Análise de Expressão Diferencial

Pablo Rodrigo Sanches

Departamento de Genética – FMRP/USP

psanches@usp.br

Roteiro de análise

1. DESeq2 (normalização e calculo da expressão diferencial)
   a) tratamento vs. controle
2. Obter os resultados:
   a) PCA
   b) Distância entre amostras
   c) Dispersão
   d) Histograma de p valor
   e) Genes significativamente DE
   f) Tabelas de genes significativamente DE

Normalização

- Descobrir como os genes se expressam diferencialmente em diferentes condições ou tecidos.
- Número de reads é correlacionado com o nível de expressão do gene
- RNA-Seq oferece uma aproximação quantitativa da abundância dos transcritos na forma de contagens.
- Contagens precisam ser normalizadas
  - Comprimento das diferentes moléculas de RNAs
  - Profundidade do sequenciamento de diferentes bibliotecas

Normalização

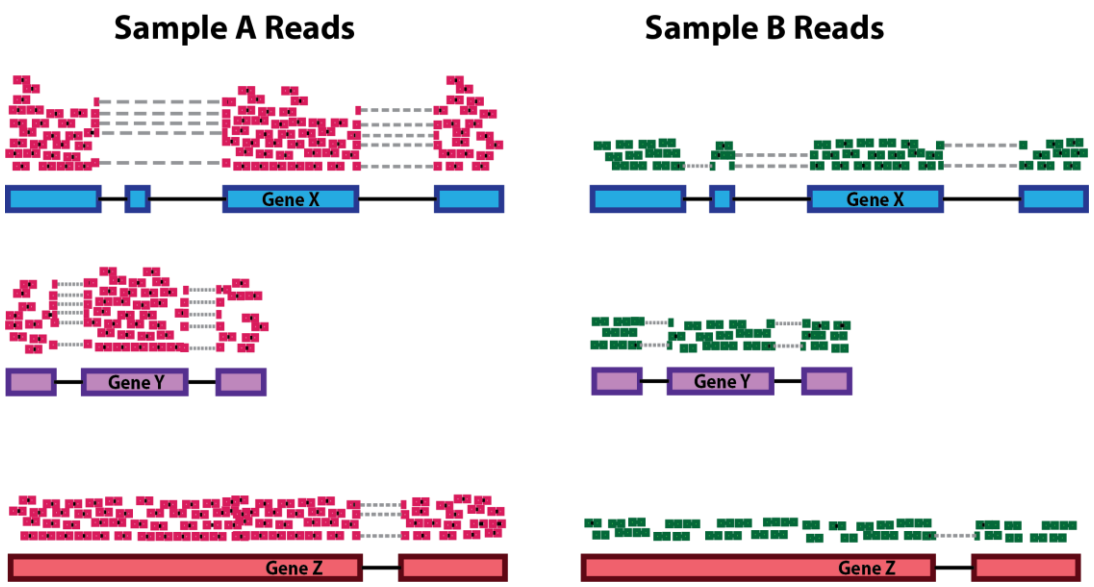
- Exemplo:
  - Biblioteca 1 → 12 milhões de reads mapeados
  - Biblioteca 2 → 16 milhões de reads mapeados

Loco	Tamanho loco (pb)	Nro. Reads Biblioteca 1	Nro. Reads Biblioteca 2
GeneA	800	24	38

É possível afirmar que temos maior expressão do GeneA na Biblioteca 2 quando comparada à Biblioteca 1 ???

Normalização

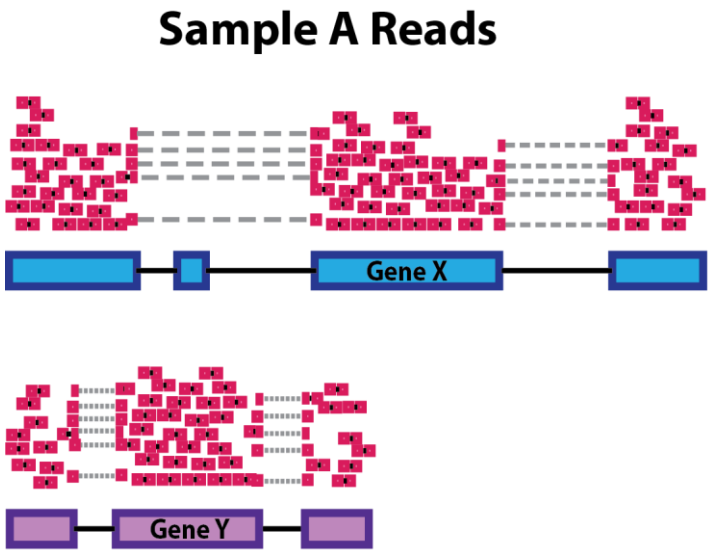
Profundidade do sequenciamento:



- No exemplo, cada gene parece ter dobrado a expressão na Amostra A em relação à Amostra B, no entanto, isso é uma consequência da Amostra A ter o dobro da profundidade de sequenciamento.

Normalização

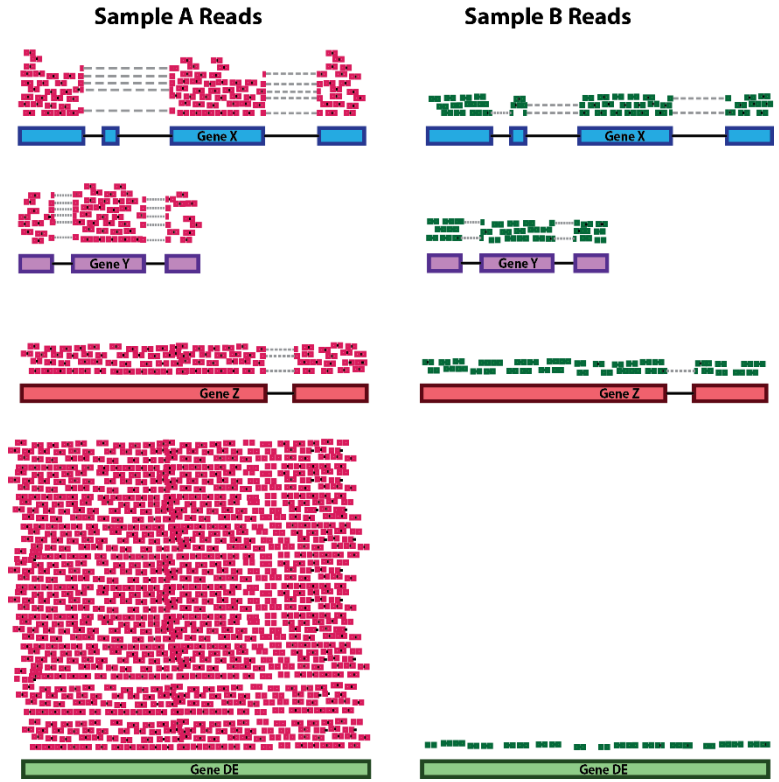
Comprimento do gene:



- No exemplo, *Gene X* e *Gene Y* têm níveis semelhantes de expressão, mas o número de leituras mapeadas para *Gene X* seria muito maior do que o número mapeado para *Gene Y* porque *Gene X* é mais longo.

Normalização

Composição de RNA:



- No exemplo, se fosse para dividir cada uma das amostras pelo número total de contagens para normalização, as contagens seriam muito enviesadas pelo Gene DE (ocupa a maior parte das contagens para uma das amostras).

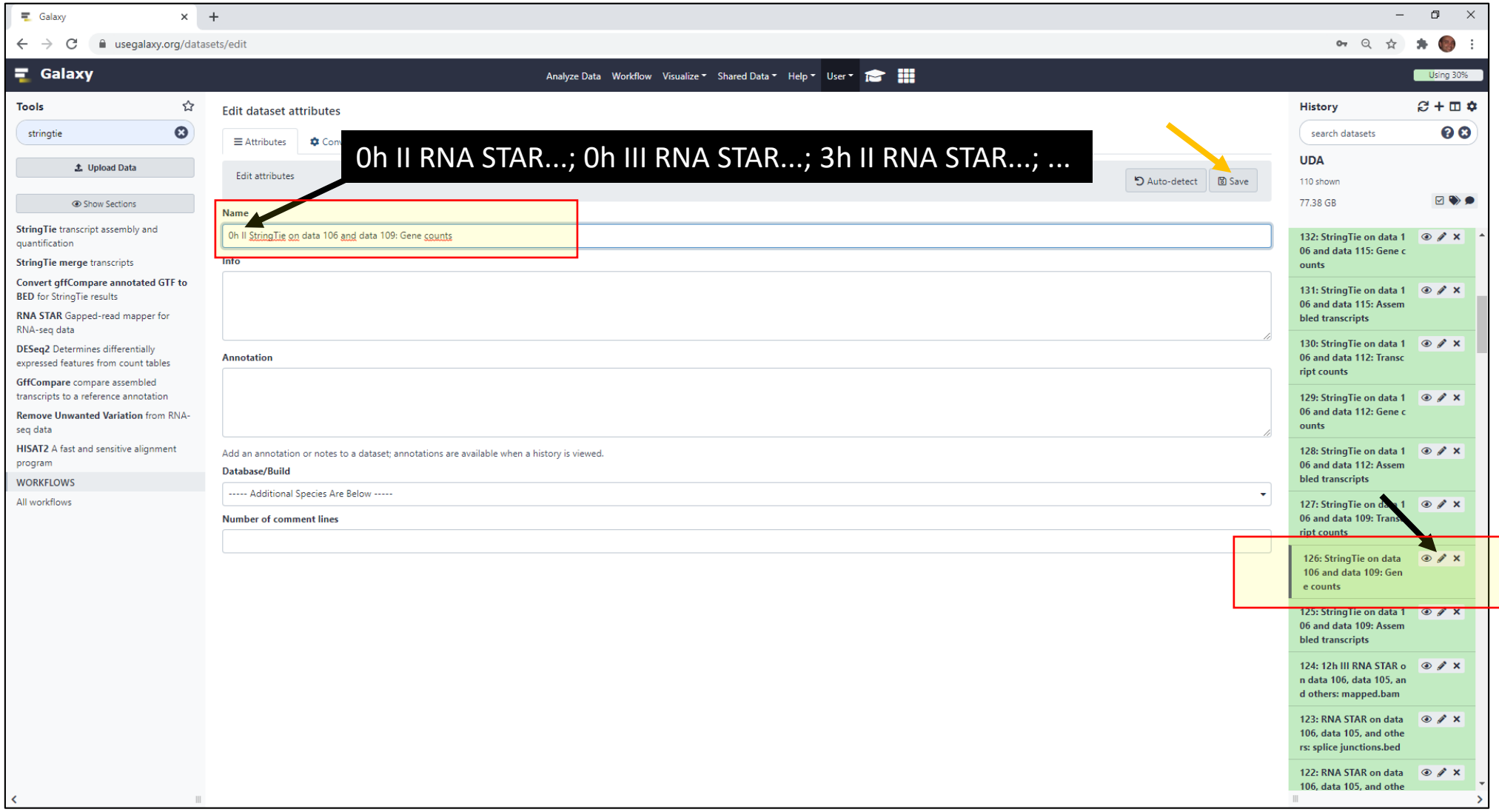
DESeq2 – Mediana das razões

- O DESeq2 executa uma normalização onde a média geométrica é calculada para cada gene em todas as amostras. A contagem de um gene em cada amostra é então dividida por essa média. A mediana dessas proporções em uma amostra é o fator de tamanho dessa amostra. Este procedimento corrige o tamanho da biblioteca e o viés da composição do RNA, que pode surgir, por exemplo, quando apenas um pequeno número de genes são altamente expressos em uma condição de experimento, mas não na outra.

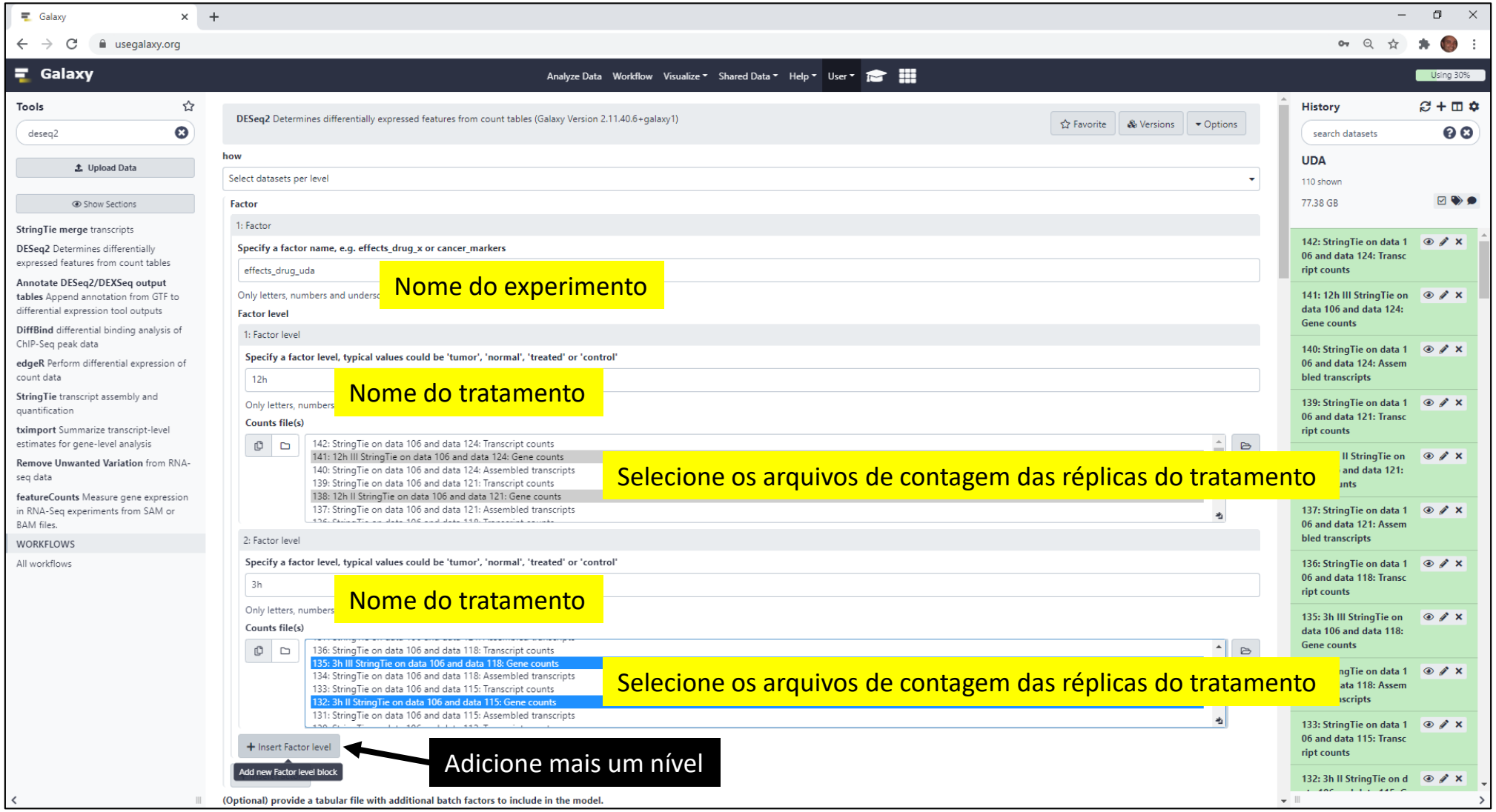
Exemplo de aplicação

	amostraA	amostraB	referência	amostraA/referência	amostraB/referência	normalizadoA	normalizadoB	FC	log2FC
geneA	2800	1000	1673,32	1,67	0,60	3429,29	816,50	0,24	-2,07
geneB	30	15	21,21	1,41	0,71	36,74	12,25	0,33	-1,58
geneC	500	750	612,37	0,82	1,22	612,37	612,37	1,00	0,00
geneD	40	80	56,57	0,71	1,41	48,99	65,32	1,33	0,42
geneE	500	1200	774,60	0,65	1,55	612,37	979,80	1,60	0,68
			Mediana	0,82	1,22				

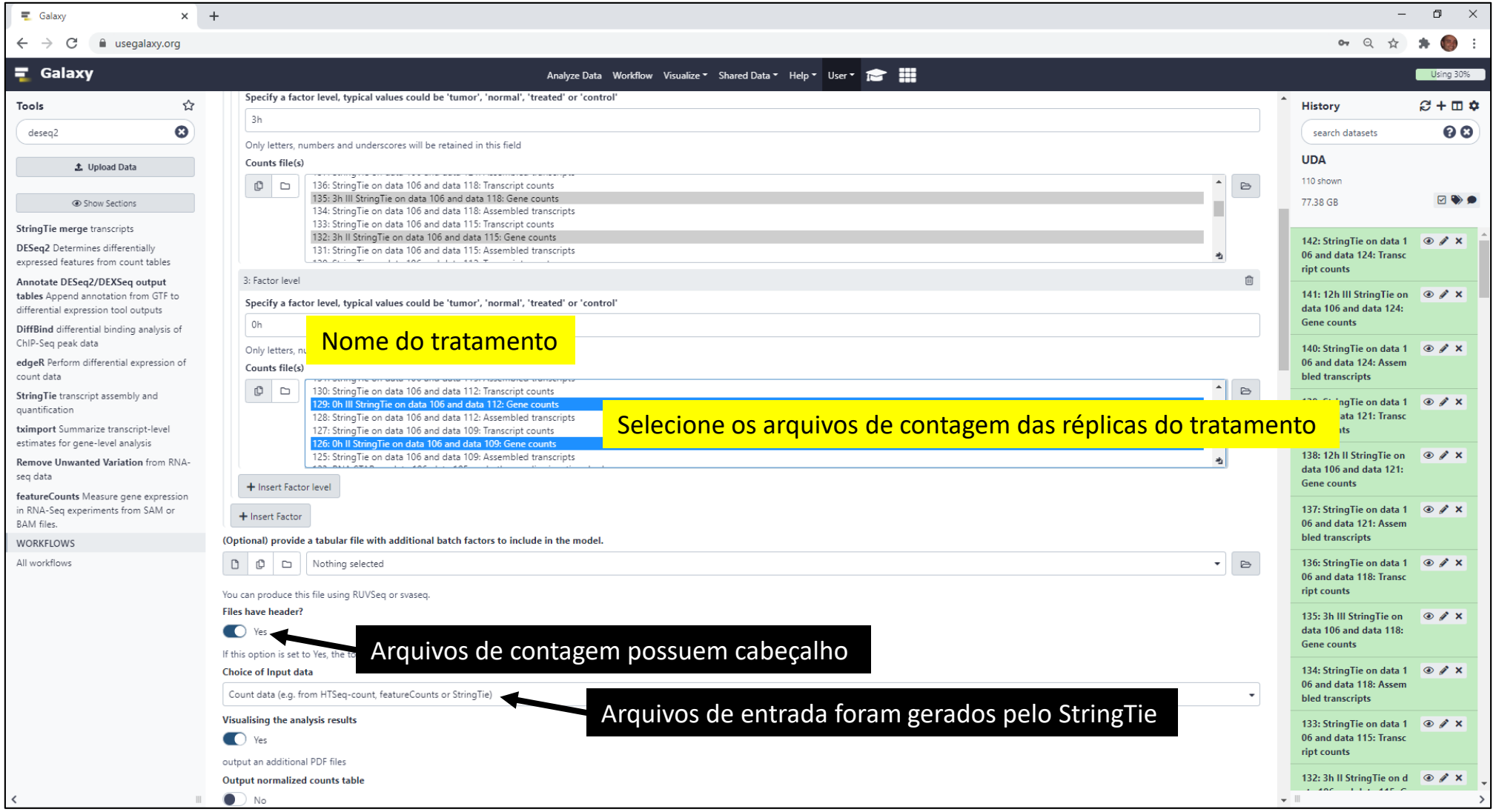
Renomear arquivos Gene counts no Galaxy (saída do software StringTie)



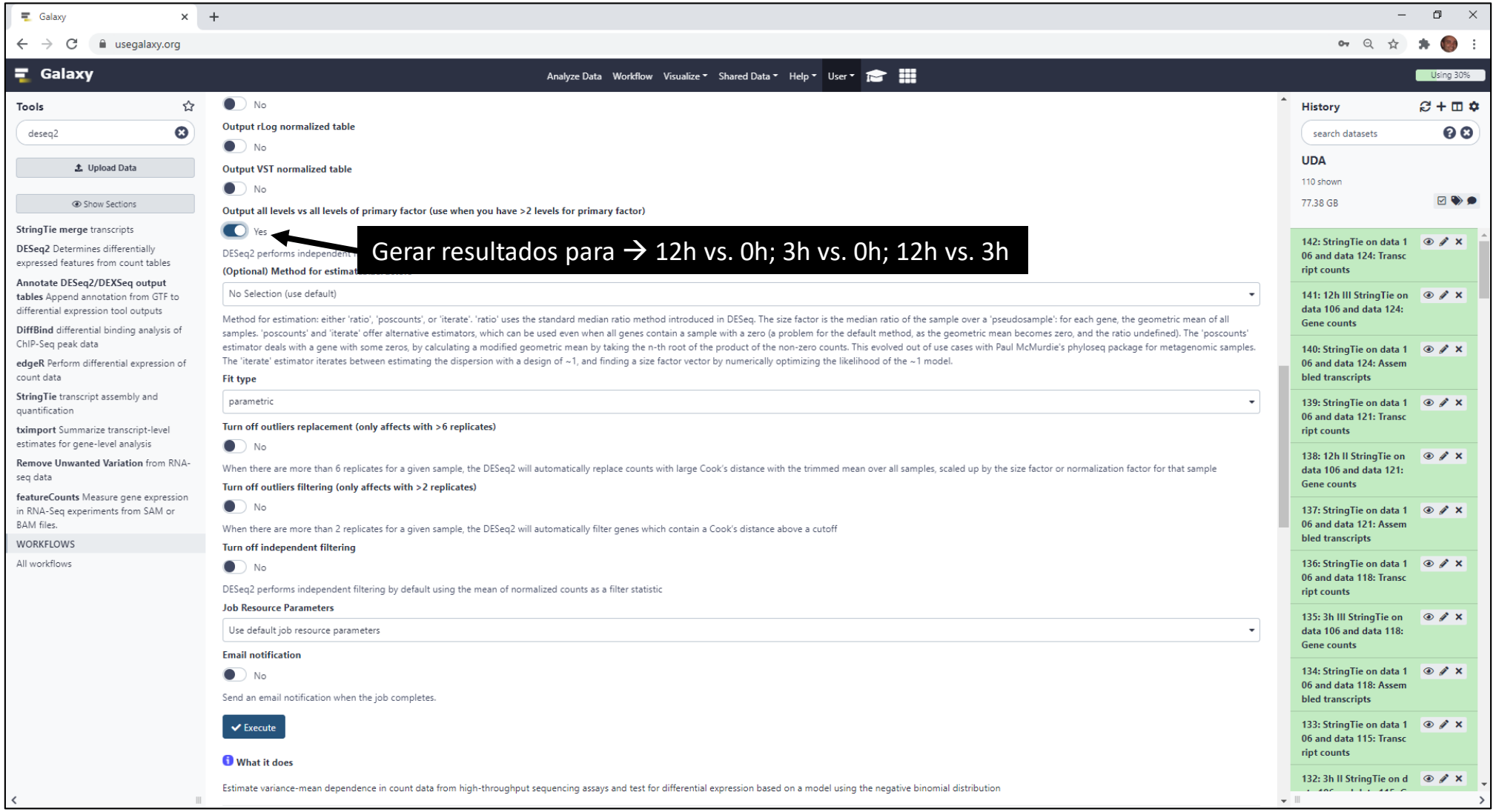
DESeq2 – Via Galaxy



DESeq2 – Via Galaxy (cont.)

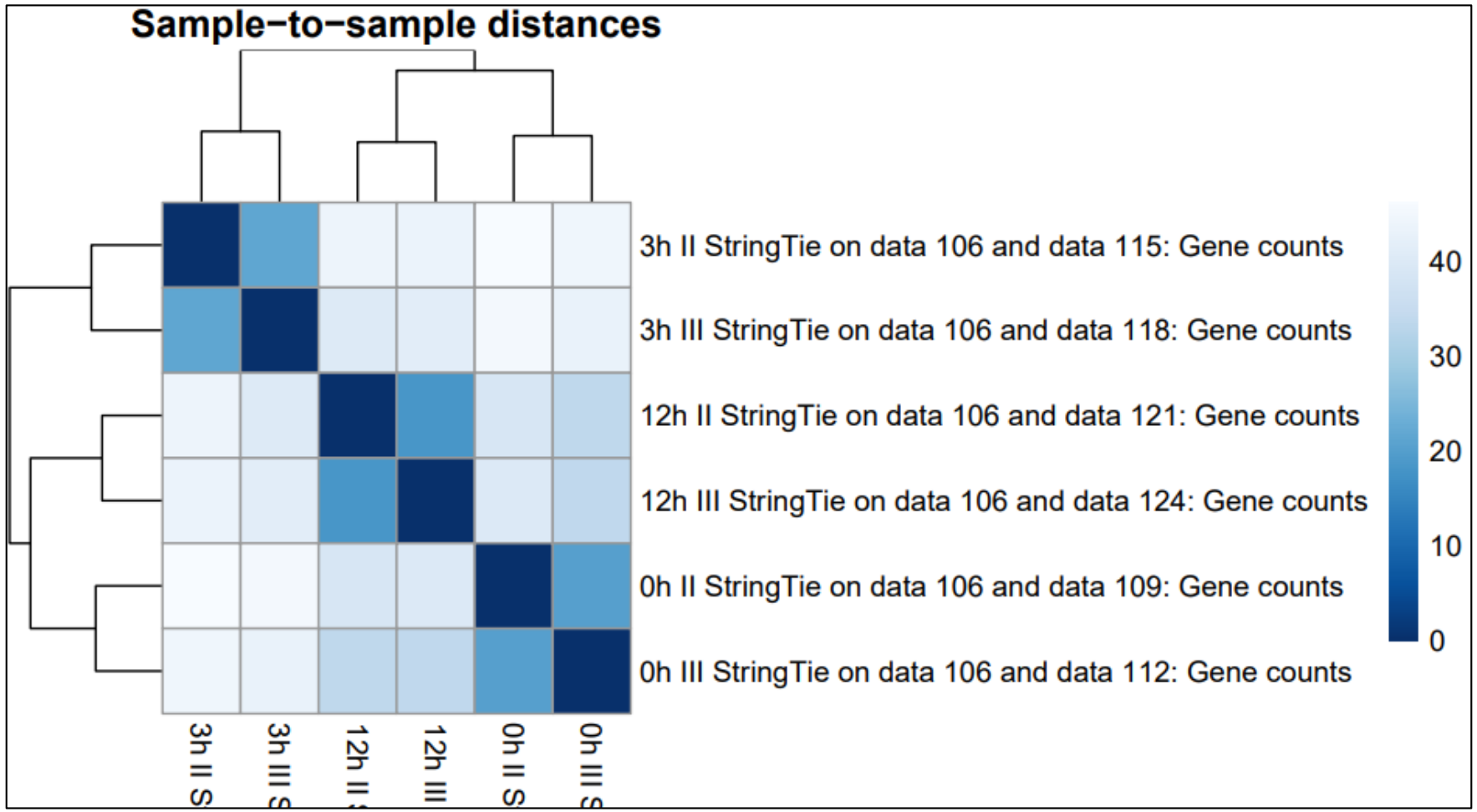


DESeq2 – Via Galaxy (cont.)

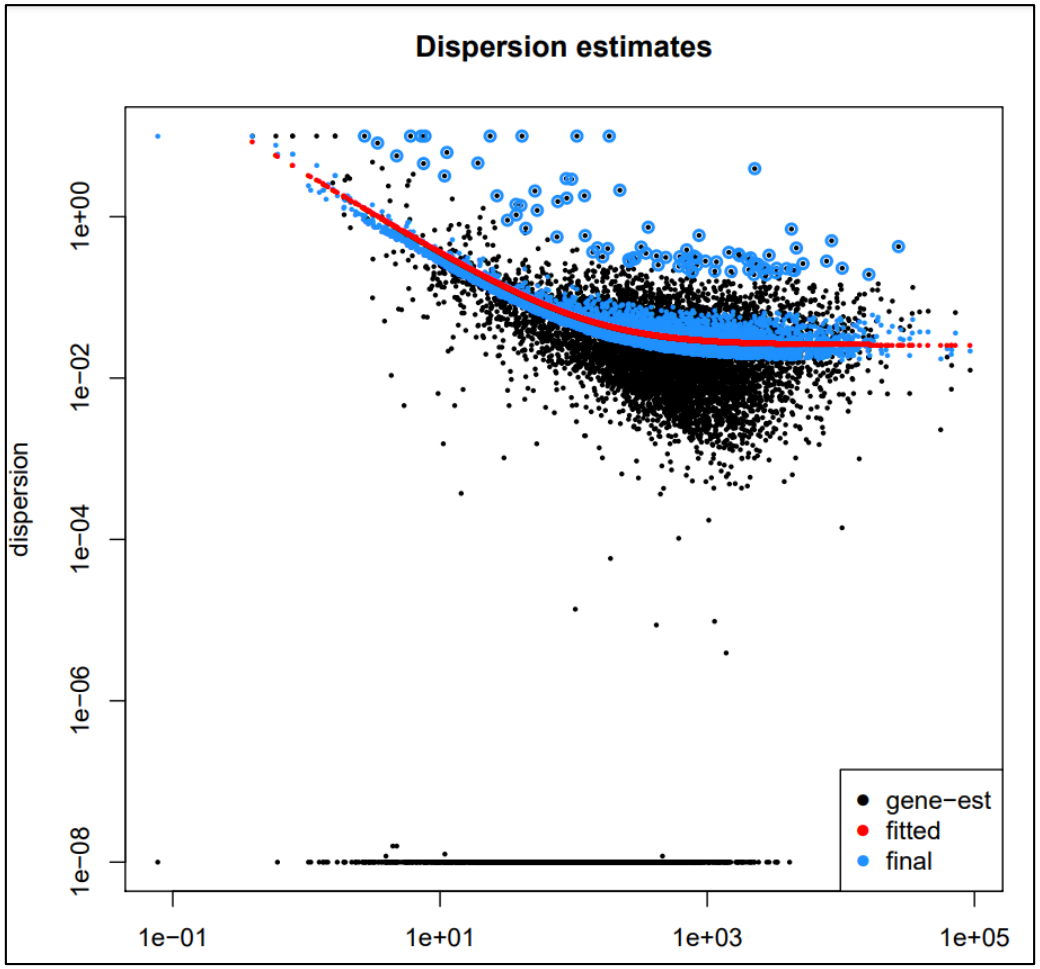


Exemplo de Resultado DESeq2 (PCA)

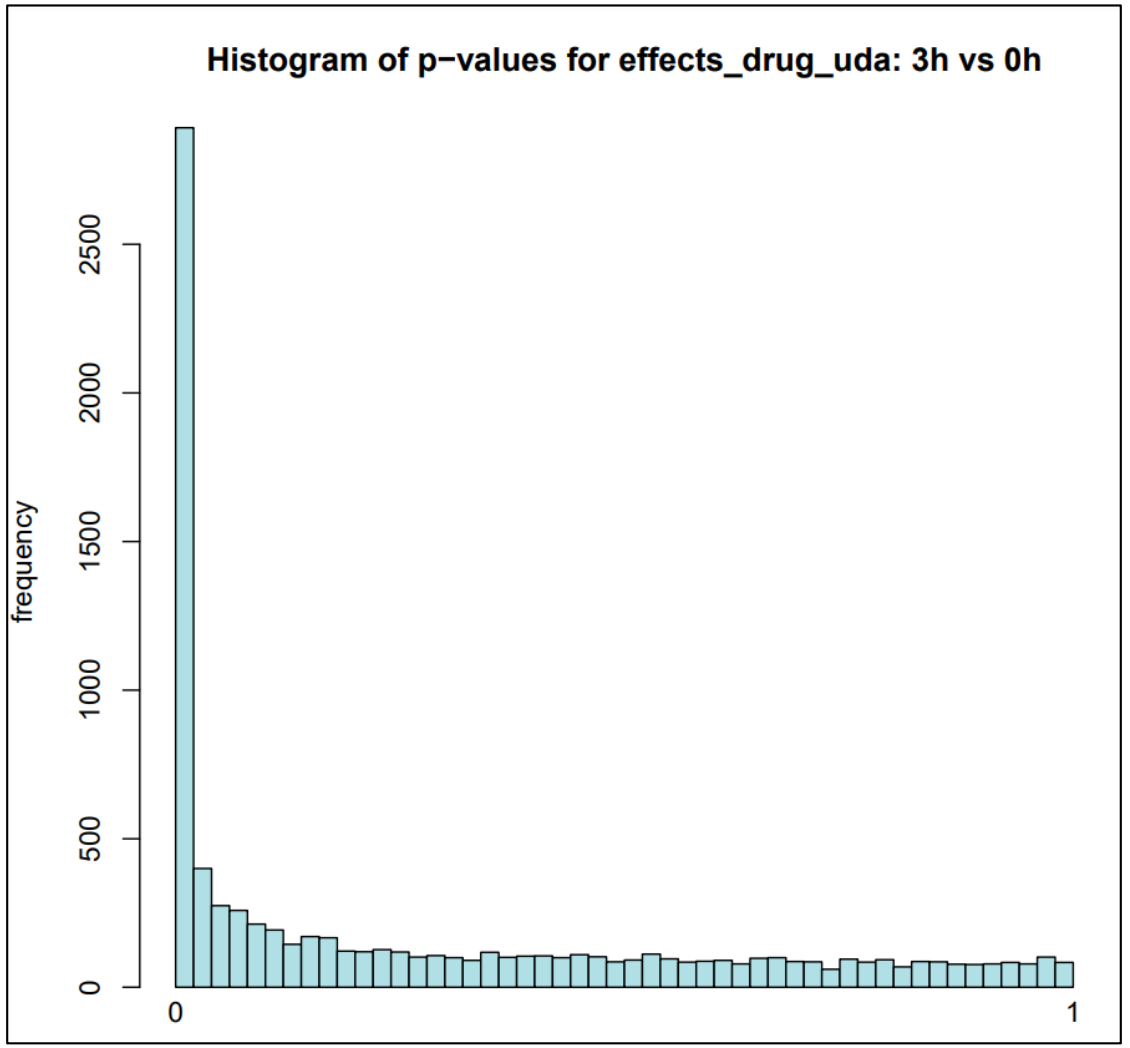
Exemplo de Resultado DESeq2 (Distância entre amostras)



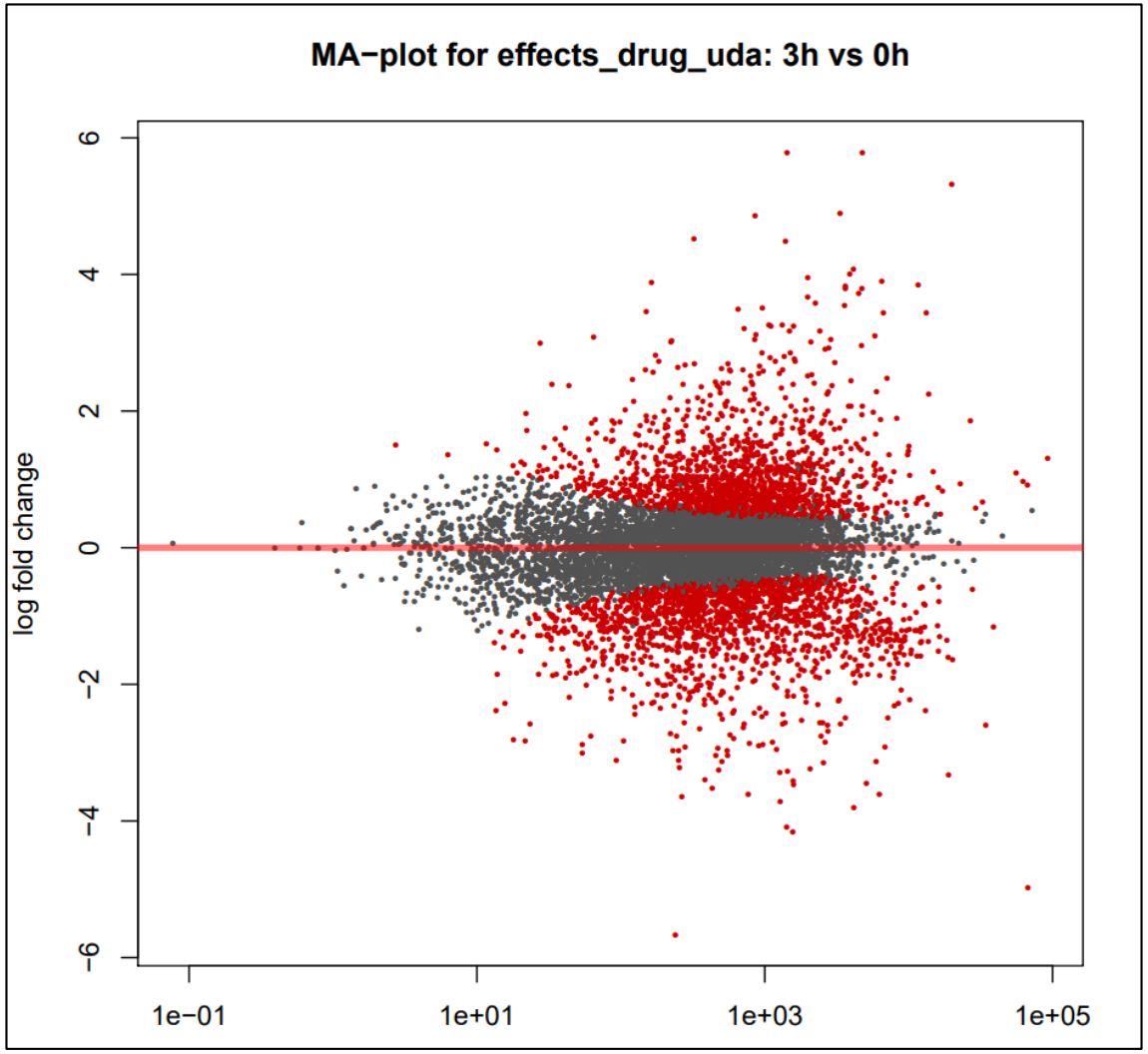
Exemplo de Resultado DESeq2 (Dispersão)



Exemplo de Resultado DESeq2 (Histograma de p valor)

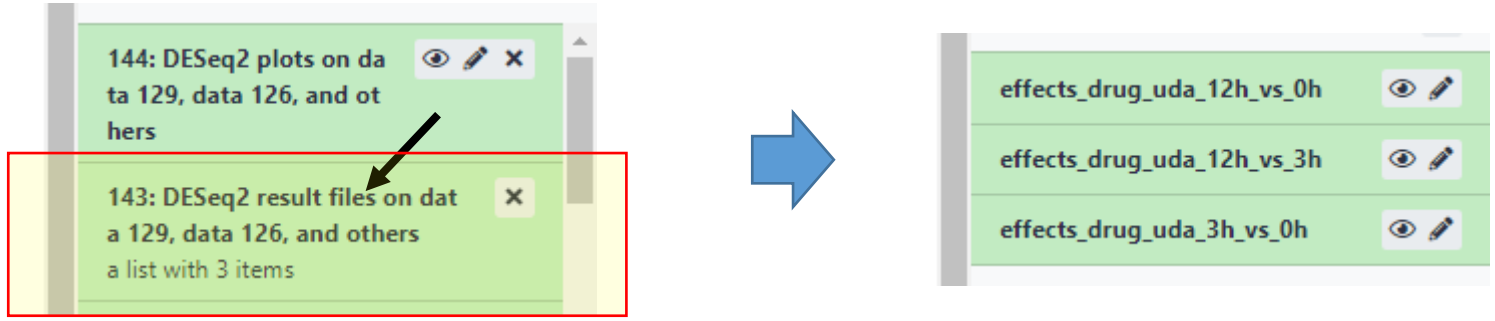


Exemplo de Resultado DESeq2 (DEG)

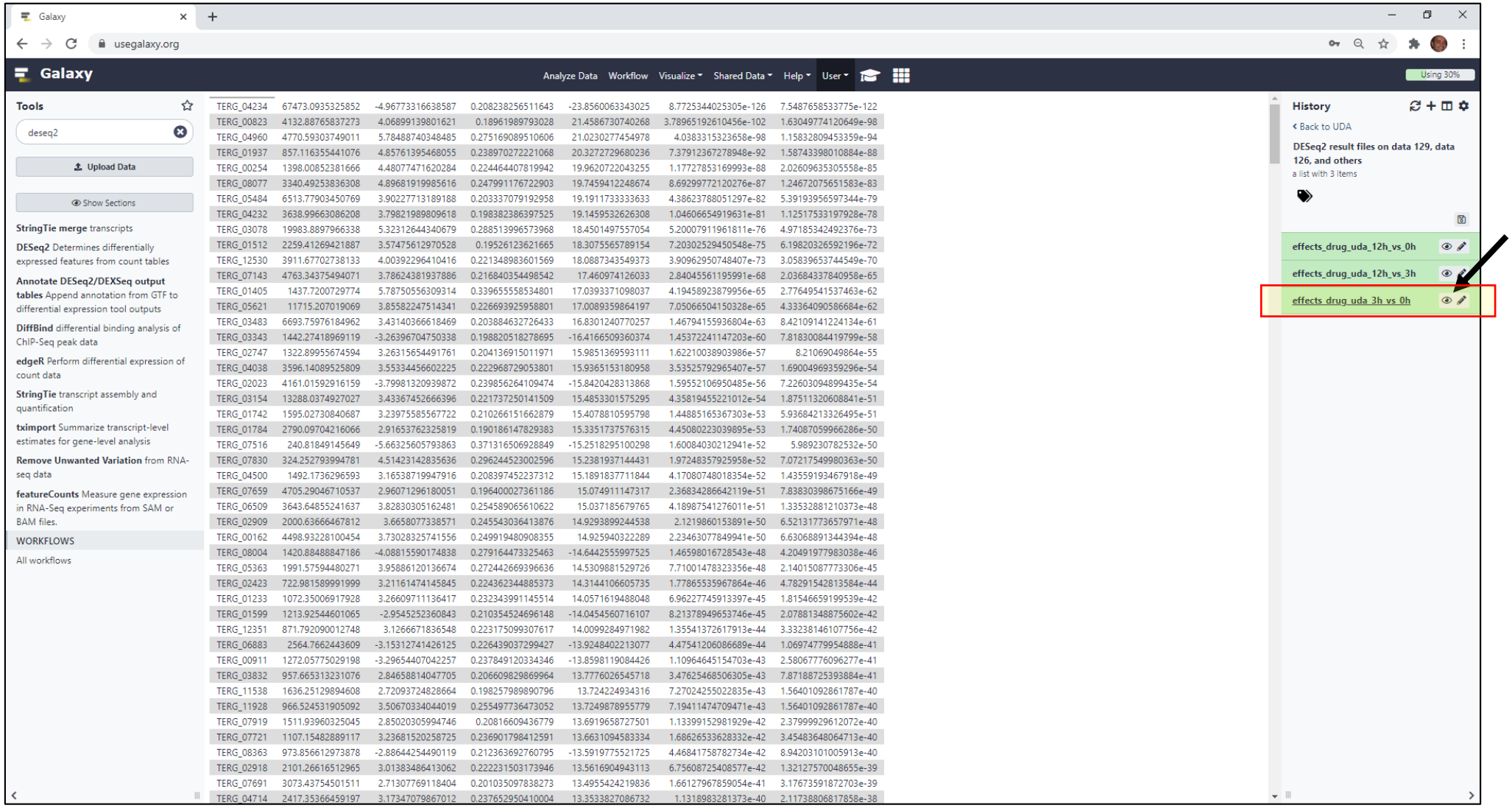


Exemplo de Resultado DESeq2 (Tabelas DEG)

- 12h vs 0h
- 12h vs 3h
- 3h vs 0h



Exemplo de Resultado DESeq2 (Tabelas DEG)



TERG_04234	67473.0935325852	-4.96773316638587	0.208238256511643	-23.8560063343025	8.7725344025305e-126	7.5487658533775e-122
TERG_00823	4132.88765837273	4.06899139801621	0.18961989793028	21.4586730740268	3.78965192610456e-102	1.63049774120649e-98
TERG_04960	4770.59303749011	5.78488740348485	0.275169089510606	21.0230277454978	4.0383315323658e-98	1.15832809453359e-94
TERG_01937	857.116355441076	4.85761395468055	0.238970272221068	20.3272729680236	7.37912367278948e-92	1.58743398010884e-88
TERG_00254	1398.00852381666	4.48077471620284	0.224464407819942	19.9620722043255	1.17727853169993e-88	2.02609635305558e-85
TERG_08077	3340.49253836308	4.89681919985616	0.247991176722903	19.7459412248674	8.69299772120276e-87	1.24672075651583e-83
TERG_05484	6513.77903450769	3.90227713189188	0.203337079192958	19.1911733333633	4.38623788051297e-82	5.39193956597344e-79
TERG_04232	3638.99663086208	3.79821989809618	0.198382386397525	19.1459532626308	1.04606654919631e-81	1.12517533197928e-78
TERG_03078	19983.8897966338	5.32312644340679	0.288513996573968	18.4501497557054	5.20007911961811e-76	4.97185342492376e-73
TERG_01512	2259.41269421887	3.57475612970528	0.19526123621665	18.3075565789154	7.20302529450548e-75	6.19820326592196e-72
TERG_12530	3911.67702738133	4.00392296410416	0.221348983601569	18.0887343549373	3.90962950748407e-73	3.05839653744549e-70
TERG_07143	4763.34375494071	3.78624381937886	0.216840354498542	17.460974126033	2.84045561195991e-68	2.03684337840958e-65
TERG_01405	1437.7200729774	5.78750556309314	0.339655558534801	17.0393371098037	4.19458923879956e-65	2.77649541537463e-62
TERG_05621	11715.207019069	3.85582247514341	0.226693925958801	17.0089359864197	7.05066504150328e-65	4.33364090586684e-62
TERG_03483	6693.75976184962	3.43140366618469	0.203884632726433	16.8301240770257	1.46794155936804e-63	8.42109141224134e-61
TERG_03343	1442.27418969119	-3.26396704750338	0.198820518278695	-16.4166509360374	1.45372241147203e-60	7.81830084419799e-58
TERG_02747	1322.89955674594	3.26315654491761	0.204136915011971	15.9851369593111	1.62210038903986e-57	8.21069049864e-55
TERG_04038	3596.14089525809	3.55334456602225	0.222968729053801	15.9365153180958	3.53525792965407e-57	1.69004969359296e-54
TERG_02023	4161.01592916159	-3.79981320939872	0.239856264109474	-15.8420428313868	1.59552106950485e-56	7.22603094899435e-54
TERG_03154	13288.0374927027	3.43367452666396	0.221737250141509	15.4853301575295	4.35819455221012e-54	1.87511320608841e-51
TERG_01742	1595.02730840687	3.23975585567722	0.210266151662879	15.4078810595798	1.44885165367303e-53	5.93684213326495e-51
TERG_01784	2790.09704216066	2.91653762325819	0.190186147829383	15.3351737576315	4.45080223039895e-53	1.74087059966286e-50
TERG_07516	240.81849145649	-5.66325605793863	0.371316506928849	-15.2518295100298	1.60084030212941e-52	5.989230782532e-50
TERG_07830	324.252793994781	4.51423142835636	0.296244523002596	15.2381937144431	1.97248357925958e-52	7.07217549980363e-50
TERG_04500	1492.1736296593	3.16538719947916	0.208397452237312	15.1891837711844	4.17080748018354e-52	1.43559193467918e-49
TERG_07659	4705.29046710537	2.96071296180051	0.196400027361186	15.074911147317	2.36834286642119e-51	7.83830398675166e-49
TERG_06509	3643.64855241637	3.82830305162481	0.254589065610622	15.037185679765	4.18987541276011e-51	1.33532881210373e-48
TERG_02909	2000.63666467812	3.6658077338571	0.245543036413876	14.9293899244538	2.1219860153891e-50	6.52131773657971e-48
TERG_00162	4498.93228100454	3.73028325741556	0.249919480908355	14.925940322289	2.23463077849941e-50	6.63068891344394e-48
TERG_08004	1420.88488847186	-4.08815590174838	0.279164473325463	-14.6442555997525	1.46598016728543e-48	4.20491977983038e-46
TERG_05363	1991.57594480271	3.95886120136674	0.272442669396636	14.5309881529726	7.71001478323356e-48	2.14015087773306e-45
TERG_02423	722.981589991999	3.21161474145845	0.224362344885373	14.3144106605735	1.77865535967864e-46	4.78291542813584e-44
TERG_01233	1072.35006917928	3.26609711136417	0.232343991145514	14.0571619488048	6.96227745913397e-45	1.81546659199539e-42
TERG_01599	1213.92544601065	-2.9545252360843	0.210354524696148	-14.0454560716107	8.21378949653746e-45	2.07881348875602e-42
TERG_12351	871.792090012748	3.1266671836548	0.223175099307617	14.0099284971982	1.35541372617913e-44	3.33238146107756e-42
TERG_06883	2564.7662443609	-3.15312741426125	0.226439037299427	-13.9248402213077	4.47541206086689e-44	1.06974779954888e-41
TERG_00911	1272.05775029198	-3.29654407042257	0.237849120334346	-13.8598119084426	1.10964645154703e-43	2.58067776096277e-41
TERG_03832	957.665313231076	2.84658814047705	0.206609829869964	13.7776026545718	3.47625468506305e-43	7.87188725393884e-41
TERG_11538	1636.25129894608	2.72093724828664	0.198257989890796	13.724224934316	7.27024255022835e-43	1.56401092861787e-40
TERG_11928	966.524531905092	3.50670334044019	0.255497736473052	13.7249878955779	7.19411474709471e-43	1.56401092861787e-40
TERG_07919	1511.93960325045	2.85020305994746	0.20816609436779	13.6919658727501	1.13399152981929e-42	2.37999929612072e-40
TERG_07721	1107.15482889117	3.23681520258725	0.236901798412591	13.6631094583334	1.68626533628332e-42	3.45483648064713e-40
TERG_08363	973.856612973878	-2.88644254490119	0.212363692760795	-13.5919775521725	4.46841758782734e-42	8.94203101005913e-40
TERG_02918	2101.26616512965	3.01383486413062	0.222231503173946	13.5616904943113	6.75608725408577e-42	1.32127570048655e-39
TERG_07691	3073.43754501511	2.71307769118404	0.201035097838273	13.4955424219836	1.66127967859054e-41	3.17673591872703e-39
TERG_04714	2417.35366459197	3.17347079867012	0.237652950410004	13.3533827086732	1.13189832813736-40	2.11738806817858e-38

Sobre as colunas da tabela de resultados DESeq2

Coluna	Descrição
1	geneid = Identificador do gene
2	baseMean = média das contagens normalizadas tomadas em todas as amostras
3	log2FoldChange = mudança log2 vezes entre os grupos. Por exemplo, valor 2 significa que a expressão aumentou 4 vezes
4	lfcSE = erro padrão da estimativa log2FoldChange
5	stat = estatística de Wald
6	pvalue = valor p do teste de Wald
7	padj = valor p ajustado de Benjamini-Hochberg (Taxa de Falsas Descobertas - FDR)

Download dos resultados