



Laboratório Ronsein de Proteômica
Instituto de Química
Universidade de São Paulo

Título: Roteiro para Análise de Dados de Proteômica no Software Perseus	Ref.: PERSEUS-SHOTGUN_BRPROT_02	Versão: 2
--	--	------------------

Preparado por: Mariana P. Massafera	Data: 05/09/2025
-------------------------------------	------------------

OBJETIVOS

Workflow para uma análise de dados *shotgun-DDA* convencionais, utilizando-se como exemplo um conjunto de 12 amostras no total, sendo 6 réplicas por grupo (controle vs. tratado).

REFERÊNCIAS

Tyanova, Stefka, and Juergen Cox. 2018. "[Perseus: A Bioinformatics Platform for Integrative Analysis of Proteomics Data in Cancer Research](#)." In, 133–48. Springer New York.

Tyanova, Stefka *et al.* 2016. "[The Perseus Computational Platform for Comprehensive Analysis of \(Prote\)omics Data](#)." *Nature Methods* 13 (9): 731–40.

Tyanova, Stefka *et al.* 2016. "[Proteomic Maps of Breast Cancer Subtypes](#)." *Nature Communications* 7 (1).

Sinitcyn, Pavel *et al.* 2018. "[Computational Methods for Understanding Mass Spectrometry–Based Shotgun Proteomics Data](#)." *Annual Review of Biomedical Data Science* 1: 207–234.

Cox, Juergen, and Matthias Mann. 2012. "[1D and 2D Annotation Enrichment: A Statistical Method Integrating Quantitative Proteomics with Complementary High-Throughput Data](#)." *BMC Bioinformatics* 13 (S16).

Rudolph, Jan Daniel, and Jürgen Cox. 2019. "[A Network Module for the Perseus Software for Computational Proteomics Facilitates Proteome Interaction Graph Analysis](#)." *Journal of Proteome Research* 18 (5): 2052–64.

Tutoriais online sobre o Perseus oferecidos nas várias edições da MaxQuant Summer School. Disponíveis em:
<<https://www.youtube.com/channel/UCKYzYTm1cnmc0CFAMhxDO8w/videos>>



DADOS EXPERIMENTAIS UTILIZADOS PARA ESTE TUTORIAL

Article: Amor, M. *et al.* Identification of regulatory networks and crosstalk factors in brown adipose tissue and liver of a cold-exposed cardiometabolic mouse model. *Cardiovasc Diabetol* **23**, 298 (2024). <https://doi.org/10.1186/s12933-024-02397-7>

PRIDE Database: <https://www.ebi.ac.uk/pride/archive/projects/PXD051640>

Description: Brown adipose tissue (BAT) has gained central attention due to its known ability to dissipate energy and counteract cardiometabolic diseases (CMDs). The cold exposure on BAT activation and the induction of thermogenesis have been extensively studied, but reports on the effects at the proteomic level in metabolic tissues of animal models mimicking CMD in humans are lacking.

Sample Processing Protocol: BAT and liver from LdlrKO mice housed at either 22°C or 5°C (n = 6/group) were homogenized in 8 M urea, Tris-HCl 0.1 M pH 8.5, supplemented with protease inhibitors at a ratio of 1:100 (# 5872S; Cell Signaling, Danvers, MA). The liver digestion and sample homogenization were processed as reported elsewhere (14). For the BAT in each sample, in 20 µg of total proteins, the corresponding volume of Ammonium bicarbonate solution 50 mM were added (final pH = 8.5, final volume 60 µL). The proteins were reduced then by incubation with 3 µl DTT 100 mM, for 30 min at 55 °C. For the protein alkylation 6 µl of iodoacetamide 150 mM were added and incubated for 20 min in the dark, at room temperature. Trypsin digestion (Merck, enzyme to protein ratio 1:20), was performed overnight at 37 °C and stopped by acidification with trifluoroacetic acid (final percentage 1%). The LC-MS/MS analysis were carried in An Ultimate 3000 nano-LC system (Thermo Fisher Scientific) connected to an Orbitrap Fusion Tribrid Mass spectrometer (Thermo Fisher Scientific) equipped with a nanoelectrospray ion source.

Link para pasta GDrive onde se encontram os arquivos *.raw e o resultado da search no software MaxQuant:

<https://drive.google.com/drive/folders/1i0N1nSUyGNCrgdiUITswhu2nYa1zsO9G?usp=sharing>

SUGESTÃO DE PROTOCOLO PARA ANÁLISE DOS DADOS

1) Carregamento dos dados no Perseus¹

- a) Generic Matrix Upload \ buscar o arquivo ProteinGroups.txt resultante de sua *search* no MaxQuant \ transferir para a coluna da direita (Main) as colunas “LFQ Intensity” de todas as amostras
- b) Clicar em “OK”

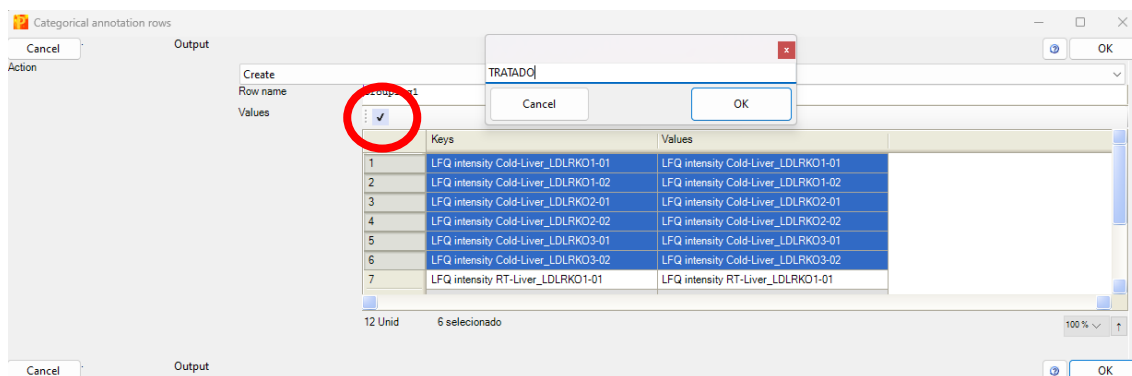
¹ Para o preparo deste protocolo foi utilizado o software Perseus versão 2.1.1.0.

2) Filtragem dos dados

- Filter rows \ Filter rows based on categorical column
- Em Column, selecionar “Reverse”
- Transferir “+” para a coluna da direita
- Em Mode, selecionar “Remove matching rows”
- Em Filter mode, selecionar “Reduce matrix”
- Clicar em “OK”
- Repetir itens a-f (um a um) para opções “Potential contaminants”, “Only identified by site” e “Taxonomy ID, -1”.

3) Nomear os diferentes grupos de amostras (ex: TRATADO; CONTROLE)

- Annot. Rows \ Categorical annotation rows
- Selecionar todas as réplicas de um mesmo grupo
- Clicar no símbolo “v” para renomear as amostras, de forma que todas as réplicas de um mesmo grupo tenham o mesmo nome.
- Repetir os passos b-c para as réplicas do outro grupo
- Clicar em “OK”



4) Transformar os dados em $\log_2(x)^2$

- Basic \ Transform \ $\log_2(x)$
- Transferir “LFQ Intensity” de todas as amostras para a coluna da direita
- Clicar em “OK”

² A matriz obtida ao término desta etapa pode ser preparada para análise por PCA – *Principal Component Analysis*. Também, ela é usada para o cálculo das proteínas exclusivas, conforme será detalhado à frente.



5) Filtragem de valores válidos

a) Filter rows \ Filter rows based on valid values

Min. valids = 3³

Mode = "In each group"

Grouping = Group 1

Values should be = Valid

Filter mode = Reduce matrix

Clicar em "OK"⁴

6) Análise estatística⁵

a) Tests \ Two-sample test

Selecionar "first group (right)" e "second group (left)"

Selecionar o tipo de teste estatístico desejado (Student's T-test)

No campo "Use for truncation", selecionar a opção desejada (mais utilizados:

Permutation-based FDR; Benjamini-Hochberg FDR)

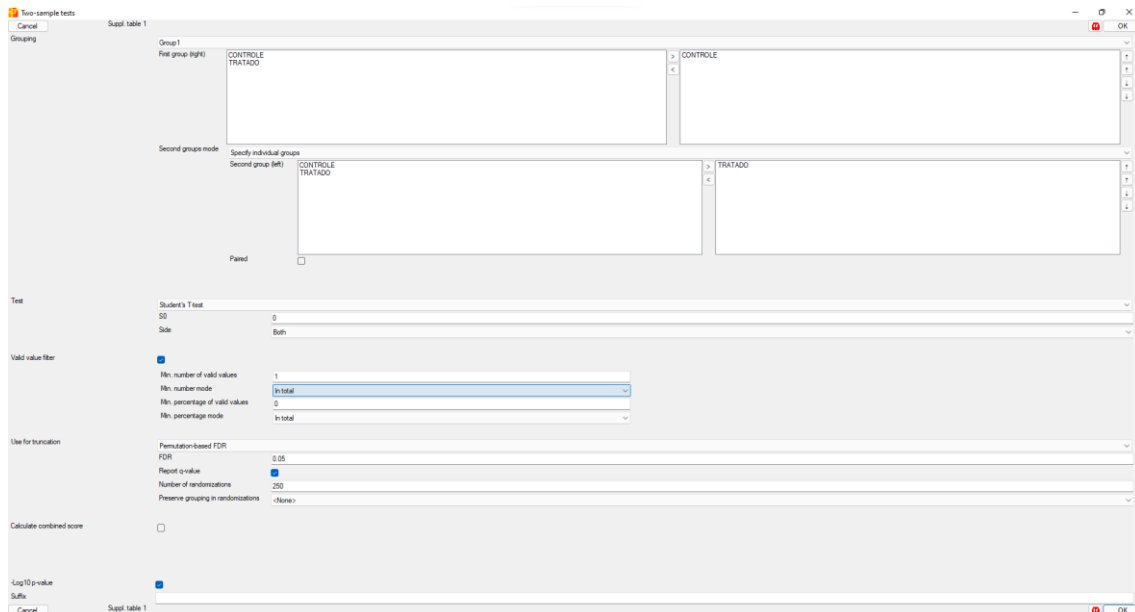
Os outros parâmetros podem ser mantidos com os valores *default*

Clicar em "OK"

³ Neste exemplo, pede-se que sejam mantidas apenas as proteínas que tiveram intensidade LFQ maior que zero em pelo menos 3 das réplicas, em cada grupo. Aqui pode-se ser mais restritivo, a depender do número de réplicas disponíveis, por exemplo.

⁴ Com a matriz resultante desta etapa, pode-se criar o *Volcano plot* (ver instruções à frente).

⁵ Exemplo para quando se tem dois grupos de amostras. Para outras configurações, escolher adequadamente o tipo de teste estatístico a ser empregado.



7) Seleção das proteínas diferencialmente abundantes

a) Filter rows \ Filter rows based on categorical column

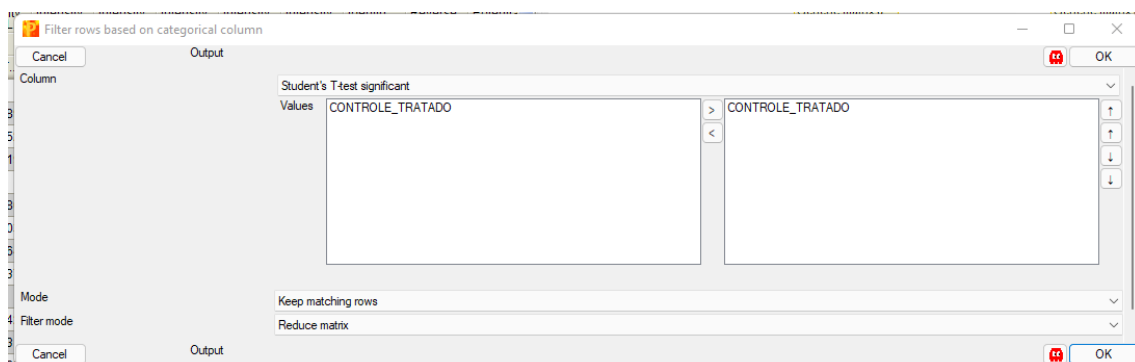
Column = [nome do teste estatístico utilizado. Neste caso = Student's T-test significant]
 Values = transferir o valor para a coluna da direita [se a coluna da esquerda estiver vazia, isso significa que não foram identificadas proteínas diferencialmente abundantes entre os grupos, com os parâmetros/filtros aplicados]

Mode = Keep matching rows

Filter mode = Reduce matrix

Clicar em "OK"

OBS: a matriz gerada conterá apenas as proteínas diferencialmente abundantes entre os grupos.



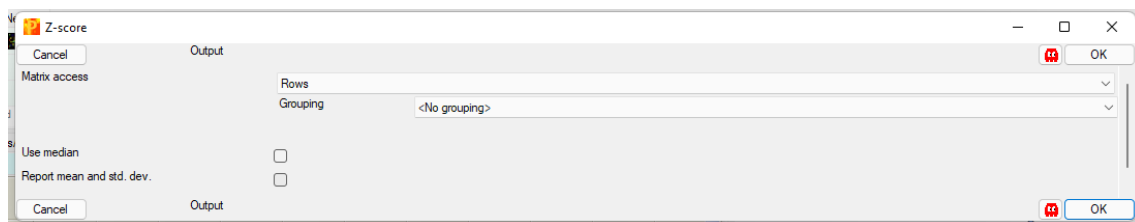
8) Normalização dos dados para construir Heatmap

a) Normalization \ Z-score

Matrix access = Rows

Grouping = <No grouping>

Clicar em “OK”

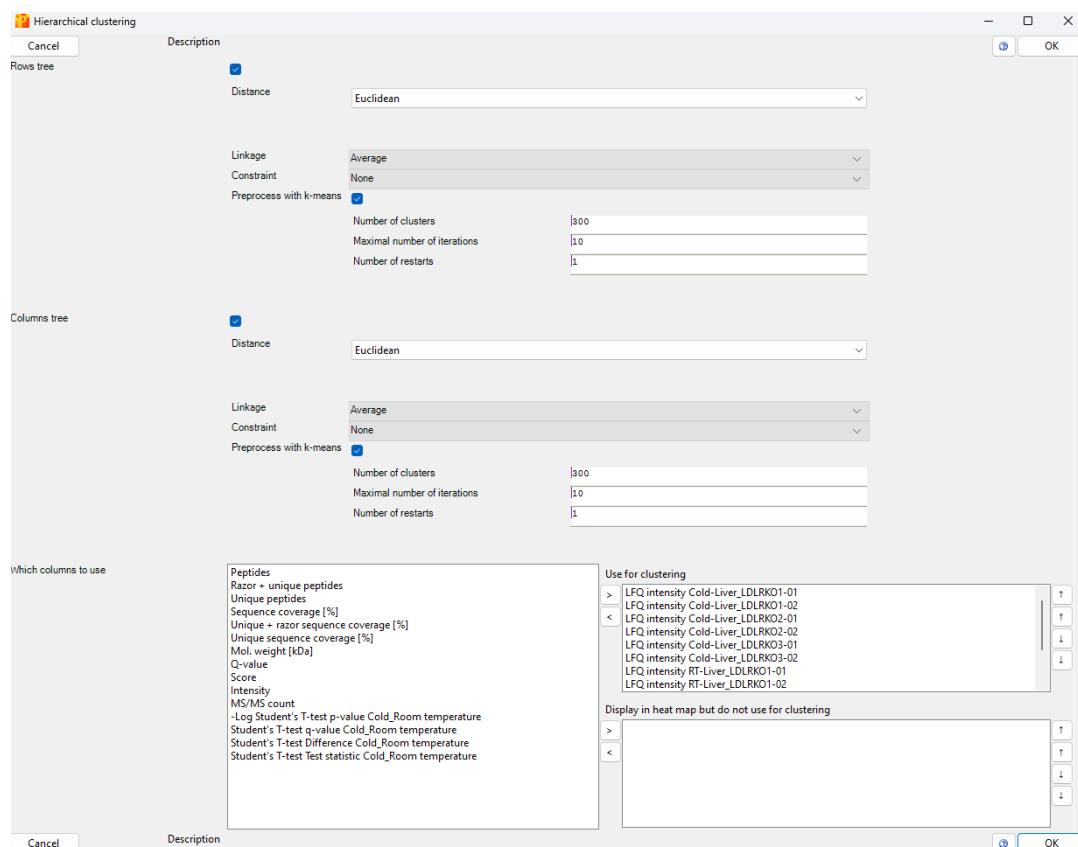


9) Construção do Heatmap [apenas com as proteínas diferencialmente abundantes]

a) Clustering/PCA \ Hierarchical clustering

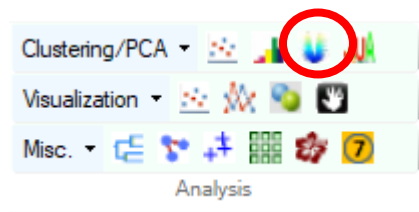
Manter os parâmetros com os valores *default*

Clicar em “OK”



10) Construção do *Volcano plot*

- OBS: utilizar a matriz obtida ao término do item “Filtragem de Valores Válidos”
- Na aba “Analysis”, selecionar ícone com imagem de um gráfico Volcano
- Alterar os parâmetros conforme abaixo:
 Grouping = Group 1
 First group (right) = TRATADO
 Second group (left) = CONTROLE
 S0 = 0.5
 Demais parâmetros, manter os valores *default*
 Clicar em “OK”



11) Análise por PCA – *Principal Component Analysis*

- Utilizar a matriz obtida após transformação dos dados em $\log_2$
- Para este tipo de análise, deve-se remover 100% dos valores não-válidos (NaN)
- Filter rows \ Filter rows based on valid values

Min. valids = Percentage

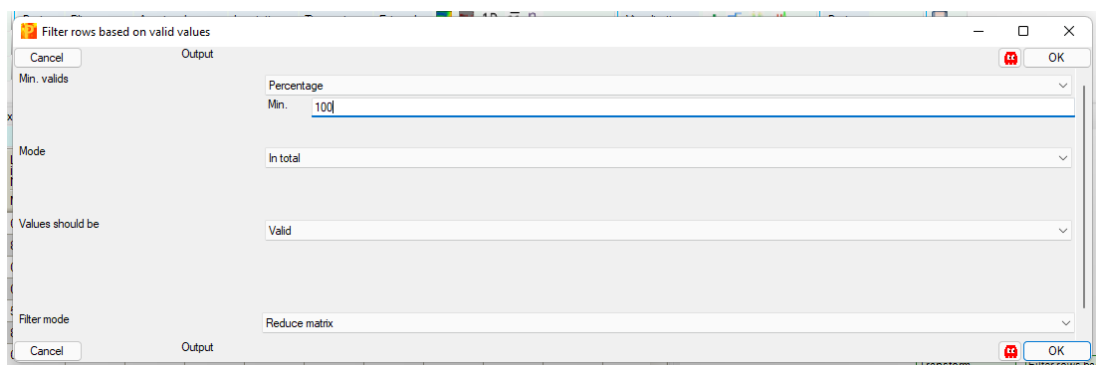
Min. = 100

Mode = In total

Values should be = valid

Filter mode = Reduce matrix

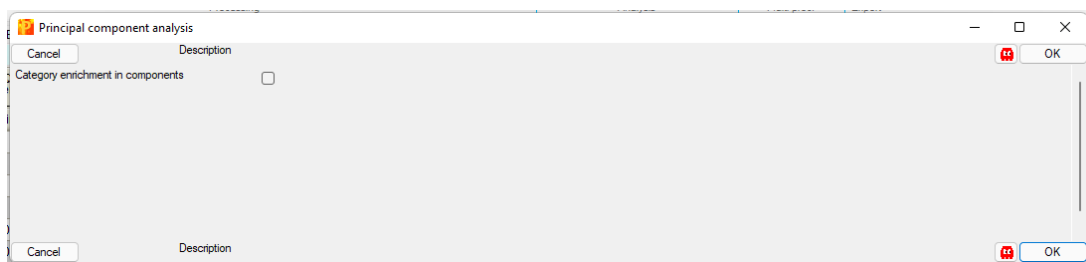
Clicar em “OK”



d) Clustering/PCA \ Principal component analysis

Utilizar os parâmetros default

Clicar em “OK”



12) Análise das proteínas exclusivas de cada grupo

a) Utilizar a matriz obtida após transformação dos dados em $\log_2$

b) Filter rows \ Filter rows based on valid values

Min. valids = 3 ⁶

Mode = “In at least group”

Grouping = Group 1

Values should be = Valid

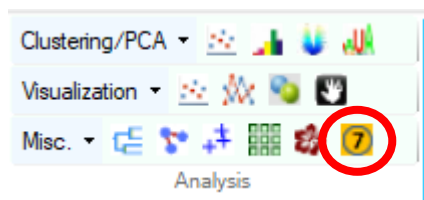
Filter mode = Reduce matrix

Clicar em “OK”

c) Para a matriz resultante, aplicar Annotate rows \ Average groups, mantendo os valores *default*. Clicar em OK

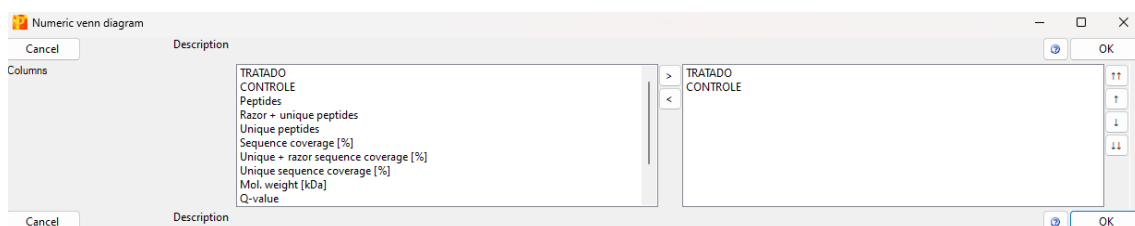


d) Com a matriz resultante selecionada, clicar no ícone “Numeric venn diagram”

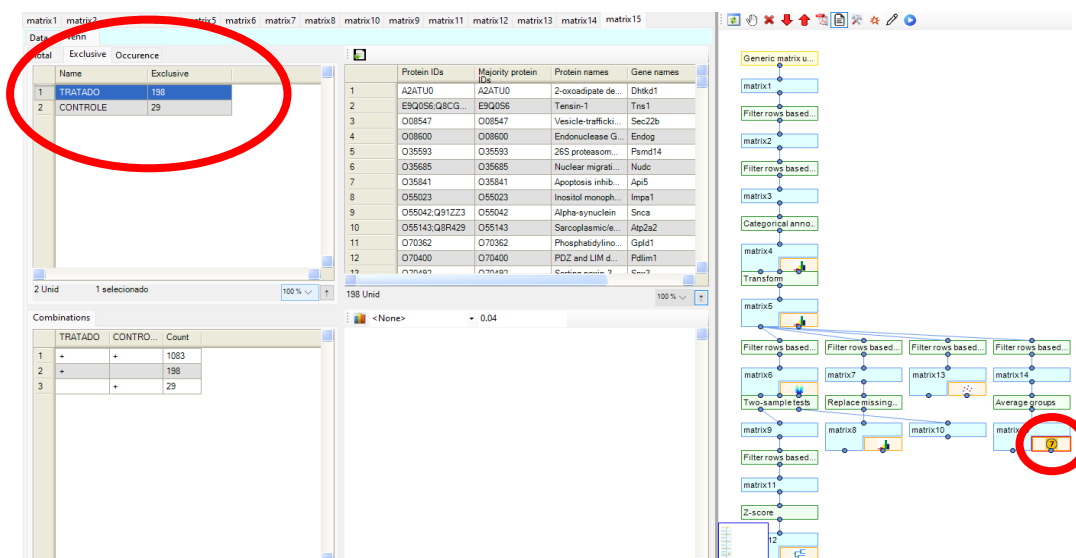


⁶ Neste exemplo, pede-se que sejam mantidas apenas as proteínas que tiveram intensidade LFQ maior que zero em pelo menos 3 das réplicas, em pelo menos um dos grupos.

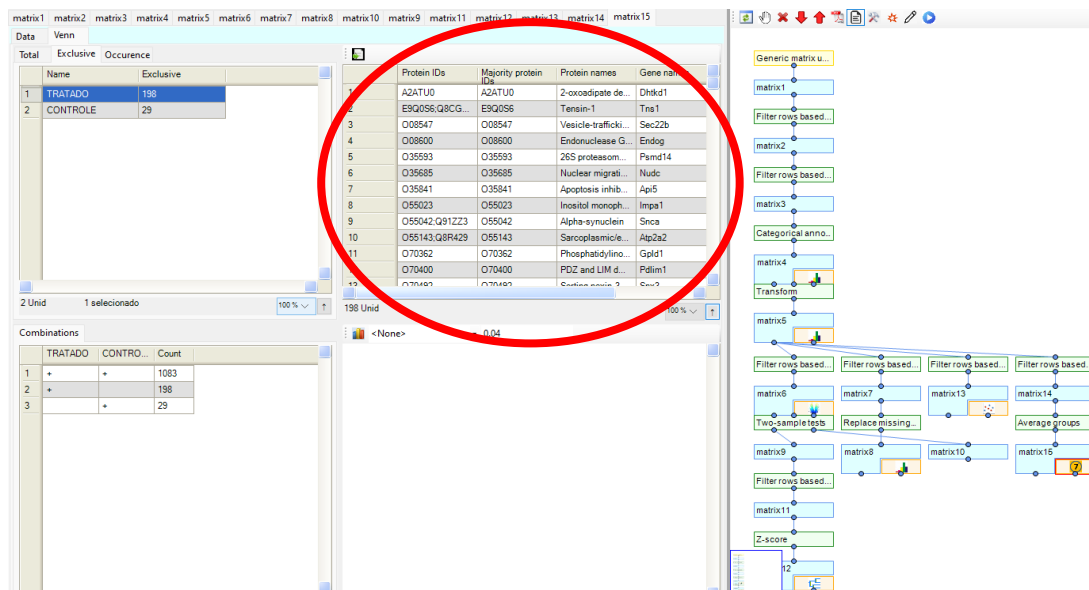
- e) Selecionar os dois grupos e transferi-los para a caixa da direita, clicando na seta para a direita
- f) Clicar em OK



- g) Clicar na imagem do resultado do diagrama de Venn (mostrado na imagem abaixo, do lado inferior direito) para visualizar, do lado esquerdo superior (aba "Exclusives"), o número de proteínas exclusivas de cada grupo, de acordo com o critério selecionado (≥ 3 identificações em um grupo, *versus* zero identificações no outro grupo).



- h) Ao clicar em cada um dos grupos, a lista de proteínas exclusivas do grupo é mostrada na janela à direita. Essa lista pode ser usada para a análise de enriquecimento (item subsequente)



13) Análise de enriquecimento (funções, compartimentos, vias metabólicas, etc)

a) Para as análises de enriquecimento, pode-se utilizar as proteínas⁷:

- de cada um dos clusters obtidos no Heatmap
- exclusivas de cada grupo
- diferencialmente abundantes, obtidas no gráfico Volcano

b) As principais plataformas disponíveis para o enriquecimento são:

- String (<https://string-db.org/>)
- UNIPROT (<https://www.uniprot.org/id-mapping>)
- WebGestalt (<https://www.webgestalt.org/>)
- Cytoscape (<https://cytoscape.org/>) [requer download]

⁷ Usar preferencialmente a coluna "majority protein IDs" da matriz para o enriquecimento.