QUANTITATIVE DATA INTERPRETATION

- 1. The facility provides the quantitative data in the excel sheet with name of the protein identified or genes and its abundance value in the respective column as per sample. The most general format would look like the below mentioned table. In this table there was two group of sample (control vs Treated) in triplicate and facility had acquired the data using DIA (Data independent workflow) and reported the protein abundance values for each of the samples from different groups. If user has data in some other format, then user need to transform its data in the following format for analysis using this pipeline of analysis.
- 2. (Note: absolute abundance values required, not log transformed)

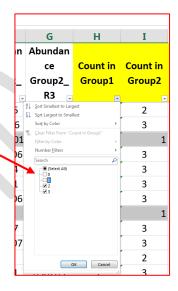
A	В	С	D	E	F	G
	Abundan	Abundan	Abundan	Abundan	Abundan	Abundan
Carros	ce	ce	ce	ce	ce	ce
Genes	Group1_	Group1_	Group1_	Group2_	Group2_	Group2_
	R1	R2	R3	R1	R2	R3
rsmF		75861.7			59158.6	75152.6
rsmG	81262.2	542694	262755	173445	266169	489787
rsmH	61023.4	120446	74140.9	54994.1	180622	230509
rsml		292793	216863		295803	256072
rsmJ		150699	102968		153065	165143
rssB				49036.3		187844
rstA	174472	1.06E+06	518260	108080	410738	803927
rsuA	225422	290118	130653	80319.9	259841	237177
rsxC	198574	442890	254043	288612		148187
rsxG	118480	72081.7	54304.6	89666.1		60048.2
ruvA	57107.1	71061.3	61080.6		110591	52039.5
ruvB						44963
sad	231802	168840	224339		355740	261521
sapA		47088.4			115030	54220.1
sbcB	114832	125472	74786.9			139984
sbcD		54945.7	43998		89298.8	103192
sbmC	144756	163906	130291	51549.4	211601	58729.5
sbp	300425	1.10E+06	886680	492698	432399	156461
sdaA	332156	677762	524046	314457	409767	1.37E+06
sdaB	461096	264817	414467	482938	420396	614893
sdhA	1.19E+06	1.11E+06	1.99E+06	231627	315826	184849
sdhB	531509	626629	1.34E+06	117018	167540	98029.5
sdhE	518351	312523	272510	254527	101190	147702

3. Filtering of the data table: It might be possible that all the replicate might not have the abundance value in every sample of the group so we always recommend that user should accept the proteins which have abundance value in at least 70% (2/3) of the samples in each group. To filter out those proteins/genes (eg. rsmF, rssB) which are not having the abundance value in replicate or groups, the numbers of data points in each of the group have to be counted to filter the genes with no replicate data points by using the COUNT function in Excel:

SL	JM T : X V	<i>f_*</i> =0	COUNT(B2:I	D2)		mgninerie			NUMBER	1811
	Α	В	С	D	E	F	G	н	I	J
	Genes	Abundan ce Group1_	Abundan ce Group1_	Abundan ce Group1_	Abundan ce Group2_	Abundan ce Group2_	Abundan ce Group2_	Count in Group1	Group2	FC (Group1/ Group2)
1	·	R1 💌	R2 💌	R3 💌	R1 💌	R2 💌	R3 💌	Τ,	Τ.	dioup -
2	aas	167712	375380	307465	230284	345511	336414	=COUNT(E	32:D2)	1.072484
3	accA	319661	279121	173606	221931	322457	379815		value1, [valu	e2],) 553

Such examples should be filtered

	Genes	Abundan ce Group1_ R1	Abundan ce Group1_ R2	ce Group1_	ce Group2_	ce Group2_	ce Group2_	Count in Group1	Count in Group2
	acpS		185495	168046		147245	160310	2	2
	acrA	87987.9	46440.6	70558	49577.7	86254.6	63247.9	3	3
acs		87849.7	35622	51822.4		19701		3	1
	acul	987995	986137	1.08E+06	812161	1.15E+06	1.45E+06	3	3
	add	859218	817237	813385	94681.4	585724	923337	3	3
	ade	65055	117547	94185.9	24582	130191	169395	3	3
	adhE	1.78E+06	1.23E+06	1.33E+06	1.43E+06	1.22E+06	1.07E+06	3	3
adhP		73996.4		55953.3	35927.6			2	1
	adiA	1.32E+06	927853	1.35E+06	1.66E+06	657797	609483	3	3
	adk	1.39E+07	2.03E+07	1.13E+07	1.14E+07	1.64E+07	1.62E+07	3	3
	agaR	42462.6	40631.6	44860.9	110522		76164.5	3	2



5. Fold change Calculation/ Differential expression of Protein (DEP): To calculate the fold change expression of the protein the protein abundance values in each groups should be averaged and then the average abundance value need to be used for differential expression of Gene. In order to calculate the DEP use average abundance value of one protein in each group and divide the average abundance value from the remaining group of samples as mentioned below.

SUM	sum • : × • fr =AVERAGE(E2:G2)/AVERAGE(B2:D2)											
	А	В	С	D	E	F	G	Н	Ι	J		
		Abundan	Abundan	Abundan	Abundan	Abundan	Abundan			FC		
	Genes	се	ce	ce	ce	ce	ce	Count in	Count in	(Group1/		
	Genes	Group1_	Group1_	Group1_	Group2_	Group2_	Group2_	Group1	Group2	Group2)		
1		- R1 -	R2 💌	R3 💌	R1 💌	R2 -	R3 🚽	Л	.Τ.			
2	aas	167712	375380	307465	230284	345511	336414	3	3	=AVERAGE		
3	accA	319661	279121	173606	221931	322457	379815	3	3	1.19655		
4	accB	6.97E+06	4.93E+06	6.11E+06	1.23E+07	6.16E+06	4.88E+06	3	3	1.29774		
5	accC	2.31E+06	2.51E+06	2.02E+06	344377	2.55E+06	2.68E+06	3	3	0.81465		

4.

6. **Statistical Significance (P value) calculation using student t-test:** P-values were calculated using T-Test after selecting the replicate in the group and then using the formula in the formula tab.

SUM	▼ : X ✓	f _x	=TTES	бт(в2	:D2,E	2:G2,	2,3)					
	А	В	С	D	Е	F	G	Н	Ι	J	К	L
1	Genes	ce Group1_	Abundan ce Group1_ R2 🔻	ce	Abundan ce Group2_ R1	Abundan ce Group2_ R2 💌	Abundan ce Group2_ R3 🔻	Count in Group1		FC (Group1/ Group ²⁾	T.Test (Pvalue)	Log2(FC)
2	aas	167712	375380	307465	230284	345511	336414	3	3	1.072484	=TTEST(B2:D2,E2:G2,2,3)	0.100956
3	accA	319661	279121	173606	221931	322457	379815	3	3	1.196553	TTEST(array1, array2, tails, type)	0.258884
л	2000	6 075106	1 035106	6 14 106	1 325±07	6 165-06	1 00ETUR	2	2	1 2077/1	0 50007767	0 276002

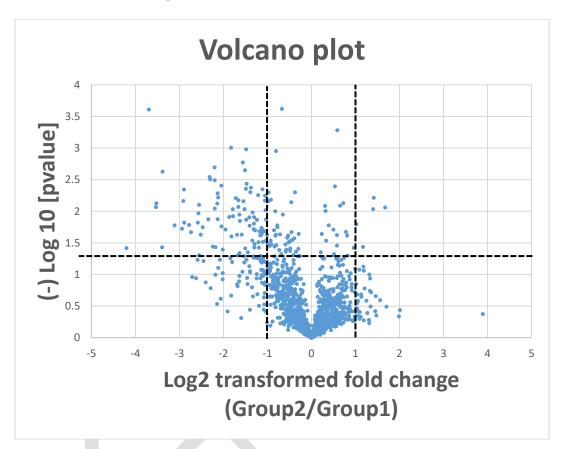
- Array1 The first group data set.
- Array2– The second group data set.
- Tails Specifies if it is a one-tailed or two-tailed test.
 - If tails = 1, T-TEST uses the one-tailed distribution.
 - If tails = 2, T-TEST uses the two-tailed distribution.
- **Type** The type of t-test to perform:
 - Type 1: Performs a paired t-test
 - Type 2: Two-sample equal variance t-test Unpaired.
 - Type 3: Two-sample unequal variance t-test Unpaired.
- 7. Log transformation of differential expression of proteins and p-value Log transformation of fold change and p-values were done to the base of 2 and 10 respectively. [Note: for p value (negative logarithm to base 10 is to be calculated)]. The above log transformed values were used for the generation of Volcano Plot as mentioned below. P value cut-off of less than 0.05 was used to determine the significant proteins and log 2 fold change of ≥1 was considered as up-regulated proteins whereas log2 fold change of ≤ -1 is considered as down regulated.

						-								
SUM	▼ : X v	' f _x	=LOG	(J2,2)										
	А	В	С	D	E	F	G	Н	I	J	К		L	м
1	Genes	Abunda ce Group1	ce Group1_	ce Group1_	ce Group2_		ce Group2_	Count in Group1	Count in Group2	FC (Group1/ Group2)	T.Test (Pvalue)	Lo T	g2(FC)	Log10- Pvalue v
2	aas	167712	375380	307465	230284	345511	336414	3	3	1.072484	0.790810545	=LC	DG(J2,2)	
3														
	accA	31966:	279121	173606	221931	322457	379815	3	3	1.196553	0.469835816	L	OG(numb	er, [base],
SUM	ACCA	f _x	- LOG1 c	0(К2) D	E	F	G F		3	1.196553	0.469835816 К	L	OG(numb	er, [base])
SUM	• : X •	f _x = B Abundan	-LOG1 c Abundan 4	O(K2) D Jundan Al	E bundan Ab	F undan Abu	G F	4 1	J			L	M	N
	↓ × ✓ A Genes	fx B Abundan ce Group1_	C C Abundan Ce Group1_	D bundan Al ce 5roup1_ G	E bundan Ab ce roup2_ Gri	F rundan Abu ce Group2_ Gro	G F Indan ce Cour bup2_ Gro	H I nt in Cour up1 Grou	tin (Grou	с 1 р1/	K T.Test (Pvalue)	L Log2(FC)	M Log10- Pvalue	N Signific nt
1	▼ : × ✓ A Genes	∫x B Abundan ce Group1_ R1 ▼	C Abundan Ce Group1_ R2 v	D bundan Al ce Group1_ G R3 v	E bundan Ab ce roup2_ Gr R1 💌	F rundan Abu ce Gro roup2_ Gro R2 ¥	G F Indan Ce Cour Sup2_ Gro R3 T	H I ntin Cour up1 Grow	tin F(p2 Grou T Grou	ն 191/ 19 ³¹	K T.Test (Pvalue)	L Log2(FC)	M Log10- Pvalue	N Significa nt
	↓ × ✓ A Genes	fx B Abundan ce Group1_	C Abundan ce Group1_0 R2 v 375380	D bundan Al ce Group1_ G R3 v 307465 2	E bundan Ab ce roup2_ Gr R1 230284 34	F uundan Abu ce G oup2_Gro R2 V I 45511 33	G F Indan ce Cour bup2_ Gro	H I Intin Cour up1 Grou 3 3 3	tin F0 ip2 Grou J.072	C 191/ 1922 2484	K T.Test (Pvalue)	L Log2(FC) =-1	M Log10- Pvalue	N Signific nt

Volcano Plot for Differentially Expressed Proteins

X-axis: Log2 (Fold Change) and Y-axis: (-) Log10 [P-values] are needed for making a volcano plot.

P value cut-off of less than 0.05 was used to determine the significant proteins and log 2 fold change of ≥ 1 was considered as up-regulated proteins whereas log2 fold change of ≤ -1 is considered as down regulated.



Horizontal line: pvalue cut-off \leq 0.05, Vertical lines: Log 2 fold change cut-off \geq 1 or \leq -1 are added manually from shapes

8. The **significance** was concluded from p-value.

	apovana a	1.011				Significant			1101112-01	-			Jugico		
SUM	• : × √	f _x	=IF(K2	2<=0.(05,"TI	RUE",	"FALS	SE")							
	A	в	С	D	E	F	G	н	I	J	К	L	м	N	0
		Abunda ce	n Abundan ce	Abundan ce	Abundan ce	Abundan ce	Abundan ce		Count in	FC			Log10-		
	Genes	Group1	_ Group1_	Group1_	Group2_	Group2_	Group2_	Group1	Group2	(Group1/	T.Test (Pvalue)	Log2(FC)	Pvalue	Significant	OverExpressed
1	*	r R1	• R2 •	R3 🔻	R1 💌	R2 🔻	R3 💌	Τ.	τ,	Group		v v			¥ ¥
2	aas	167712	2 375380	307465	230284	345511	336414	3	3	1.072484	0.790810545	0.100956	=IF	(K2<=0.05,"TRUE","F	ALSE") No
3	accA	31966	L 279121	173606	221931	322457	379815	2	3	1.196553	0.469835816	0.258884	0 32805	(logical test lvalue if	true], [value_if_false])

If p-value<= 0.05, SIGNIFICANT (TRUE) If p-value >0.05, NON-SIGNIFICANT (FALSE)

Similar function can also be used for the UP and Down for up regulated and down regulated proteins, if the same formula is used on log2 (FC) column and cut off for up regulated and down regulated should be used as ≥ 1 or ≤ -1 for up regulated and down regulated proteins respectively then separate list of the DEP (UP and Down) can be generated for the GO (Gene Ontology) GO and Heat map analysis may be performed on DEP using Shiny GO and Morpheus.

Resources

- ShinyGO For Gene Ontology and enrichment analysis <u>http://bioinformatics.sdstate.edu/go/</u>
- 2. Morpheus For heat map and clustering analysis https://software.broadinstitute.org/morpheus/
- 3. BoxPlotr For boxplot representation of data http://shiny.chemgrid.org/boxplotr/
- 4. ClustVis For principal component analysis <u>https://biit.cs.ut.ee/clustvis/</u>
- 5. VolcaNoseR For volcano plot https://huygens.science.uva.nl/VolcaNoseR2/
- 6. ggVolcanoR for volcano and upset plot https://ggvolcanor.erc.monash.edu/
- 7. Venny 2.0 for Venn Diagrams https://csbg.cnb.csic.es/BioinfoGP/venny.html