MacVector 17

for Mac OS X

RNASeq Gene Expression Analysis Tutorial

Mailector Inc

Software for Scientists

Copyright statement

Copyright MacVector, Inc, 2019. All rights reserved.

This document contains proprietary information of **MacVector**, **Inc** and its licensors. It is their exclusive property. It may not be reproduced or transmitted, in whole or in part, without written agreement from **MacVector**, **Inc**.

The software described in this document is furnished under a license agreement, a copy of which is packaged with the software. The software may not be used or copied except as provided in the license agreement.

MacVector, Inc reserves the right to make changes, without notice, both to this publication and to the product it describes. Information concerning products not manufactured or distributed by **MacVector, Inc** is provided without warranty or representation of any kind, and **MacVector, Inc** will not be liable for any damages.

This version of the RNASeq Analysis tutorial was published in January 2019.

Contents

CONTENTS	3
INTRODUCTION	4
SAMPLE FILES	4
TUTORIAL	4
Create and Populate an Assembly Project	4
Run a Bowtie Assembly	6
Analyzing the Bowtie Results	8
Exporting Data to Excel	12

Introduction

One common use of Next Generation Sequencing (NGS) technology is to analyze the relative expression levels of all known genes of an organism in a single experiment. mRNA is extracted from the organism and randomly sequenced using NGS to generate millions of reads. These can then be aligned against a sequenced genome to determine how many reads align to each known gene, resulting in data that can be used to estimate the relative expression levels of each gene.

MacVector incorporates the popular Bowtie algorithm which is a blazingly fast assembler than can align millions of reads to a reference genome in just a few minutes with minimal memory requirements. This tutorial shows you how to perform one of these analyses with sample data that is installed along with MacVector.

Sample Files

You can find the data files for this tutorial in the following location;

/Applications/MacVector/Tutorial Files/Contig
Assembly/RNASeq/

The data can also be downloaded from;

https://macvector.com/downloads.html

Tutorial

Create and Populate an Assembly Project

You must have the *Assembler* module enabled for this tutorial. You can check if you have an active *Assembler* license by choosing **MacVector I About MacVector**. You should see a "splash screen" something like this;



If the logo simply reads "*MacVector*" and not "*MacVector* with *Assembler*" then you do not have a current license to run *Assembler* and you should contact us at <u>support@macvector.com</u> to obtain a temporary license.

Select File I New I Assembly Project to create an empty Assembly Project document. Click on the Add Ref toolbar button, navigate to the /Applications/MacVector/Tutorial Files/Contig Assembly/RNASeq/folder and choose Campylobacter jejuni IA 3902. Then click on the Add Reads toolbar button and choose the file RNASeq.fastq.gz from the same folder.

Your project should look like this;



Note that the Status column indicates the type of data contained in each imported file with *REF* indicating a reference sequence. By default, imported fastq files are assign as *Illumina* files (MacVector will also automatically identify and flag paired-end read files). If your data comes from a different source (IonTorrent, PacBio, Oxford Nanopore etc) you can simply double-click on the row to change the data source.

Also note that MacVector can directly read gzipped files (these typically have a .gz extension). There is no need to unzip the files prior to analysis. This can save a lot of disk space with large data sets.

Double-click on the *Campylobacter jejuni* item to open up a sequence document window. Switch to the Features tab.

•••		Camp	ylobacter_je	juni_IA3902 -	— Features				
2]• ~			Q Description			
DNA	Locked Text V	liew Prefs Rep	lica Topolog	y Create Edit	Delete Join	Filter			
Edit	or	Мар	Features	Annotatio	ins				
Туре	Start	∧ Stop	C Desc	ription					
			/tran /tran	sl_table=11 slation=MQEN`	YGASNIKVLKGLE	AVRKRPGMYIGDTNIG			
gene	2579	4888	/gen /locu	e=gyrB s_tag=CJSA_0	003				
CDS	4916	004 periplasmic protei 359.1 ILALFLSASWAQN	n ILEINPDTGLIIDPDSPLV						
gene	4916	5257	C /locu	s_tag=CJSA_0	004				
CDS	5260	6498	C /code /locu /proc /prot /tran /tran	on_start=1 s_tag=CJSA_0 luct=molybdop ein_id=ADC276 sl_table=11 slation=MKQNI	005 oterin oxidoreduct 360.1 DQKENRRDFLKNI	tase family protein GLGLFGISVLSNFSFEN			
gene	5260	6498	C /locu	s_tag=CJSA_0	005				
CDS	6709	8010	/code /locu /proc /prot /tran /tran	/codon_start=1 /locus_tag=CJSA_0006 /product=Na+/H+ antiporter family protein /protein_id=ADC27661.1 /transl_table=11 /translation=MTLLTNPIIISVVLMTLLCLFRFNVLLSLLISAL					
gene	6709	8010	/locu	s_tag=CJSA_0	006				
CDS	8144	12634	/code /gene /locu /proc	on_start=1 e=gltB s_tag=CJSA_00 luct=glutamate	007 synthase (NADP	H) large subunit			

Note how the reference sequence is annotated with **CDS** and **gene** features. This will be important later when we need to calculate the number of RNASeq reads that align to each feature.

Run a Bowtie Assembly

Close the sequence window and return to the Assembly Project. Select both items in the list (hold down the <shift> key to select the second item), then click the Bowtie toolbar button.

Bowtie Options	Read pre-processing
Preset: Sensitive Type of Alignment: Local No gaps within first bases of read: Number of Threads: 4 Threads	 Discard reads less than Trim ends with quality less than Trim N's from ends Discard short reads that contain any N's
Use paired-end alignments Minimum insert size: Maximum insert size: 400 Orientation: Forward - Reverse	Generate child contigs Check this box if you are using the Reference as a scaffold to assemble related reads, or if you want to 'drill down" into individual alignments. For other tasks (e.g. SNP analysis or RNA-Seq expression analysis) leave this unchecked.
	Defaults Cancel OK

This is where you can change the parameters for the Bowtie analysis. Typically, the defaults settings work just fine. In this case, our reads are not paired-end, but if you do have paired-end reads, make sure the appropriate checkbox is selected.

Click on the OK button to start the Bowtie alignment.

During the analysis, a job progress sheet will open;

Status: Running Running Bowtie indexing and analysis (step 4	of 12)									
Elapsed time: 11s										
You can close this dialog and track progress using the Job Manager (see the Windows menu).	Close Stop									

The job should complete in less than a minute on a reasonably modern machine.

Click on the View button to close the dialog and show the results.

The Bowtie results are encapsulated in a named job item in the project. If you click on the disclosure item to the left of the job name you will see (a) a "Reference Contig" that you can click on to open up the alignment in the *Contig Editor* and (b) any reads that do not align are collected into an Unaligned_Reads file. While these are typically failed or contaminant reads, there may be times when these are the reads you want e.g. if your reference contained just rRNA and tRNA genes are you wanted to generate a fastq file enriched in mRNA.

•) 🔴 🔵 Untitled — Project															
C		- Ale and a second seco	87	÷.	OF A	9	0	P	~	Ę	ेष	8	Q~Nar	ne		>>
Add	d Reads	Add Seqs	Add Ref	Add Contig	Remove	Rese	t Prefs	Replica	Phred	Cros	ssMatch	Phrap		Fi	lter	
	Projec	t 「	Proper	ties	Coverag	je										
	Name					~ S	atus		Length		#	С	lipL	0	ClipR	Start
100	Camp	oylobacte	r_jejuni_l	A3902		R	F		1635	045				1	1635045	
	RNAS	Seq.fastq.	gz			llu	ımina Uı	npaired		100	2000	000				
Æ	▼ Bowti	ie 1 - 14:5	58 - Jan '	16, 2019												
	Ur	naligned_l	Reads_1.f	q.gz		llu	ımina Uı	npaired		100	23	367				
ŧ۳,	Ca	ampyloba	cter_jeju	ni_IA3902 C	ontig 1				1635	046	1976	633				

Analyzing the Bowtie Results

Double-click on the **Contig 1** item to open the *Reference Contig* document window. Switch to the Editor tab if necessary.



By default, the consensus is shown in the middle of the window. If you want to see the consensus at the top, immediately underneath the reference sequence (as shown above), click on the Prefs toolbar button.

The reads are shown aligned to the reference. You can scroll through the entire assembly (1.6 Mbp and \sim 198,000 reads) if you wish.

Select the Map tab.

•••		Can	npylobacter_jeju	ni_IA3902 Cont	ig 1 — Map		
	- - - -	× 0.	· 🚯 😫	$\geq \otimes$			1
Locked Prefs	Replica Add Seqs	Digest 100%	Preview Create	Edit Delete			Range
Editor	Мар	Features	Annotations	Summary	Coverage	SNPs	VCF
			Campylobact	er_jejuni_IA3902	-		
	R)) 4 (0) 4 (1)						
M () () ()	B)] (0) (0))						
994	400 198800 298200	397600 497000 5	96400 695800 795	200 894600 99400	0 1093400 1192800	1292200 1391600	1491000 1590400
			LAN AN AN A DI TAN AN AN	NAME AND A DAY OF A	NI MALA ANG ING I UKAN		
				IN MARINE TANKE AND AND AND AND A	HA HAATO WANTING ING IN MAAD	III NI IN III WARAN IN	NA KUMATNAN IJIN I DUKUN
	CHERCE HAN DE MAN						# (13 C) (1 1 1 1 1
SNPs							
INDELs		\wedge					
7000							
6000							
5000							
3000							
2000							
0							1
-		$\mathbf{\nabla}$					
(DAN) PROBANISH DI DISCONDE DA	n - El de la presenta portez den de Calaba de Balana. N - El de la - Hel a portez den de Calaba de Balana.	ana mara in o kana kana kana hana na sa Di Kata kana kana kana kana kana basa kana	N X XV DO FRINCINGO (MARANI INDO XXXXX) I N X XV DO FRINCINGO (MARANI INDO XXXXXX) I	TERE DE DE MERINE DO	na name anders in des de solaris de Na Indes anders in de de solaris de	H BANDIINNA BADANA I H BANDIINNA BADA	100 1 3 1 1 5 100 100 100 10 10 1
LINE OF ORDER OF O							

This shows a map of the entire genome, along with coverage information for the reads. You can see there are several spikes of high coverage. Lets zoom into the circled area.

Carefully click to the left of the circled area, hold down the mouse button, drag to the right of the area and let go.

If you repeatedly drag, eventually you will see that there is a peak of coverage immediately under a single green **gene** feature;



Let the mouse pointer hover over the green arrow representing the gene. The annotations for the **gene** are displayed in a tooltip.

In this case, the annotation is not too informative, as there is simply a cryptic /locus_tag qualifier with the value "CJSA_Cj5SB". However, it is simple to run an internet BLAST search to find out more information;

Click on the green Feature arrow so select it, then choose the **Database I Online Search for Similar Sequences (BLAST)** menu item. Set up the dialog as shown below.

Note that MacVector remembers the region you selected when you clicked on the green arrow,

<u>NCBI Websit</u>	e and Data Usage Policies and Disc	laimers
Program: blastn ᅌ	Region: from 404273	to 404392
Database: nr Perform gapped alignm 	Expect: 10	O
More Choices	Defaults	Cancel OK
Click OK, wait for the job sheet. Switch to the BLAS	o to complete, then click ST Map results tab.	OK in the resulting
Can	npylobacter_jejuni_IA3902 Contig 1 — Results	3
C BLAST Description List C BLAST Alighed Sequer		1
	—	
CRM98_01310	55 othersonnal BINA	
b		acetylmuramoyl-L-alanineD-glutamate ligase
CRM98_03290	rain FDAARGOS_422 chromosome, complete genome [gi 12	bypothetical protein
23S ribosomal RNA	5S ribosomal RNA	CHM98_03280 5
🕻 🗣 ent 2 from 590488 to 590369 of Campylobacter jejuni st	train FDAARGOS_422 chromosome, complete genome [gij12	266938355jgbjCP023867.1j] score 240 e-value 4.1
CRM98_08320	5S ribosomal RNA	hypothetical protein
23S ribosomal RNA		CP

The BLAST Map displays a graphical representation of the region around the primary alignment. You can see the gene we selected at the top and the *High-scoring Segment Pairs* of the matching database sequences aligned underneath, complete with annotations on the matching region along with ~2kb on either side. It is immediately apparent that the gene aligns to 5s rRNA genes, exactly the sort of gene we would expect to express high levels of RNA in the experiment.

Close the BLAST results window and switch to the Summary tab of the reference contig window.

Editor	Мар	Features	Annotations	Summary
Summary report f	or Campylobacte	r_jejuni_IA3902	Contig 1	
Number of segmer	ts: 3025			
Total residues o	overed by reads	: 1069028		
longest consensu	IOT COVERED DY R	eads: 566017		
Average length o	of consensus seg	ments: 353		
Number of aligne	d reads: 197633			
Number of unique	reads aligned:	197633		
Number of unalig	ned reads: 2367			
Average read ler	reads: 200000 offh: 100			
Average coverage	depth: 15			
Average quality	value for conse	nsus: 14		
Number of conser	isus residues of	poor quality («	< 40): 1635047	
Regions with no	coverage:			
1-3 (3)				
335-336 (2)				
511-729 (219)				
847-1054 (208)				
1155-1230 (76)				
1565-1569 (5)				
1670-1690 (21)				
2034-2326 (293)				
4004-4932 (49) 5033-5087 (55)				

This tab summarizes the results of the Bowtie assembly. Because the data is RNASeq, the entire genome does not have coverage, as you would expect. In fact, there were 3,025 separate aligned segments, representing just over 1 million of the 1.6 million bases in the genome,

Campylobacter_jejuni_IA3902 Contig 1 — Coverage	Switch to the Coverage tab.												
Campylobacter_jejuni_IA3902 Contig 1 — Coverage Coverage SNPs VCF Editor Map Features Annotations Summary Coverage SNPs VCF Coverage report for Campylobacter_jejuni_IA3982 Contig 1 Coverage SNPs VCF Average coverage depth for Campylobacter_jejuni_IA3982 Contig 1 Coverage SNPs VCF Average coverage depth for Campylobacter_jejuni_IA3982 Contig 1 Coverage SNPs VCF Mame Type Start Stop Length Depth # Reads RPKM TPM dnaA CDS 1 1323 1323 0 13 74.85 62.44 dnaA CDS 1 13223 132.21 164.04 135.04													
Map Features Annotations Summary Coverage SNPs VCF Coverage report for Campylobacter_jejuni_IA3902 Contig 1 Image: Coverage depth in select features: Name Type Start Stop Length Depth # Reads RPKM TPM dinal CD5 1 1323 1323 0 13 74.85 62.44 dinal CD5 1 1324 1366 1 22 164.04 135.64	Campylobacter_jejuni_IA3902 Contig 1 — Coverage												
Editor Map Features Annotations Summary Coverage SNPs VCF Coverage report for Campylobacter_jejuni_IA3982 Contig 1	Locked Prefs Replica Dots												
Coverage report for Campylobacter_jejuni_IA3902 Contig 1 Average coverage depth in select features: Name Type Start Stop Length Depth # Reads RPKM Type Start Stop 1323 0 1374.85 62.44 dnaN CD5 1483 2550 1668 23 135.64	Editor Map Features Annotations Summary Coverage SNPs VCF												
Average coverage depth in select features: Name Type Start Stop Length Depth # Reads RFKM TPM dnaA CDS 1 1323 1323 0 13 74.85 62.44 dnaA CDS 1 483 2550 1068 1 23 164.04 136.84	Coverage report for Campylobact	Coverage report for Campylobacter jejuni IA3902 Contig 1											
Average coverage depth in select features: Name Type Start Stop Length Depth # Reads RPKM TPM dnaA CDS 1 1323 1323 0 13 74.85 62.44 dnaA CDS 1 6183 2.550 1668 1 2.3 164.04 136.64													
Name Type Start Stop Length Depth # Reads RPKM TPM dnaA CDS 1 1323 1323 0 13 74.85 62.44 dnaA CDS 1 686 1 23 164.04 136.64	Average coverage depth in select features:												
dnaA CDS 1 1323 1323 0 13 74.85 62.44 dnaN CDS 1483 2550 1068 1 23 164.04 136.84	Name	Туре	Start	Stop	Length	Depth #	Reads	RPKM	TPM				
dnaN CDS 1483 2550 1068 1 23 164.04 136.84	dnaA	CDS	1	1323	1323	Θ	13	74.85	62.44				
	dnaN	CDS	1483	2550	1068	1	23	164.04	136.84				
gyrB CDS 25/9 4888 2310 / 188 619.92 51/.14	gyrB	CDS	2579	4888	2310	7	188	619.92	517.14				
putative periplasmic protein CDS 4916 5257 342 1 8 1/8.18 148.64	putative periplasmic protein	CDS	4916	5257	342	1	8	1/8.18	148.64				
molybdopterin Oxidoreductase T CDS 5260 6498 1239 1 24 147.55 123.08	molybdopterin oxidoreductase f	CDS	5260	6498	1239	1	24	147.55	123.08				
Na7/n+ antiporter family prote CDS 0/09 0010 1302 0 10 50.50 40.00	alte	CDS	0/09	12624	1302	2	142	242 54	40.00				
glib (D) 0144 16054 4451 5 145 242.54 202.55 concerved hypothetical protein (DS 126/4 16254 2455 0 6 6 5 8 83	conserved hypothetical protein	CDS	12644	1/395	1752	9	145	69 56	202.33				
oltho CDS 14398 1584 1446 4 63 331 87 276 84	g1th	CDS	14398	15843	1446	4	63	331 87	276 84				
rnhB CDS 15844 16419 576 7 46 608.31 507.46	rnhB	CDS	15844	16419	576	7	46	608.31	507.46				
COMEA CDS 16452 16691 240 25 64 2031.24 1694.46	COMEA	CDS	16452	16691	240	25	64	2031.24	1694.46				
rbr CDS 16756 17403 648 54 366 4302.27 3588.96	rbr	CDS	16756	17403	648	54	366	4302.27	3588.96				
ilvD CDS 17563 19239 1677 5 89 404.25 337.23	ilvD	CDS	17563	19239	1677	5	89	404.25	337.23				
putative integral membrane pro CDS 19251 19775 525 4 26 377.23 314.69	putative integral membrane pro	CDS	19251	19775	525	4	26	377.23	314.69				
conserved hypothetical protein CDS 19867 21093 1227 1 24 148.99 124.29	conserved hypothetical protein	CDS	19867	21093	1227	1	24	148.99	124.29				
ExsB CDS 21170 21844 675 3 23 259.55 216.51	ExsB	CDS	21170	21844	675	3	23	259.55	216.51				
dsbI CDS 21865 23391 1527 4 69 344.19 287.13	dsbI	CDS	21865	23391	1527	4	69	344.19	287.13				
dba CDS 23403 23570 168 3 8 362.72 302.58	dba	CDS	23403	23570	168	3	8	362.72	302.58				
methyl-accepting chemotaxis pr CDS 236/6 25454 1//9 4 83 355.38 296.46	metnyl-accepting cnemotaxis pr	CDS	236/6	25454	1//9	4	83	355.38	296.46				
CCDA-1 LDS 25444 26558 915 16 159 1323.63 1104.18	CCPA-1 fumarulacatoacatato hudrolaco	CDS	25444	26358	915	16	159	1323.63	1104.18				
Tunary lace to ace to ace to 2002 2000 0/9 2 2/ 2007 195.10	RNA pseudouridylate synthese f	CDS	20422	27300	846	2	27	233.97	195.10				
	nurB	CDS	28393	29721	1329	å	45	257 92	215 15				
nrdA CDS 29737 32106 2370 4 129 414.60 345.86	nrdA	CDS	29737	32106	2370	4	129	414.60	345.86				
sodium/dicarboxylate_symporter_CDS3214533530138610149818.87683.10	sodium/dicarboxylate symporter	CDS	32145	33530	1386	10	149	818.87	683.10				
thyX CDS 33650 34273 624 2 17 207.52 173.11	thvX	CDS	33650	34273	624	2	17	207.52	173.11				
pyrG CDS 34393 36024 1632 4 71 331.38 276.44	pyrG	CDS	34393	36024	1632	4	71	331.38	276.44				
recJ CDS 36011 37582 1572 1 32 155.06 129.35	recJ	CDS	36011	37582	1572	1	32	155.06	129.35				
ansA CDS 37678 38673 996 1 12 91.77 76.56	ansA	CDS	37678	38673	996	1	12	91.77	76.56				
hypothetical protein CDS 44967 45494 528 0 2 28.85 24.07	hypothetical protein	CDS	44967	45494	528	Θ	2	28.85	24.07				
hypothetical protein CDS 45484 46170 687 0 0 0.00 0.00	hypothetical protein	CDS	45484	46170	687	0	0	0.00	0.00				
type II restriction-modificati CDS 46231 50004 3774 0 11 22.20 18.52	type II restriction-modificati	CDS	46231	50004	3774	0	11	22.20	18.52				
putative cytoplasmic protein CDS $50001 52127 2127 0 2 7.16 5.97$	putative cytoplasmic protein	CDS	50001	52127	2127	0	2	7.16	5.97				
putative periptasmic protein CDS 52202 52303 /02 / 53 5/5.06 4/9./4 MES family drug resistance tra CDS 52900 54102 1203 1 14 88.64 73.95	MES family drug resistance tra	CDS	52202	54102	1203	1	53 14	3/3.00	4/9./4				

٦

This is by far the most useful tab for RNASeq expression analysis. There are a number of columns;

Name: this is the preferred name of the feature. For **CDS** features it is typically the contents of the /gene qualifier, but MacVector will use other qualifiers if /gene is not present.

Type: the type of feature. By default MacVector only displays **CDS** and **gene** features but other feature types can be requested.

Start: the start location of the feature.

Stop: the stop location of the feature.

Length: the length of the feature.

Depth: the average depth of coverage across the entire length of the feature (rounded down).

Reads: the total number of reads that aligned to the feature.

RPKM: Reads Per Kilobase of transcript per Million mapped reads. This is a common calculation used to normalize the data to facilitate comparison of expression levels between genes. It is calculated as follows;

- Count up the total reads in a sample and divide that number by 1,000,000 this is our "per million" scaling factor.
- Divide the read counts by the "per million" scaling factor. This normalizes for sequencing depth, giving you reads per million (RPM)
- Divide the RPM values by the length of the gene, in kilobases. This gives you RPKM.

TPM: Transcripts Per Kilobase Million. This is a variation on RPKM that is calculated slightly differently;

- Divide the read counts by the length of each gene in kilobases. This gives you reads per kilobase (RPK).
- Count up all the RPK values in a sample and divide this number by 1,000,000. This is your "per million" scaling factor.
- Divide the RPK values by the "per million" scaling factor. This gives you TPM.

The advantage of using TPM is that this normalizes the data between different experiments so that you can directly compare the values for the same gene between different runs.

Exporting Data to Excel

The data in the Coverage tab is formatted to simplify exporting the columns into Microsoft Excel for further analysis. Specifically, the

columns are tab-separated so that when you copy and paste into Excel, each value gets pasted into a separate cell.

Carefully select the text starting at the **Name** column header. Hold down the <shift> key and scroll to the bottom of the window. Still holding down <shift>, click just after the last character of the bottom line, so that the entire table gets selected. Choose **Edit I Copy**

This copies the entire text table to the clipboard. Now we can switch to Microsoft Excel and paste the data into a new workbook.

Open Microsoft Excel and (if necessary) create a new blank workbook. Select the top-left cell. Choose **Edit I Paste Special...**

Older versions of Microsoft Excel (e.g. Office 2008) would correctly paste the tab-separated values directly, but Excel 2016 and later do not paste correctly, requiring you to go through this workaround;

	Paste Spe	ecial
Source: Paste: Paste link:		
As:		
Text		
Result	Inserts the contents of the Clipboard as a Text format.	Display as icon
		Cancel OK

Even though "Text" is the only possible option, you must go through this dialog to get the required behavior.

Click OK

The data gets pasted into the workbook, with each data point in a separate cell;

•		ഹ എ ച		Workboo	k1	Q	Search She	et		(i) •	
	Home Insert P	age Layout Fo	ormulas Da	ata Rev	iew Vie	W			🛂 Shan	e ^	
P	aste		• A• A• • A •	Alignment	% • Number	Conditional Formatting * Format as Table * Cells Styles * Cell Styles *					
LA	$1 \overline{\mathbf{v}} \times \sqrt{f_{\mathbf{X}}} \text{Name} \qquad \qquad \mathbf{v}$										
4	A	В	С	D	E	F	G	Н	1		
1	Name	Туре	Start	Stop	Length	Depth	# Reads	RPKM	TPM		
2	dnaA	CDS	1	1323	1323	0	13	74.85	62.44		
3	dnaN	CDS	1483	2550	1068	1	23	164.04	136.84		
4	gyrB	CDS	2579	4888	2310	7	188	619.92	517.14		
5	putative periplasmic prote	in CDS	4916	5257	342	1	8	178.18	148.64		
6	molybdopterin oxidoreduc	tase f CDS	5260	6498	1239	1	24	147.55	123.08		
7	Na+/H+ antiporter family p	orote CDS	6709	8010	1302	0	10	58.5	48.8		
8	gltB	CDS	8144	12634	4491	3	143	242.54	202.33		
9	conserved hypothetical pro	otein CDS	12644	14395	1752	0	16	69.56	58.03		
10	gltD	CDS	14398	15843	1446	4	63	331.87	276.84		
11	rnhB	CDS	15844	16419	576	7	46	608.31	507.46		
12	comEA	CDS	16452	16691	240	25	64	2031.24	1694.46		
13	rbr	CDS	16756	17403	648	54	366	4302.27	3588.96		
14	ilvD	CDS	17563	19239	1677	5	89	404.25	337.23		
15	putative integral membran	e pro CDS	19251	19775	525	4	26	377.23	314.69		
16	conserved hypothetical pro	otein CDS	19867	21093	1227	1	24	148.99	124.29		
17	ExsB	CDS	21170	21844	675	3	23	259.55	216.51		
18	dsbl	CDS	21865	23391	1527	4	69	344.19	287.13		
19	dba	CDS	23403	23570	168	3	8	362.72	302.58		
20	methyl-accepting chemota	xis pr CDS	23676	25454	1779	4	83	355.38	296.46		
21	ccpA-1	CDS	25444	26358	915	16	159	1323.63	1104.18		
22	fumarylacetoacetate hydro	olase CDS	26422	27300	879	2	27	233.97	195.18		
23	RNA pseudouridylate synth	nase f CDS	27413	28258	846	0	0	0	0		
24	purB	CDS	28393	29721	1329	3	45	257.92	215.15		
25	nrdA	CDS	29737	32106	2370	4	129	414.6	345.86		
26	sodium/dicarhoxylate sym	norter CDS	32145	33530	1386	10	149	818 87	683 1		
-	Sheet1 +	-									
	Ready					= -		·	+ 10	00%	

Now you can repeat this procedure with multiple datasets representing time-points, drug treatments, different growth conditions etc., and use the built-in functions of Excel for advanced analysis and comparison between runs.