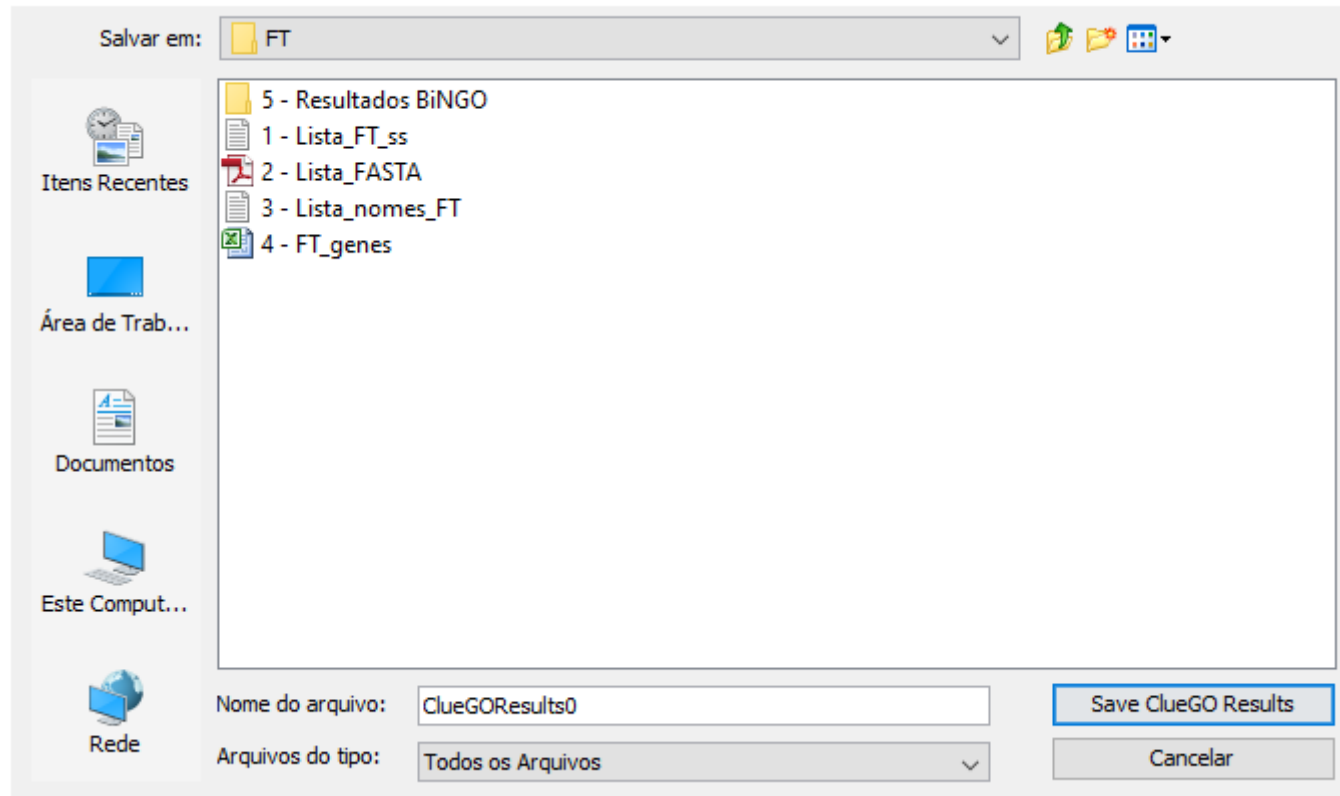
- Analysis Mode:** ClueGO: Functional Analysis, CluePedia: Genes/miRNAs, Preselected Functions.
- Load Marker List(s):** Homo Sapiens [9806], # Automatic #, File (selected), Network, Shape: Ellipse.
- Visual Style:** Groups, Significance.
- ClueGO Settings:** Ontologies/Pathways (GO Biolog... 1... 13.05... selected), Evidence (All checked).

The bottom of the interface features a toolbar with a red box highlighting the "Detach View" icon (a small square with a diagonal line). Below the toolbar, there is a "Table Panel" and a "Results for Analysis0" section with tabs for ClueGO Results (Cluster #1), Cluster #1, ClueGO Log, and CluePedia. A "Memory" indicator is visible in the bottom right corner.












Nesta tela dá para dar zoom nos processos biológicos



Direcionar os resultados para a pasta



Os resultados / output na pasta

Nome	Data de modificação	Tipo	Tamanho
 ClueGOResults0 BinaryGeneTermMatrix	12/11/2021 16:56	Documento de Te...	1 KB
 ClueGOResults0 ClueGOLog	12/11/2021 16:56	Documento de Te...	2 KB
 ClueGOResults0 EdgeAttributeTable	12/11/2021 16:56	Documento de Te...	3 KB
 ClueGOResults0 Functional Groups With ...	12/11/2021 16:56	Documento de Te...	1 KB
 ClueGOResults0 Genes With Correspondi...	12/11/2021 16:56	Documento de Te...	1 KB
 ClueGOResults0 KappaScoreMatrix	12/11/2021 16:56	Documento de Te...	2 KB
 ClueGOResults0 NodeAttributeTable	12/11/2021 16:56	Documento de Te...	8 KB
 ClueGOResults0 Overview Specific Cluste...	12/11/2021 16:56	Arquivo PNG	15 KB
 ClueGOResults0 Overview Specific Cluste...	12/11/2021 16:56	Microsoft Edge H...	17 KB
 ClueGOResults0 Specific Cluster #1	12/11/2021 16:56	Arquivo PNG	15 KB
 ClueGOResults0 Specific Cluster #1	12/11/2021 16:56	Microsoft Edge H...	36 KB

Os arquivos interessantes:

- ClueGOResults0 BinaryGeneTermMatrix
 - Esse arquivo mostra os processos biológicos (PB) nas linhas e os genes nas colunas e monta uma matriz associando os genes com cada PB
- ClueGOResults0 Functional Groups With Genes
 - Esse arquivo tem os resultados dos processos biológicos agrupados por grupos funcionais de genes
- ClueGOResults0 NodeAttributeTable
 - Este arquivo tem os PB e o P-valor e P-valor ajustado pelo teste de Bonferroni indicando quais genes e PB são significativos e quais não são

ClueGOResults0 NodeAttributeTable

- Este arquivo contém os gene, os processos biológicos, o P-valor, o P-valor ajustado para Bonferroni e outras informações...
- Para selecionar os genes geralmente usa-se o p-valor ajustado que chamamos de FDR pra evitar falso positivo
- Isso é usado quando existem muitas variáveis, nesse caso, os genes
- Por mais que as análises de redes tenham menos genes, é interessante padronizar
- “The Bonferroni correction adjusts probability (p) values because of the increased risk of a type I error when making multiple statistical tests.” (Armstrong, 2014. DOI: 10.1111/opo.12131)

Busca das sequências promotoras no site NCBI

Busca dos fatores de transcrição no site TFM-Explorer

FATORES DE TRANSCRIÇÃO

1º passo

- No site NCBI: <https://www.ncbi.nlm.nih.gov/>
- Pesquisar o nome do gene no NCBI para a espécie estudada
- Clicar na opção do gene identificada para a espécie

The screenshot shows the NCBI Gene database search results for 'ABL2 cattle'. The search bar at the top contains 'Gene' and 'ABL2 cattle', with a red box highlighting the search input area. Below the search bar, there is a 'COVID-19 Information' banner. The main content area displays the search results for 'ABL2 - ABL proto-oncogene 2, non-receptor tyrosine kinase' in 'Bos taurus (cattle)'. A red box highlights the gene name. The results include the gene name, 'Also known as: ABLL', 'Gene ID: 511845', and links to 'RefSeq transcripts (8)', 'RefSeq proteins (8)', and 'PubMed (2)'. There are buttons for 'Orthologs', 'Genome Data Viewer', 'BLAST', and 'Download'. The 'RefSeq Sequences' section is partially visible at the bottom. On the right side, there are sections for 'Filters: Manage Filters', 'Results by taxon' (listing 'Top Organisms' like 'Bos taurus (1)', 'Bos indicus x Bos taurus (1)', and 'Bos indicus (1)'), 'Find related data' (with a 'Database' dropdown and a 'Find items' button), and 'Search details' (showing the search query: '(ABL2[All Fields] AND ("Bos taurus" [Organism] OR CATTLE[All Fields])) AND alive[prop]').

3º passo

- Na nova página, verificar as posições iniciais dos genes nos quadrinhos “from” e “to” no canto direito
- Se o gene é Forward, a posição inicial dele está no quadrinho “from”, significa que o gene começa nesta posição
- Se o gene é Reverse, a posição inicial dele está no quadrinho “to”, significa que o gene começa nesta posição

The screenshot shows the NCBI Nucleotide database interface. The main content area displays the sequence for **Bos taurus isolate L1 Dominette 01449 registration number 42190680 breed Hereford chromosome 16, ARS-UCD1.2, whole genome shotgun sequence**. The sequence is shown in FASTA format. On the right side, there is a panel titled **Change region shown** with a red box around it. This panel has two radio buttons: **Whole sequence** (unselected) and **Selected region** (selected). Below the radio buttons, there are input fields for **from: 60454698** and **to: 60558562**. There is an **Update View** button below these fields. Below the **Change region shown** panel, there is a **Customize view** panel with a **Display options** section. The **Show reverse complement** checkbox is checked, and the **Show gap features** checkbox is unchecked. There is an **Update View** button below this panel. At the bottom of the page, there are sections for **Analyze this sequence** (with a **Run BLAST** button), **Pick Primers**, **Highlight Sequence Features**, and **Find in this Sequence**.

4º passo

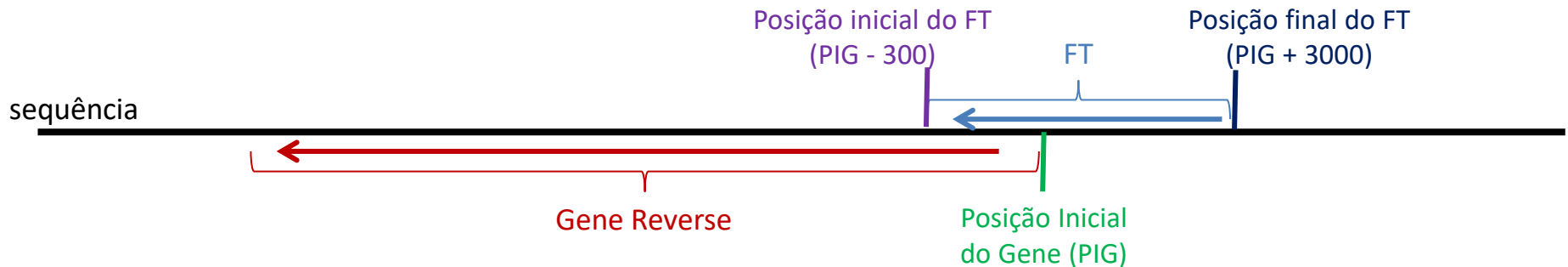
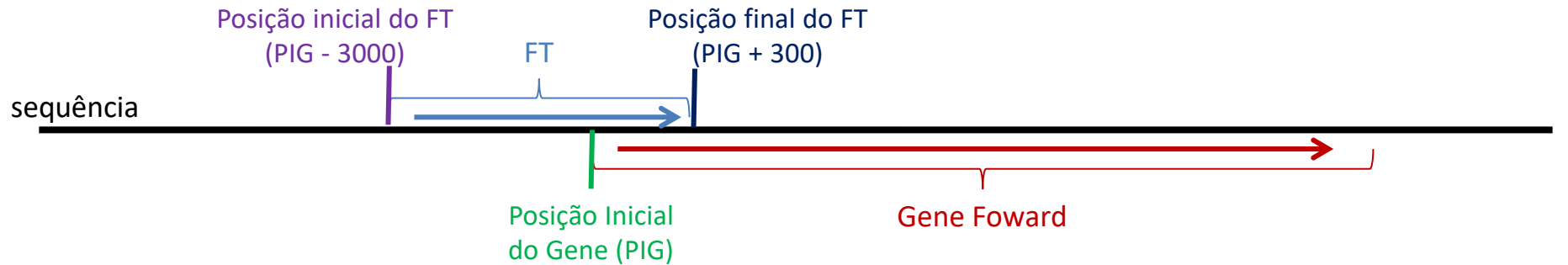
Anotar/digitar as posições iniciais de cada gene e salvar em um arquivo (por precaução)

Para as análises seguintes, precisaremos do fator de transcrição (FT) de cada gene.

Como os FT se situam em uma posição anterior ao início do genes, precisamos calcular as posições iniciais e finais dos fatores de transcrição de cada gene

Para genes Foward: assumimos que o FT começa 3000 pb antes e termina 300pb depois

Para genes Reverse: assumimos que o FT começa 3000 pb depois e termina 300pb antes



Bp: base pairs – pares de bases; O tamanho do FT entre 3000 e 300bp pode ser diferente em outros trabalhos

4º passo

- Na tabela do Excel abaixo é possível calcular as posições iniciais e finais dos fatores de transcrição
- Isso também pode ser feito na calculadora ou criar uma planilha no Excel à parte
- Na planilha abaixo, clicar duas vezes na tabela
- Pegar o valor da posição inicial do gene no 3º passo (por isso é importante saber se é Forward ou Reverse)
- Colar na cédula verde (abaixo da cédula com título 'Posição inicial do GENE')
- Dar ENTER
- Ao lado vão aparecer as posições iniciais e finais do FT

Para genes Foward			
Posição inicial do GENE		Posição inicial do FT	Posição final do FT
0		-3000	300
Para genes Reverse			
Posição inicial do GENE		Posição inicial do FT	Posição final do FT
60558562		60558262	60561562

Na tabela, temos o exemplo do gene ABL2

- É um gene Reverse
- Sua posição inicial é 60558562
- Assim
- ✓ A posição inicial do FT é 60558262
- ✓ A posição final do FT é 60561562

- É bom deixar essas informações anotadas em uma planilha
- É importante salvar essas informações, principalmente considerando que este é um exemplo pequeno com poucos genes, mas podem ser identificados mais de 40, 60, 80 genes, etc..

Gene	Sentido	Pos. Original	Nova pos. inicial	Nova pos final
ABL2	Reverse	60558562	60558262	60561562
TMEM245
LOC101907107
PEAK1