

DNASTAR RNASEQ TEMPLATED

Jean-Yves Sgro

February 21, 2017

Contents

1	Acknowledgement	2
2	Introduction	2
2.1	This Tutorial	2
3	Set-up	2
3.1	Lasergene DNASTAR	2
3.2	Download Data	3
4	Part A: Setting up a templated RNA-Seq project in SeqMan NGen	4
4.1	Launch SeqMan NGen	4
4.2	Choose where to work	4
4.3	Choose workflow	5
4.4	Choose Assembly Type	5
4.5	Reference genome	5
4.6	Input Sequence Files and Define Experiments	7
4.7	Group replicates	7
4.8	Choose control	7
4.9	Set Assembly options	7
4.10	Assembly output	9
4.11	Start Assembly	9
4.12	Finish Assembly	9
4.13	Save Project	11
5	Part B: Analyzing the RNA-Seq results in ArrayStar.	11
5.1	Launch ArrayStar	11
5.2	Get Started	11
5.3	Organize data	13
5.4	Quick Gene Set	13
5.5	Set List	15
5.6	Add information columns	15
5.7	Fold Change	15
5.8	Sort genes	19
5.9	Show selected genes	19
5.10	Identify possible operon structures	20
5.11	More potential operons	20
6	Appendix	22
6.1	Appendix A: Data	22
6.2	Appendix B: SigmaB	22
6.3	Appendix C: Sigma B Operon	23
6.4	Appendix D: Listeria monocytogenes	23
7	Resources	23
	REFERENCES	24

1 Acknowledgement

This tutorial is based on the Templated RNA-Seq Workflow¹ on the DNASTAR Tutorial web page². The tutorial is called “templated” as sequence reads are aligned to the genome template.

Note: A separate tutorial exists for cases where there is no template genome.

2 Introduction

Templated RNA-Seq uses next-gen sequencing to show the presence of RNA at a particular moment. RNA can be identified and quantified by alignment to the genome.

This tutorial is meant to become familiar with the DNASTAR software for next-gen sequencing. In this tutorial we will use two DNASTAR software:

- **SeqMan NGen** will be used to align RNA-Seq data onto the genome.
- **ArrayStar** will be used to analyse the completed RNA-Seq alignment assembly.

Note: While **SeqMan NGen** exists for both Mac and Windows, **ArrayStar** is a *Windows-only* software (as it is based on the Microsoft .Net framework.)

2.1 This Tutorial

In this tutorial, we will compare stationary phase RNA from wild-type *Listeria monocytogenes* cells with that from mutant cells that do not express sigma B, a major transcriptional regulator (see Oliver et al. (2009) and Appendix below.)

The tutorial is split in two parts following the software used:

- **Part A:** Setting up and running a templated RNA-Seq project in **SeqMan NGen**
- **Part B:** Analyzing the RNA-Seq results in **ArrayStar**

Choose which OS you prefer to work in. However, note that Part B can only be run within Windows and the file(s) from Part A would need to be transferred on the Windows side unless there is a way to share a directory.

For this tutorial it is therefore advised to run both Parts A and B under Windows.

3 Set-up

3.1 Lasergene DNASTAR

For this tutorial we need access to the DNASTAR software which is installed on the class iMacs.

- If you need to install the software on your lab or laptop computer follow instructions on the Biochemistry department web page for “Available Software”³ or open a Biochem IT job request on the Job Board.
- If you are within the Biochemistry department with a *wired* computer such as the iMacs in the classroom you can simply launch the software to use **SeqMan NGen** or **ArrayStar**.

¹http://www.dnastar.com/t-support-tutorials.aspx#/seqman_ngen_tutorials/#!/Documents/tutorial3templatedrnaseqworkflow.htm

²<http://www.dnastar.com/t-support-tutorials.aspx>

³<https://biochem.wisc.edu/intranet/it/available-software>

- If you are on a *wireless* computer you will need to connect to the Biochemistry network by VPN⁴

3.2 Download Data

The data for the workshop can be found on the DNASTAR Tutorial web page as T3_Templated_RNA-Seq.zip⁵

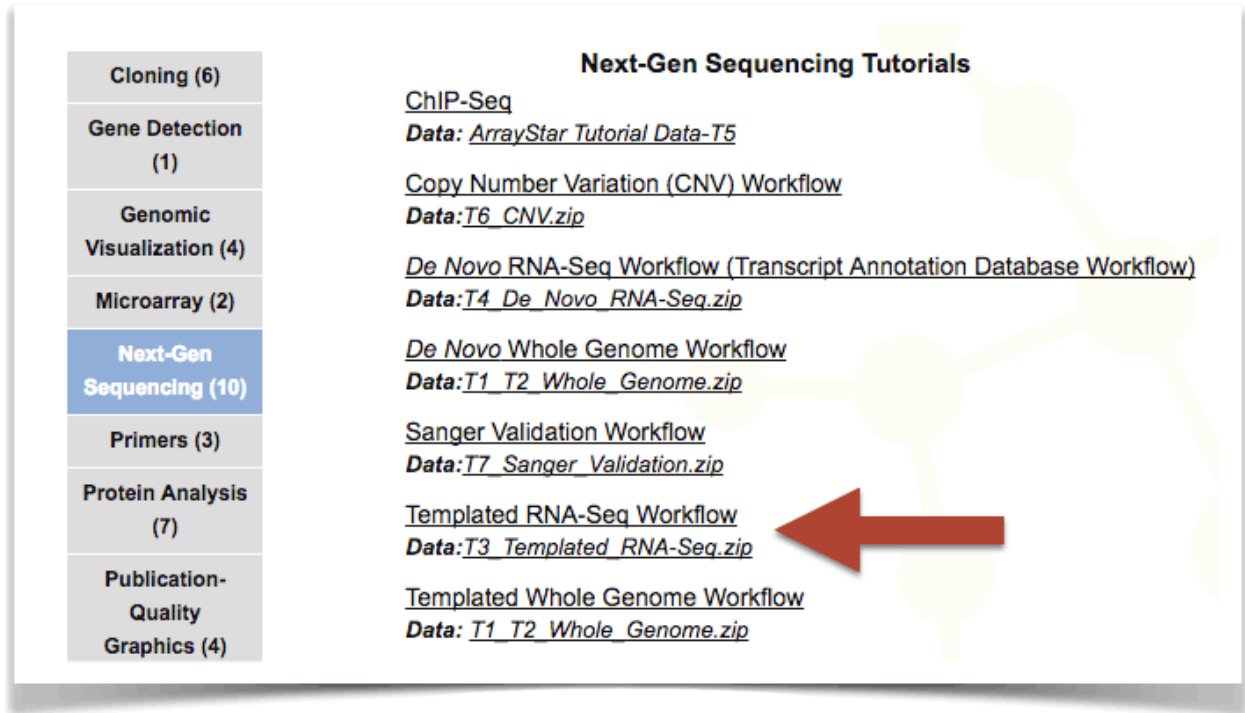


Figure 1: “Finding T3_Templated_RNA-Seq.zip on DNASTAR Tutorial web page”

☉ **TASK: Download the data to your desktop and unzip it.**

The resulting directory will contain 5 files:

```
Listeria monocytogenes 4b F2365.NC_002973.6.gbk
sigB_1.fastq
sigB_2.fastq
wt_1.fastq
wt_2.fastq
```

- The .gbk file is a GeneBank sequence file for the complete genome of the bacteria **Listeria monocytogenes**.
- The fastq files are sequencing reads. There are two replicates each for the wild type (wt) and for an *isogenic* strain lacking **Sigma B**. (See also **Appendix A** below.)

⁴<http://vpn.wisc.edu/>

⁵https://s3.amazonaws.com/star-deploy/SupportingDataFiles/LG14/SeqMan+NGen+Tutorial+Data/T3_Templated_RNA-Seq.zip

4 Part A: Setting up a templated RNA-Seq project in SeqMan NGen

4.1 Launch SeqMan NGen

On both Mac and Windows you can launch individual DNASTAR software by finding them on the hard drive or for example the “Start” menu under Windows.

However, the **DNASTAR Navigator** consolidates all DNASTAR software in one place and may make it easier to launch any of the desired software.

SeqMan NGen is located under **Genomics** within the Navigator.

Click on SeqMan NGen to launch.

4.1.1 Macintosh

Perhaps the easiest way would be to use “Spotlight Search” (top right “magnifying glass icon”) and start typing the name of the software. For SeqMan NGen it should appear after the first few letters are typed.

However, for this tutorial it is recommended to use Windows since the second part of the tutorial requires a Windows-only software.

4.1.2 Windows

Click on the “Start” button (bottom left - looks like a 4 white squares) and scroll down to the letter **D** where DNASTAR should be listed.

Click on the downward pointing arrow on the right hand side of the name and find the software you need. *e.g.* DNASTAR Navigator 14 or directly SeqMan NGen.

4.2 Choose where to work

The first screen after the launch offers 3 choices:

- Assemble on local computer
- Re-run local assembly
- Assemble on the DNASTAR cloud

☉ **TASK:** Click **Assemble on local computer**.



Figure 2: “Three choices on first screen**“

4.3 Choose workflow

On the next screen “Choose Assembly Workflow” screen, select **Transcriptome / RNA-Seq** and press Next.

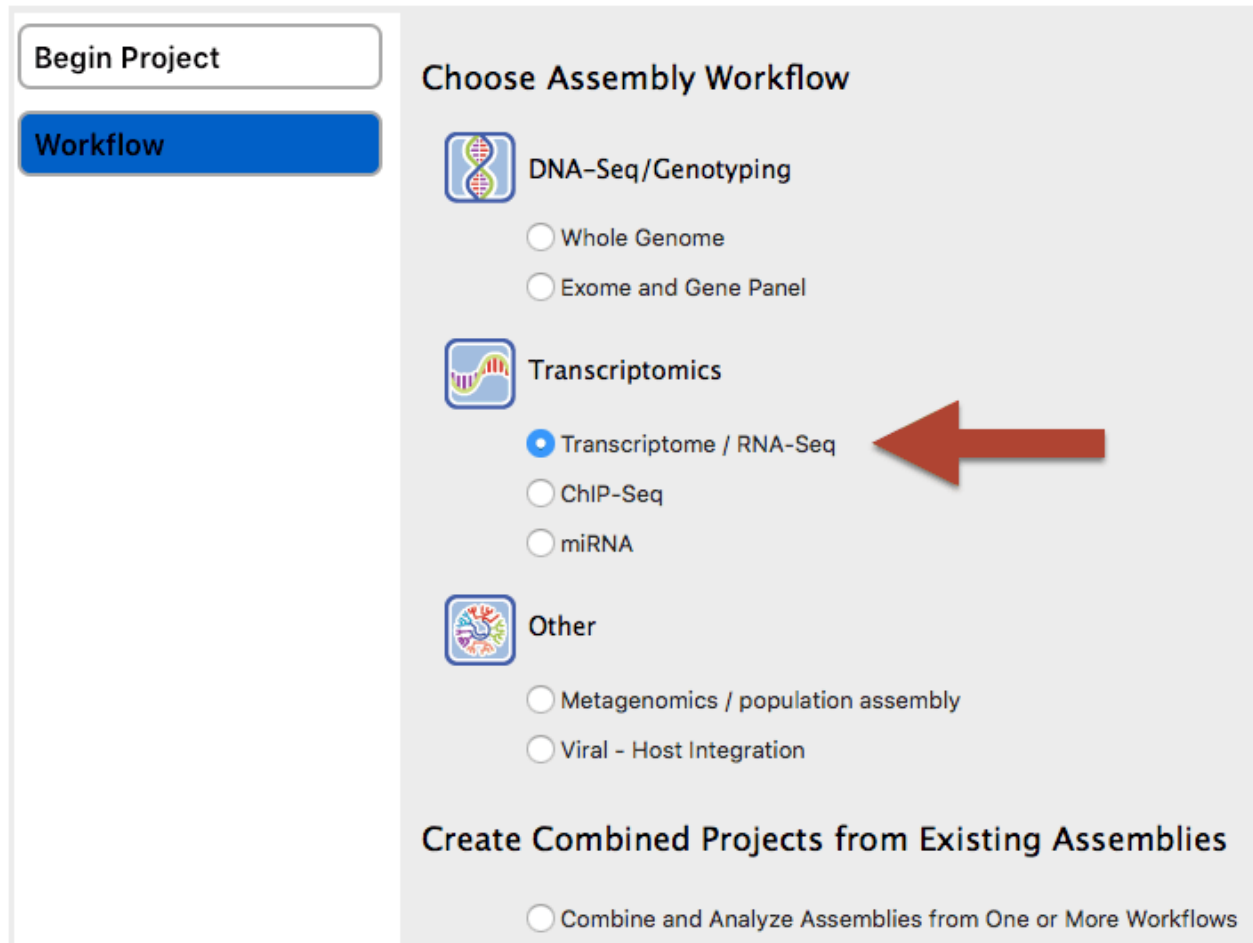


Figure 3: “select **Transcriptome / RNA-Seq**”

4.4 Choose Assembly Type

In the “Choose Assembly Type” screen, select **Reference based assembly** and click Next.

4.5 Reference genome

- In the Input Reference Sequences screen add the reference sequence *Listeria monocytogenes 4b F2365.NC_002973.6.gbk* by pressing the **Add** button.
- Then select the file and click **Open**.
- Click **Next**

Note: If a reference sequence had not been provided with the tutorial data, you could have downloaded an *L. monocytogenes* genome here using the **Download NCBI Genomes** button.

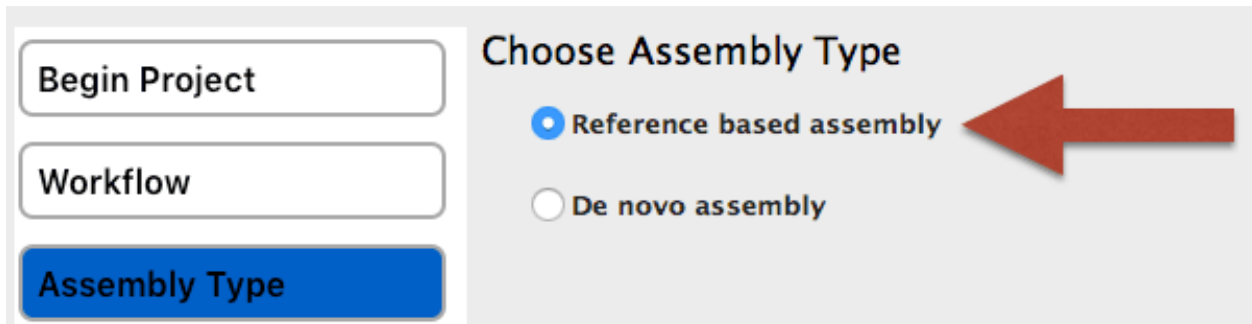


Figure 4: “select **Reference based assembly**”

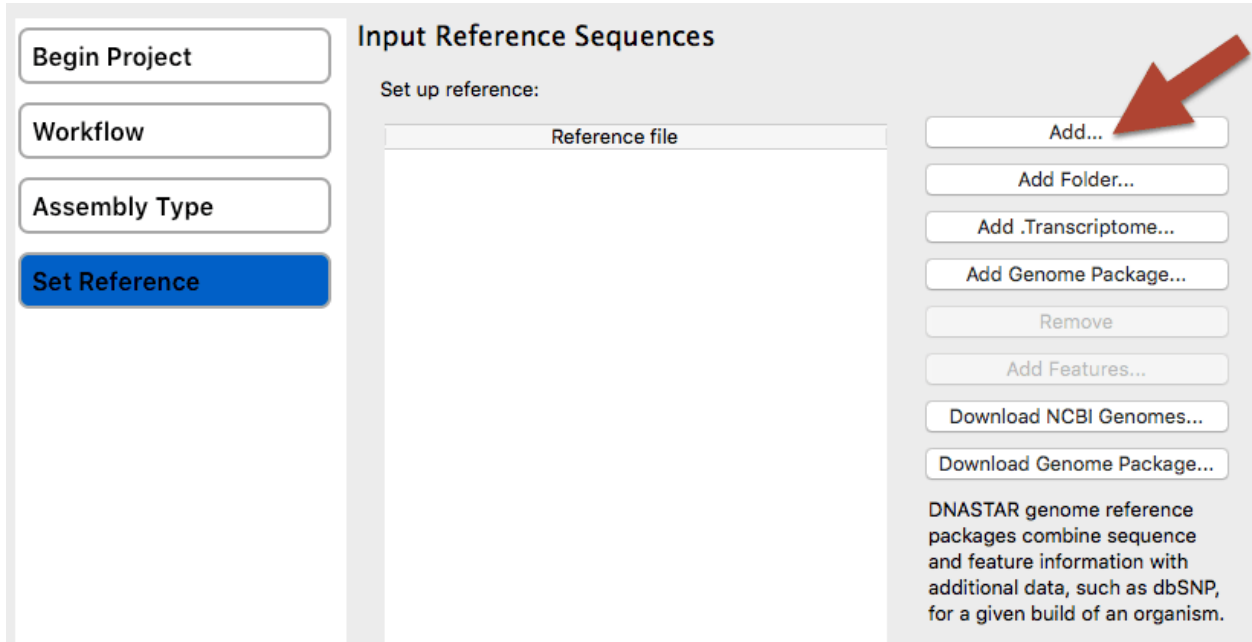


Figure 5: “Press **Add** and select reference sequence.”

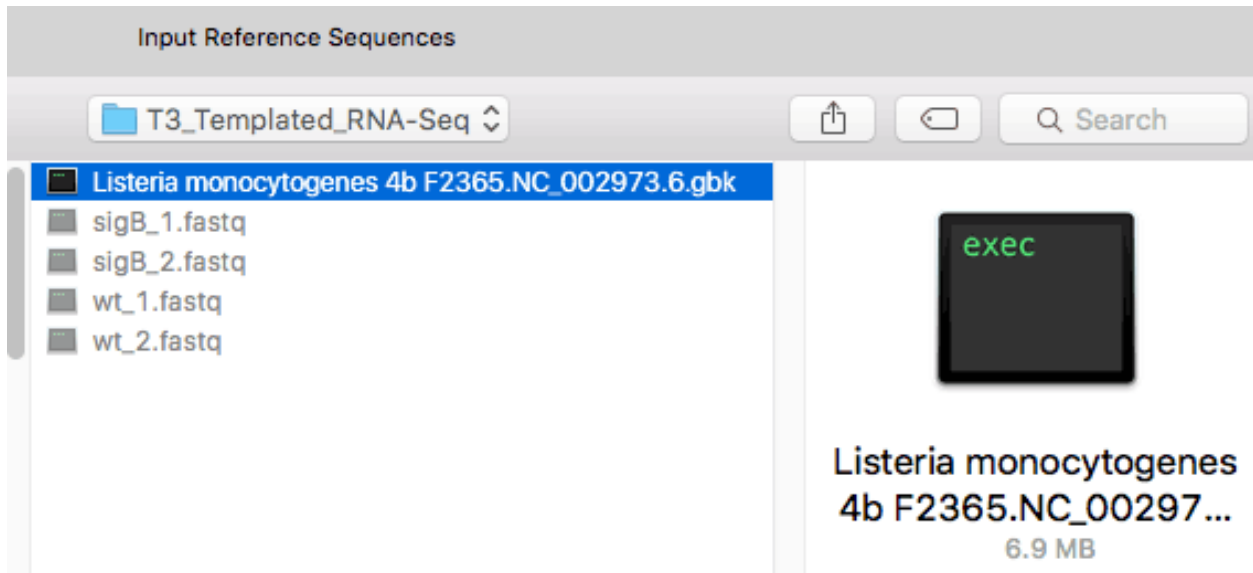


Figure 6: “Select file and click **Open**.”

4.6 Input Sequence Files and Define Experiments

In the Input Sequence Files and Define Experiments or Individual Replicates screen: (See illustration below.)

1. Set the **Read technology** to Illumina, and *uncheck* the **paired-end data** box.
2. Check the **Run Multi-sample data as separate assemblies** box.
3. Check the **Samples have replicates** box. When you do so, ***note that the “Experiment” column header below has changed to “Individual Replicate.”***
4. Using the procedure described in the previous step, use the **Add** button to add all four .fastq files from the tutorial data folder.
5. Name each of the four files by clicking on “ENTER NAME” and type in the name **sigB_1**, **sigB_2**, **wt_1** or **wt_2**, as appropriate for that row.
6. Click **Next**.

4.7 Group replicates

In the Group Individual Replicates into Replicate Sets screen:

1. Select the two **sigB** replicates and click on the **Group Selected** button. In the dialog, name the set **sigB** and click **OK**.
2. Do the same for the two **wt** replicates, naming the set “wt.”
3. Click **Next**.

4.8 Choose control

In the Set Up Experiments screen, check the **Is Control** box to the right of **wt**. Then click **Next**.

4.9 Set Assembly options

In the Assembly Options screen, check **Haploid** (since this is a bacterial genome)

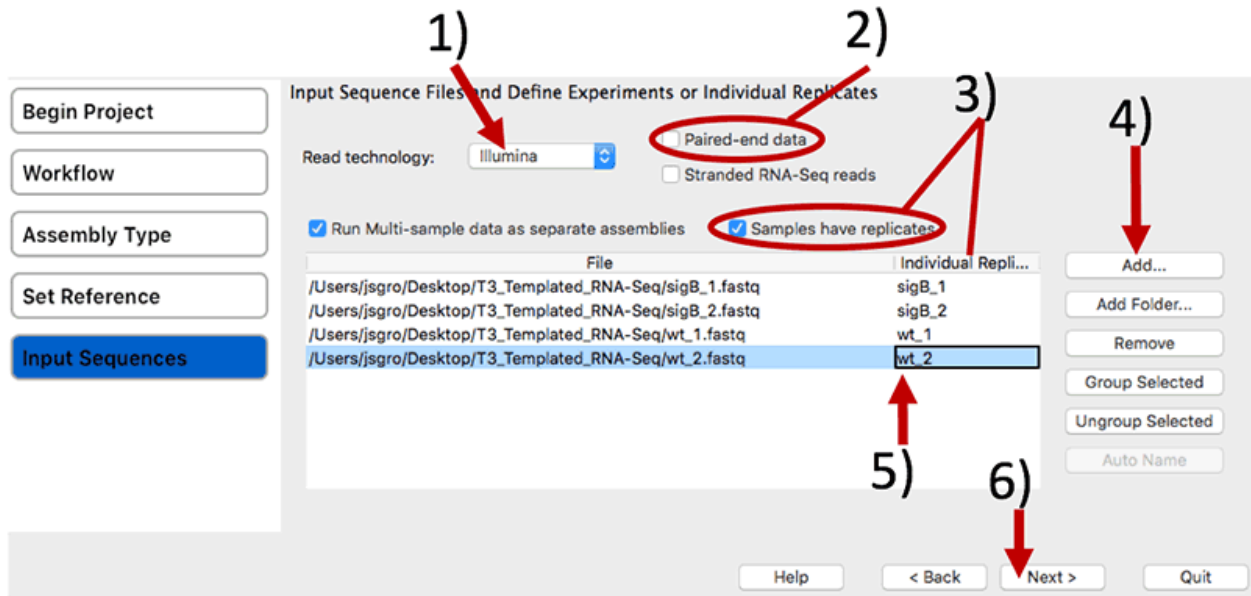


Figure 7: “Follow all steps to set-up files and define experimental details.”

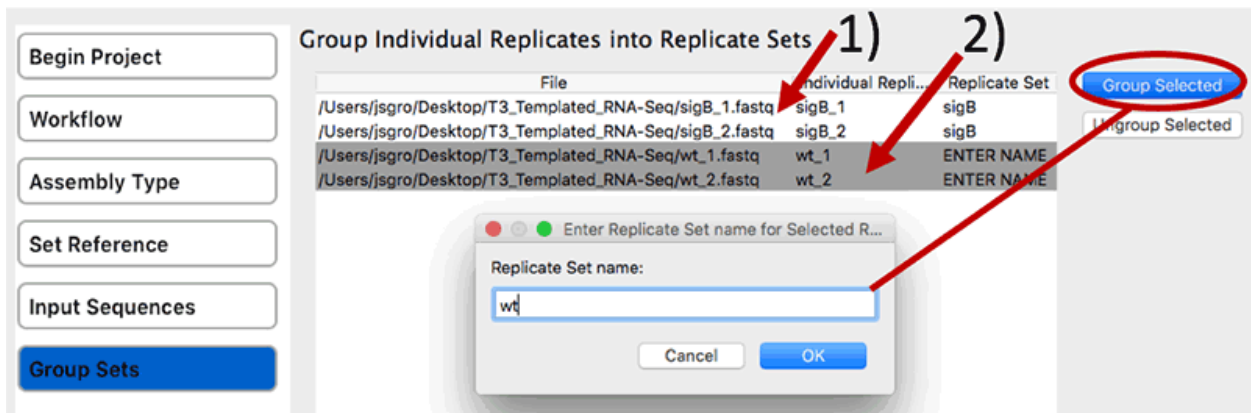


Figure 8: “Group replicates to define experiment.”

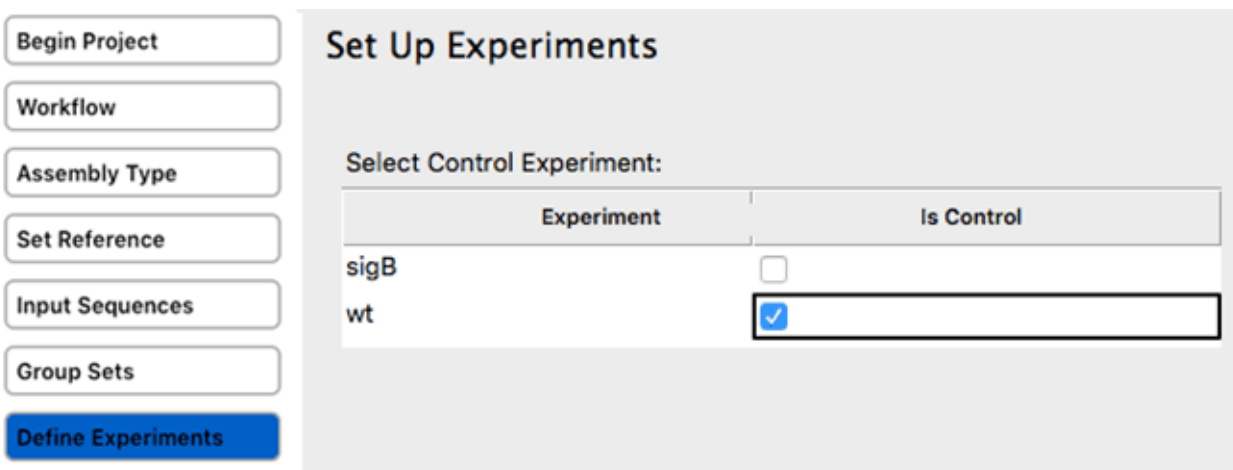


Figure 9: “Group replicates to define experiment.”

There is nothing else to change on that screen.

Then click **Next**.

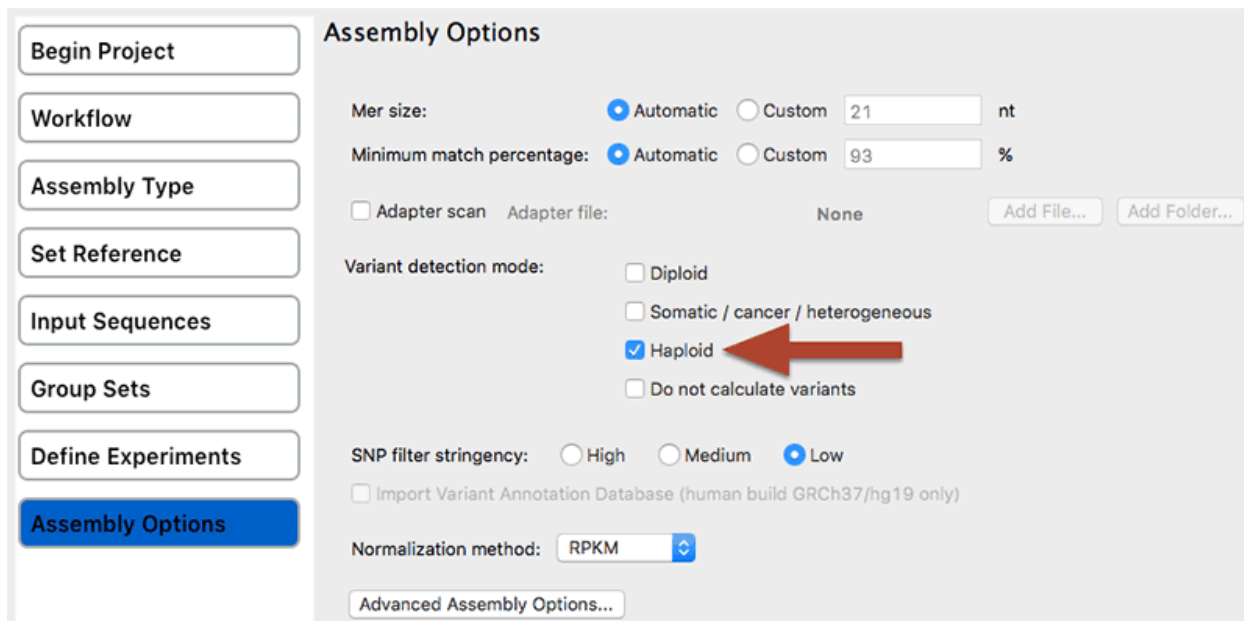


Figure 10: “Assembly options: check Haploid.”

4.10 Assembly output

In the Assembly Output screen:

1. Type “Templated RNA-Seq” into the **Project Name** text box. This name will be assigned to all output files, including the finished assembly.
2. Use the **Browse** button to specify a **Project Folder** for your assembly output files. *Note:* For local users, an alternative way to select a location is to drag and drop a folder from the file explorer onto the Project Folder row.
3. Click **Next**.

4.11 Start Assembly

In the “Your assembly is ready to begin” screen is revealed the script created by our previous clicks. However, all you have to do is press **Start Assembly** to begin the assembly.

Assembly will be complete within **about 5 minutes** depending on hardware configuration.

4.12 Finish Assembly

Wait until being informed that assembly has finished, then click **Next**.

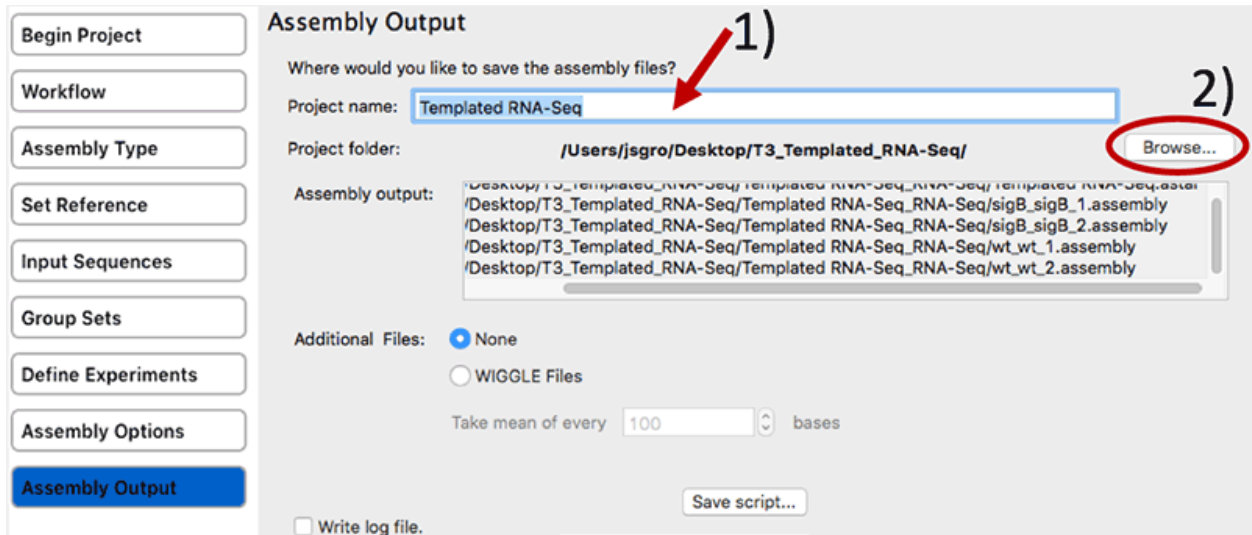


Figure 11: “Assembly options: check Haploid.”



Figure 12: “Start Assembly.”

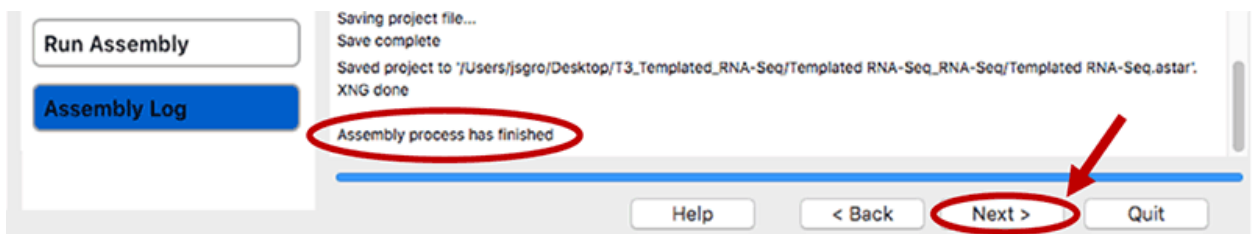


Figure 13: “Finish Assembly.”

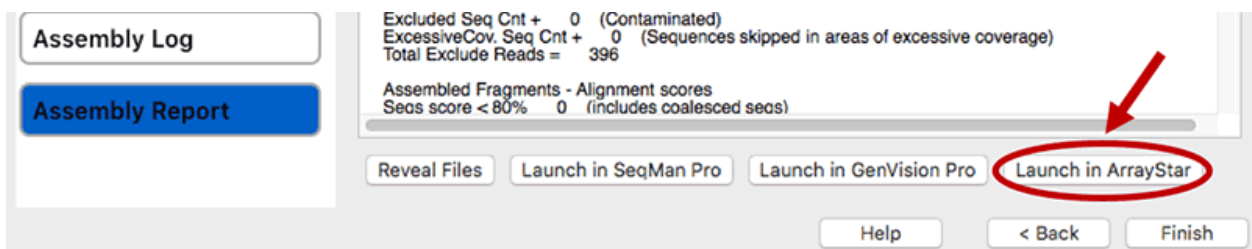


Figure 14: “Finish Assembly.”

4.13 Save Project

If you are on a Windows system the ArrayStar software will launch. See part B for continuing the analysis.

If you are on a Macintosh the transfer will not work and a warning message will appear:

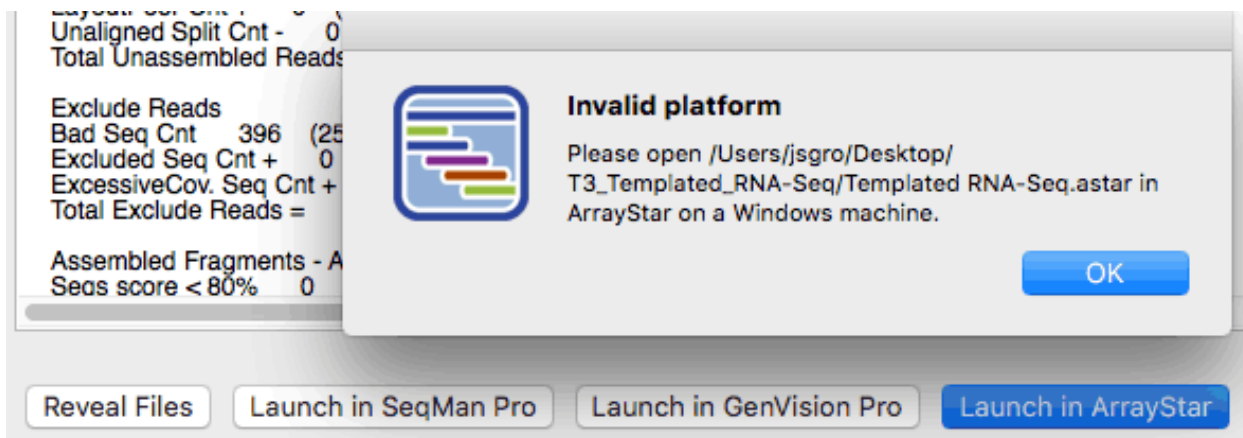


Figure 15: “Finish Assembly.”

Note: The file **Templated RNA-Seq.astar** can be transferred on the Windows side to continue the analysis.

5 Part B: Analyzing the RNA-Seq results in ArrayStar.

In Part B, we will analyze the results of the RNA-Seq assembly in ArrayStar by using a “*quick gene set*” to **locate a potential operon structure**.

An “operon” is a group of one or more genes that are transcribed as a single RNA unit.

In this section of the tutorial, we will create a “*quick gene set*,” then use the *Gene Table* to search for potential operon structures.

5.1 Launch ArrayStar

Either use the DNASTAR Navigator opened earlier (ArrayStar is listed under the **Genomics** category,) or find ArrayStar within the Windows “Start” menu on the bottom left (see beginning of tutorial above.)

5.2 Get Started

When we ran SeqMan NGen we saved a file called **Templated RNA-Seq.astar** which will serve as the start for the analysis. This file is compiled as a “project” and therefore:

Within the first panel in ArrayStar under **Get Started** choose **Open a project...** and navigate to where the file was saved (probably within a directory called “Templated RNA-Seq_RNA-Seq”)

Note: In order to be [allegedly] “helpful” Windows will hide know filename extensions. Therefore your file will appear *without* the **.astar** extension, which can be confusing.

Click **Open**

Note: It will take about 30 seconds to 1 minute to load and display the data under the “Scatter Plot” tab.



Figure 16: “Use the ‘Start’ button to launch ArrayStar.”

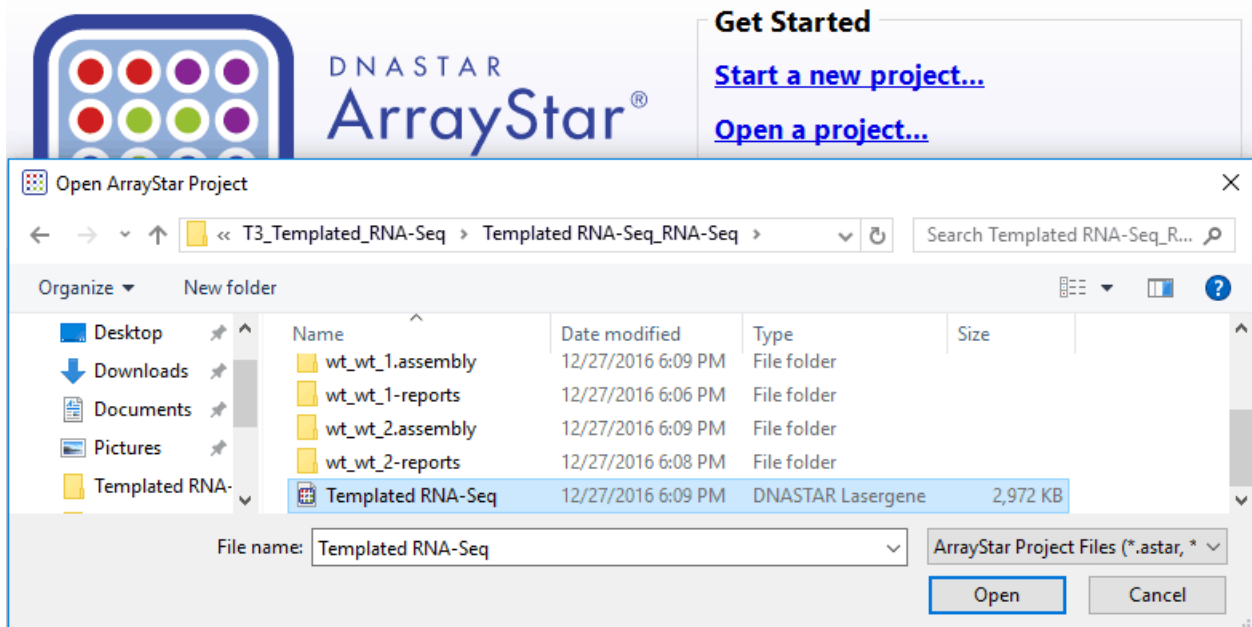


Figure 17: “Open project file Templated RNA-Seq.astar.”

5.3 Organize data

Before continuing any further click on the “Experiment List” tab

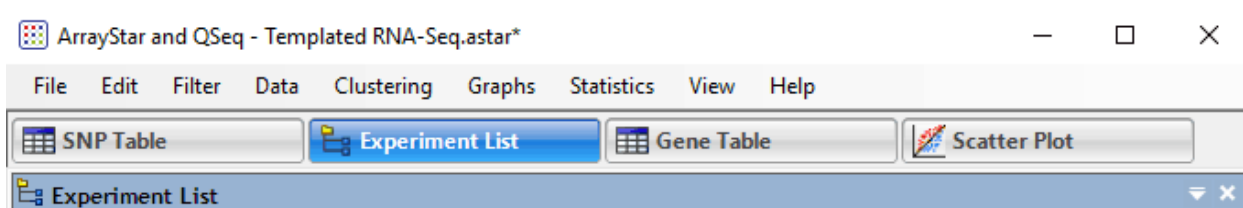


Figure 18: “Click on the”Experiment List" tab.“

Depending on how the data was loaded into ArrayStar, you will see either an RNA-Seq folder or both an RNA-Seq and Variant folder. In the latter case, select the Variant folder, then right-click on it and choose Delete. When prompted, press the Delete button.

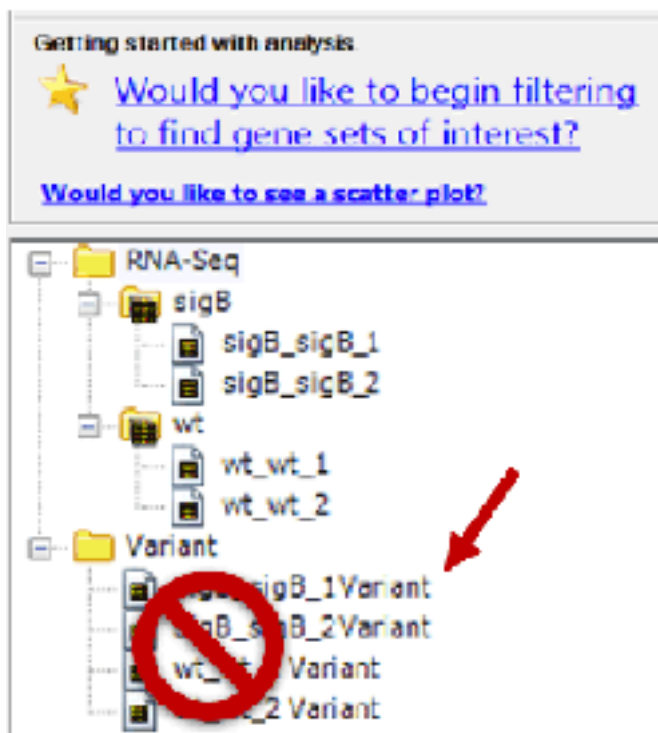


Figure 19: “Right Click on folder”Variant" to delete it.“

You will be warned with: “Are you sure you wish to delete 4 experiments? You will not be able to undo this deletion.”

The variant analysis is used as part of another DNASTAR tutorial and it is safe to remove these for our purpose: Click the **Delete** button.

5.4 Quick Gene Set

To access the Quick Gene Set Creation dialog, use the menu command **Graphs > Venn Diagrams** and then press the **Quick gene set creation** button.

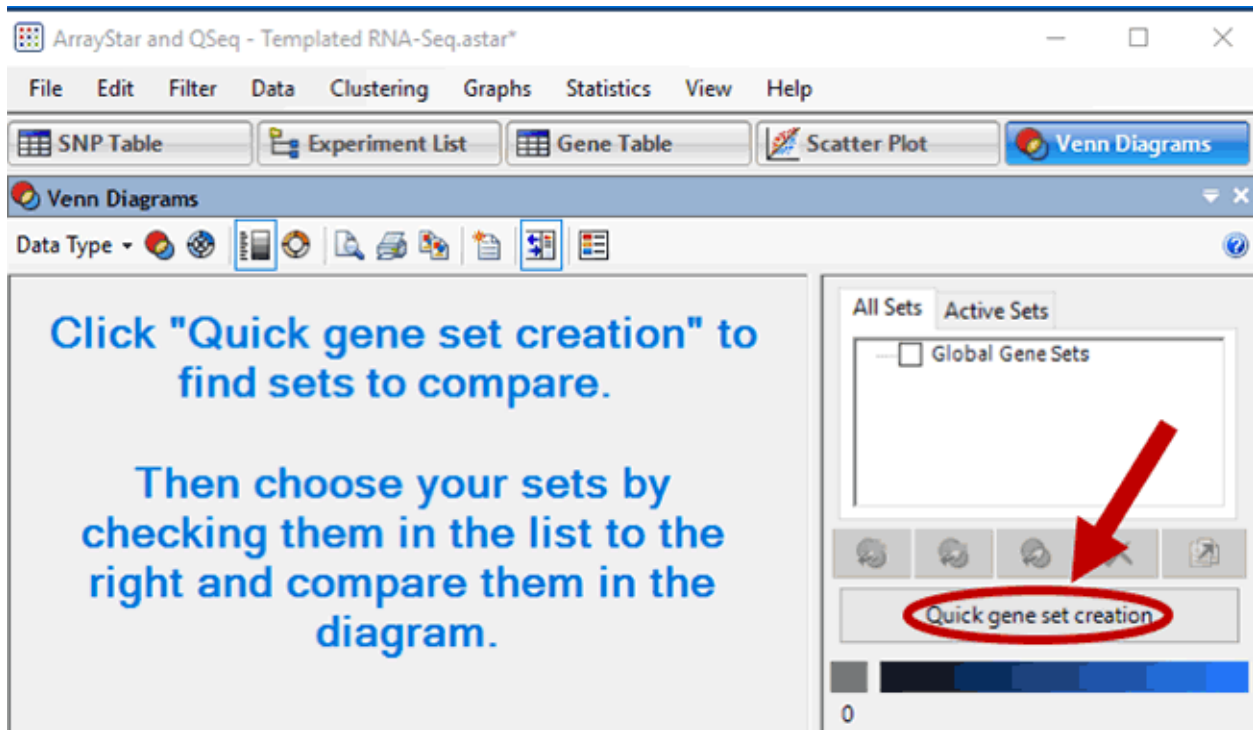


Figure 20: “Menu: **Graphs** > **Venn Diagrams** then click **Quick gene set creation**.”

This will open the “Step 1” comparison workflow window options and in the next section we will chose one.

5.4.1 Step 1- choose one comparison workflow

The window panel offers 3 different methods to compare and the experimental material:

1. Check expriments individually
2. Compare experiments to a baseline (we will choose this one below)
3. Compare all experiments pairwise

In the center section of Step 1, Compare Experiments to a Baseline, select a **Baseline Experiment** of **wt**, and then click the **Select** button just below.

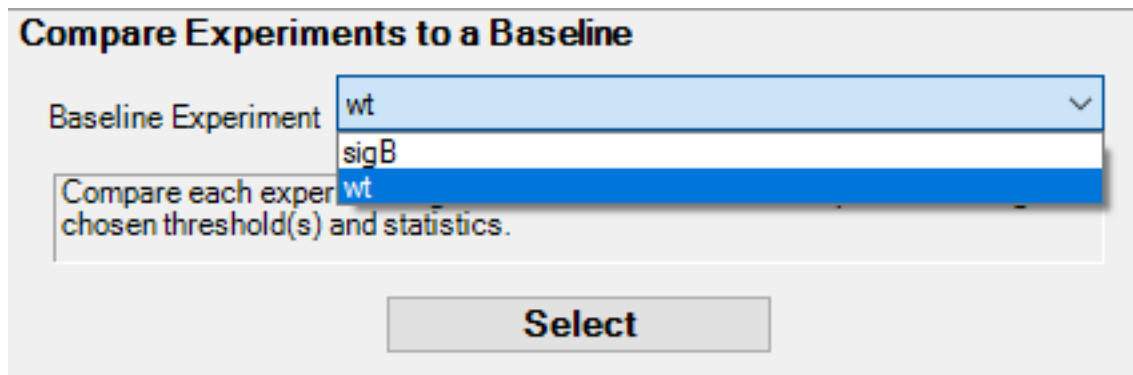


Figure 21: “Choose **wt** as the baseline.”

5.4.2 Step 2- Select experiments and genes to compare

Keep everything as the presented default: click button **Move to Step 3 (Comparisons)**.

5.4.3 Step 3

In Step 3, keep the **Signal Threshold** and **Fold Change** boxes checked, but uncheck the **P value** box.

Also remove the checkmark by the **Up** box, to the right of **Fold Change**.

The filter is now set up to find genes in the sigB mutant samples that have a ≥ 2 -fold downward change, compared to the wildtype, and an RPKM signal value ≥ 10 .

Press **Finish**.

5.5 Set List

Open the Set List by using the menu cascade **Data > Show Set List**. Note that the newly-created quick gene set is already selected and called **sigBxwt, 2 fold down, signal ≥ 10** .

5.5.1 Show gene table

DNASTAR software makes heavy use of icons that may not have menu items equivalents.

Such is the region of the ArrayStar panel called “Actions section” (see illustration below.)

In the Actions section, click the link **Select and show the table of this set’s Genes** (2nd icon from the left as illustrated below.)

While only three columns appear initially, the Gene Table can display a variety of gene name and annotation fields, notes, expression levels, and statistical calculations.

We will add some columns in the next section.

5.6 Add information columns

On the **Actions** section of icons click the **Add/Manage Columns** tool ( **Add/Manage Columns**) to open the Manage Columns dialog.

- Under Available Gene Info, select **Target Range**. Press the **> Add Column >** button to add the items to the Current Columns list.
- Click the **Gene Values** button. Then select **Signal**. Click the **Log2** radio button, then press **> Add 2 Columns >** to add them to the Current Columns list.
- Click **OK** to close the Manage Columns dialog and return to the Gene Table.

5.7 Fold Change

On the same line of icons click the **Add Fold Change** tool ( **Add Fold Change**).

Specify a Control of **wt** and an Experiment of **sigB**, then press **OK**.

The Gene Table should now contain seven columns.

Due to the choice made above all fold changes show a *down* direction.

Quick gene set creation ? X

Step 1 - Compare experiments to Baseline: wt ▲

Step 2 - 2 experiments selected for comparison. ▲

Step 3 - Select Comparisons

Signal Threshold >= 10.00

Compare all genes against this signal threshold to determine if they should be included in the results. When comparing pairs, the higher of the two experiments must exceed this threshold.

Fold change >= 2.00

Up Down

Calculate the fold change for each gene in a pair of experiments to determine if the gene should be included in the results.

P value <= 0.0500

Statistical Test Student's t-test

Correction Method FDR (Benjamini Hochberg)

Calculate the P value for each gene in a pair of replicate set experiments to determine if the gene should be included in the results.

Finish Cancel

Figure 22: “Uncheck **P value** and **Up** as marked.”

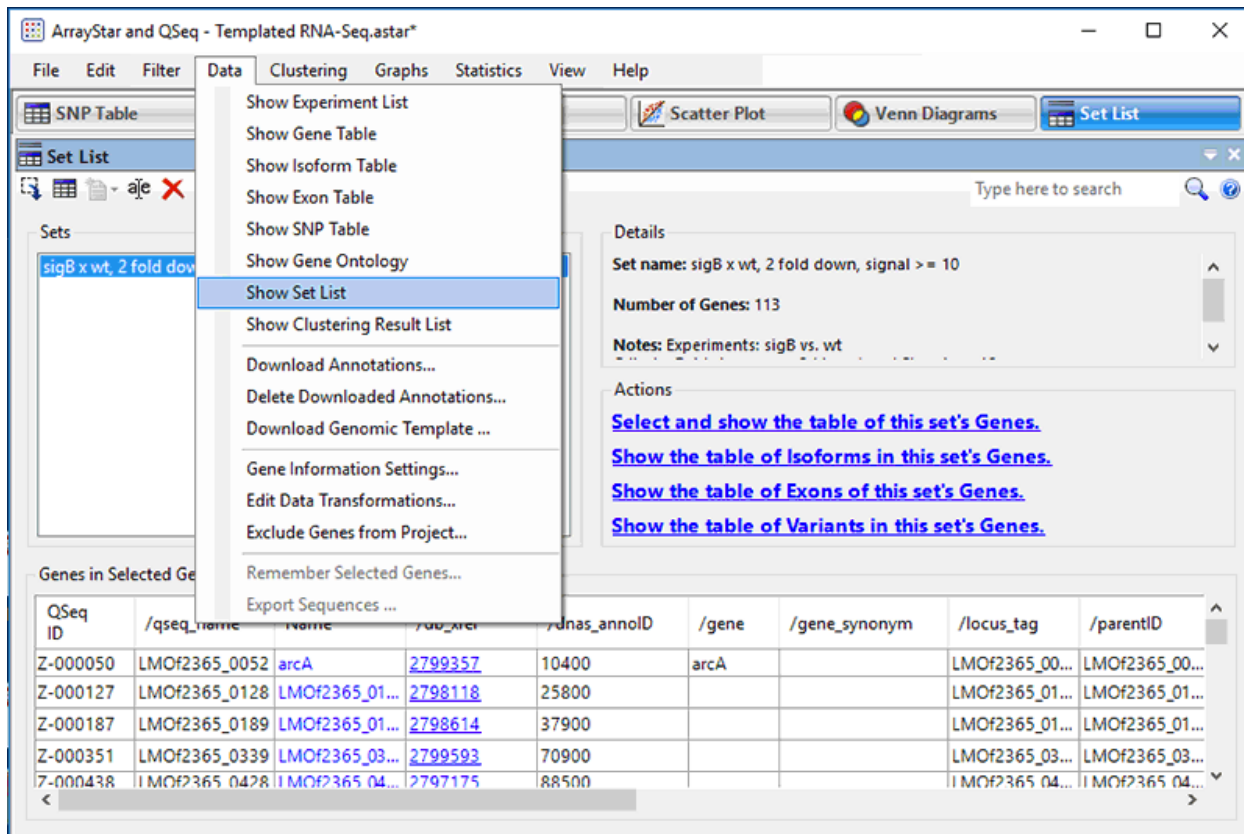


Figure 23: “Menu cascade **Data > Show Set List.**”

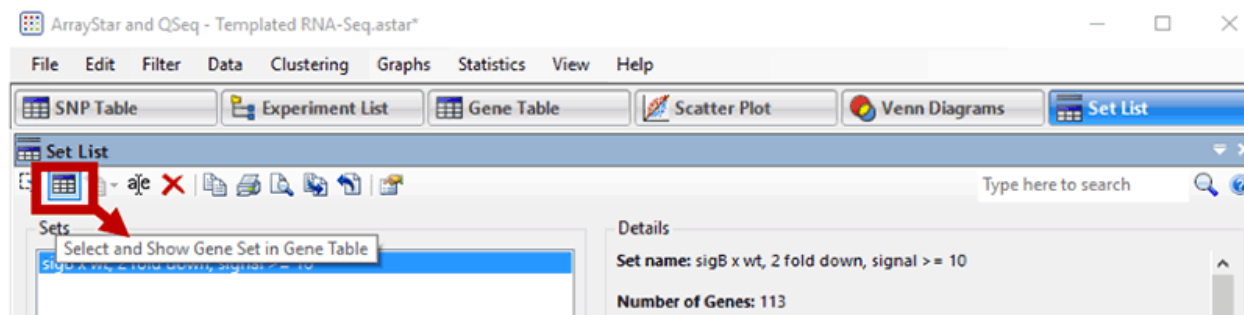


Figure 24: “Use the second button on the **Actions** section.”

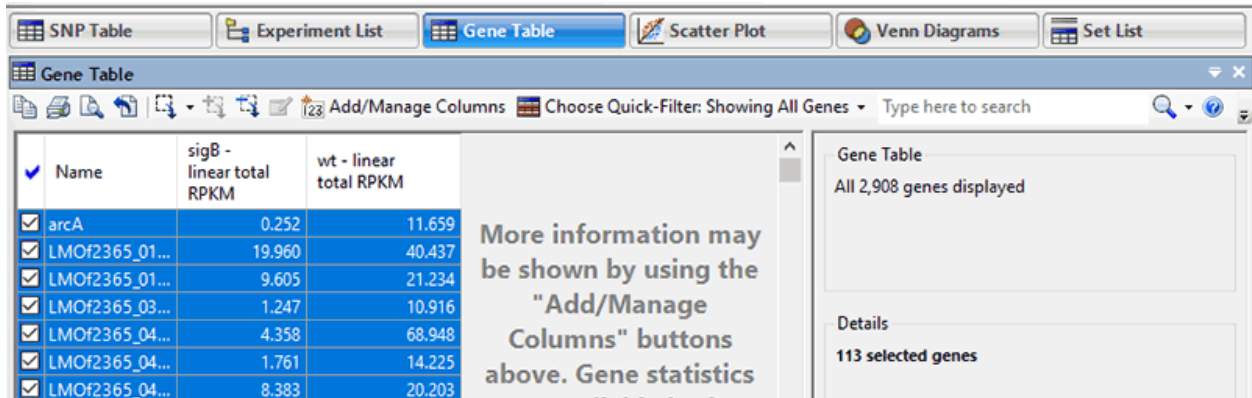


Figure 25: “Resulting Gene Table is first shown with only 3 columns.”

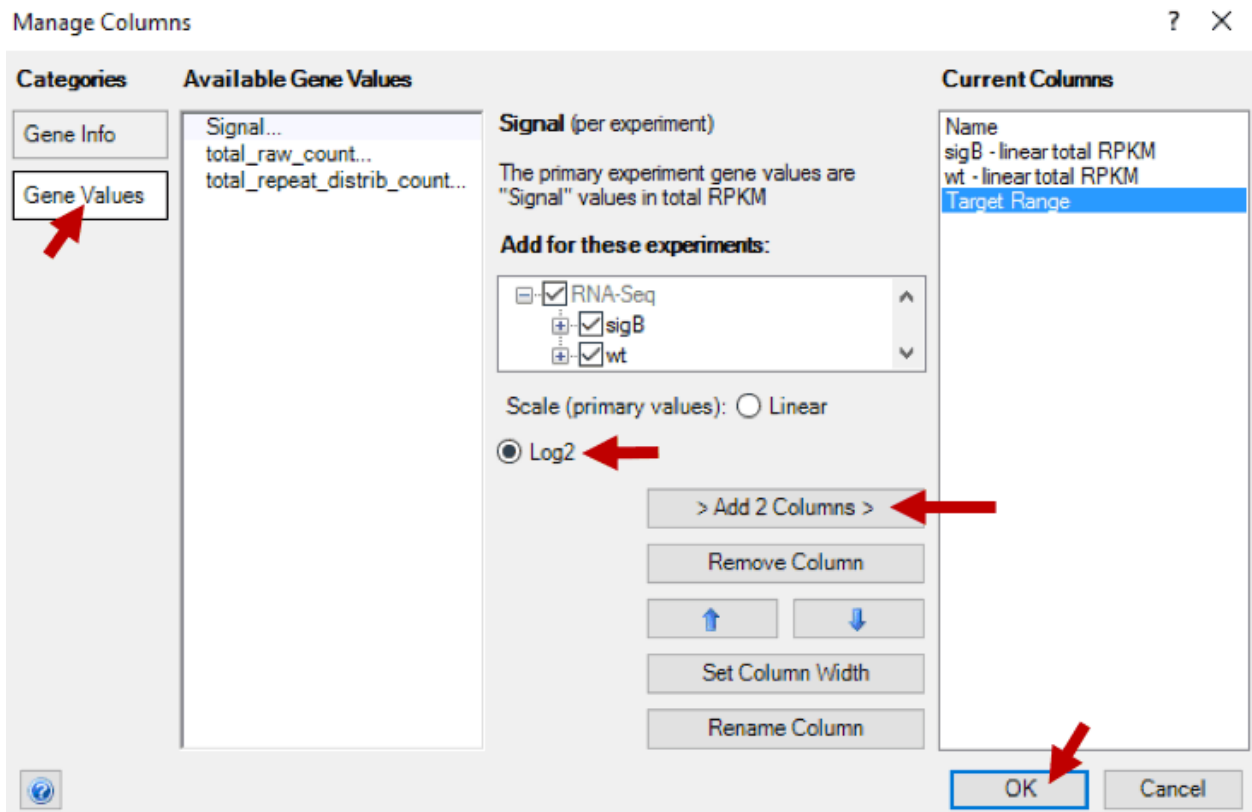


Figure 26: “Add/Manage Columns. Step to add Log2 values.”

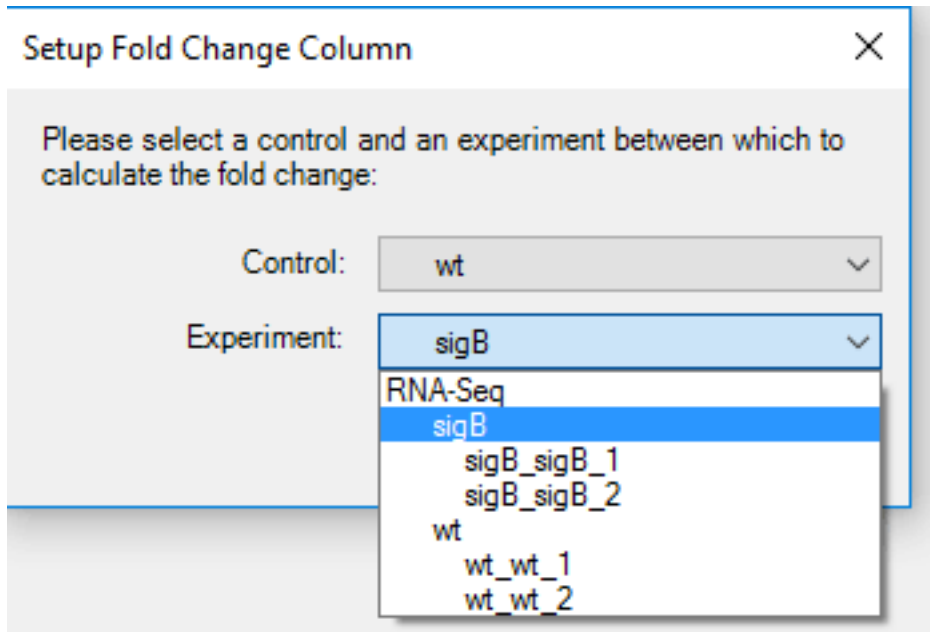


Figure 27: “Specify control and experiment samples”

5.8 Sort genes

Click once on the column header for **Target Range** to sort all of the genes in the project in *ascending order of appearance on the assembly*.


Scroll down, noting that the genes within the “quick gene set” remain selected in blue and are interspersed along the whole table as illustrated below.

<input checked="" type="checkbox"/>	Name	sigB - linear total RPKM	wt - linear total RPKM	Target Range	sigB - log ₂ total RPKM	wt - log ₂ total RPKM	Fold change
<input type="checkbox"/>	LMOF2365_0931	9.464	6.161	946211..946603	3.24239	2.62312	1.536 up
<input type="checkbox"/>	LMOF2365_0932	1.826	1.551	946606..947709	0.86847	0.63298	1.177 up
<input checked="" type="checkbox"/>	LMOF2365_0933	3.023	10.963	947792..948295	1.59604	3.45463	3.626 down
<input type="checkbox"/>	LMOF2365_0934	2.024	2.106	948409..949221	1.01740	1.07439	1.040 down
<input checked="" type="checkbox"/>	gabD	0.212	36.417	949446..950912	-2.23782	5.18655	171.774 down
<input checked="" type="checkbox"/>	LMOF2365_0936	3.371	10.467	951056..951361	1.75316	3.38779	3.105 down
<input type="checkbox"/>	LMOF2365_0937	0.273	1.457	951366..952634	-1.87447	0.54322	5.343 down

Figure 28: “Genes within the”quick gene set” remain selected in blue (arrows.)”

5.9 Show selected genes

To remove genes that are not in the “quick gene set” from the table, click on the **Choose Quick-Filter** tool

( **Choose Quick-Filter: Showing All Genes**) and select **Show Only Selected Genes**.

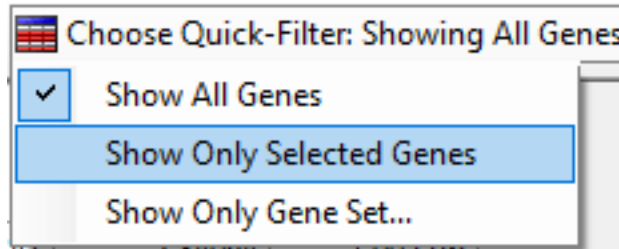


Figure 29: "Show only the"quick gene set" in the table."

5.9.1 Unselect genes

- Click on any row to select it.
- Then **Ctrl+click** (This *could* be **SHIFT+Ctrl+click** on a Mac running Windows) the same row to remove the selection from that row.
- The table should now contain *no blue highlighting*.

5.10 Identify possible operon structures

To identify possible operon structures, scroll down the Gene Table, noting sections where consecutive, or nearly consecutive, genes show similar trends in expression levels and fold changes. One candidate for an operon would be the four overlapping (or adjacent, in one case) genes starting with **LMO~~f~~2365_0912** and ending with **LMO~~f~~2365_0915**.

This also happens to be the location of the **sigB** gene (arrow) :

5.11 More potential operons

Check the list, you may find more, for example:

5.11.1 molybdo-cofactor biosynthesis genes

moeA	9.190	21.931	1072071..1073294	3.20002	4.45488	2.386	down
mobB	6.003	24.406	1073273..1073758	2.58574	4.60917	4.065	down
moaE	5.199	22.291	1073755..1074177	2.37825	4.47840	4.287	down
moaC	11.041	25.622	1074422..1074904	3.46479	4.67930	2.320	down
moaA	4.094	23.153	1074933..1075934	2.03364	4.53315	5.654	down
moaB	10.216	26.778	1076457..1075969	3.35276	4.74299	2.621	down

From Dworkin M., Falkow S., Rosenberg E., Schleifer K.-E., Stackebrandt E. (2006): Molybdopterin Cofactor Biosynthesis.

In *S. carnosus*, nine genes were identified (Fig. 8), all of which appear to be involved in molybdenum cofactor biosynthesis

(Note: As of this writing, the book PDF can be downloaded from the Springer web site <http://link.springer.com/referencework/10.1007%2F0-387-30744-3>)

5.11.2 opuC

opuCD	9.620	28.024	1437146..1436475	3.26596	4.80862	2.913	down
opuCC	4.612	17.954	1438087..1437161	2.20553	4.16620	3.892	down

✓	Name	sigB - linear total RPKM	wt - linear total RPKM	Target Range	sigB - log ₂ total RPKM	wt - log ₂ total RPKM	Fold change
✓	LMOF2365_0838	3.475	13.146	847905..848753	1.79714	3.71658	3.782 down
✓	LMOF2365_0839	4.663	12.031	848863..849420	2.22136	3.58864	2.579 down
✓	LMOF2365_0899	0.021	15.625	918946..920334	-5.60450	3.96582	760.243 down
✓	LMOF2365_0912	35.262	72.516	929818..930162	5.14004	6.18024	2.056 down
✓	LMOF2365_0913	21.493	46.366	930146..930619	4.42581	5.53501	2.157 down
✓	sigB ←	20.826	65.607	930597..931376	4.38032	6.03577	3.150 down
✓	LMOF2365_0915	28.249	73.315	931377..931976	4.82014	6.19603	2.595 down
✓	LMOF2365_0933	3.023	10.963	947792..948295	1.59604	3.45463	3.626 down
✓	gabD	0.212	36.417	949446..950912	-2.23782	5.18655	171.774 down
✓	LMOF2365_0936	3.371	10.467	951056..951361	1.75316	3.38779	3.105 down
✓	LMOF2365_0958	0.175	34.109	975291..975130	-2.51362	5.09206	194.777 down
✓	LMOF2365_0973	1.329	80.619	987897..988130	0.41047	6.33305	60.655 down
✓	LMOF2365_1015	0.149	45.222	1025760..1025392	-2.74370	5.49895	302.891 down
✓	moeA	9.190	21.931	1072071..1073294	3.20002	4.45488	2.386 down
✓	mobB	6.003	24.406	1073273..1073758	2.58574	4.60917	4.065 down
✓	moaE	5.199	22.291	1073755..1074177	2.37825	4.47840	4.287 down
✓	moaC	11.041	25.622	1074422..1074904	3.46479	4.67930	2.320 down
✓	moaA	4.094	23.153	1074933..1075934	2.03364	4.53315	5.654 down
✓	moaB	10.216	26.778	1076457..1075969	3.35276	4.74299	2.621 down
✓	LMOF2365_1147	2.641	24.972	1152338..1151931	1.40089	4.64226	9.456 down

Figure 30: “Potential operon structures.”

B) Genetic organization of the molybdo-cofactor biosynthesis genes

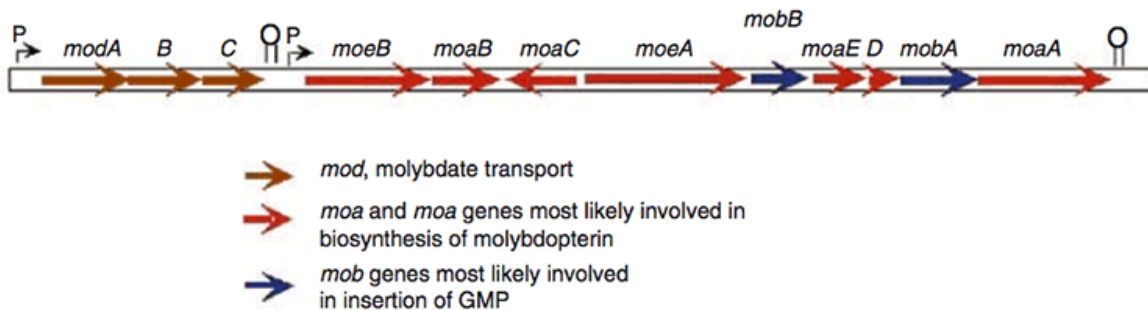


Fig. 8. Genetic organization of genes and operons involved in the dissimilatory nitrate/nitrite reduction in *Staphylococcus carnosus*.

Figure 31: “Figure 8 from Dworkin M., Falkow S., Rosenberg E., Schleifer K.-E., Stackebrandt E. (2006). *Molybdo-cofactor biosynthesis genes*.”

opuCB	7.037	15.689	1438745..1438089	2.81504	3.97171	2.229	down
opuCA	7.067	19.397	1439942..1438749	2.82103	4.27777	2.744	down

The operon, designated *opuC*, consists of four genes which are predicted to encode an ATP binding protein (OpuCA), an extracellular substrate binding protein (OpuCC), and two membrane-associated proteins presumed to form the permease (OpuCB and OpuCD). The operon is preceded by a potential SigB-dependent promoter. (Fraser et al. 2000)

6 Appendix

6.1 Appendix A: Data

The data used in the tutorial was published in (Oliver et al. 2009) and is available for download on the Gene Expression Omnibus (GEO) under accession number GSE15651⁶

Oliver et al. (2009) info:

Experiment type Expression profiling by high throughput sequencing

Summary The stationary phase stress response transcriptome of the human bacterial pathogen *Listeria monocytogenes* was defined using RNA sequencing (RNA-Seq) with the Illumina Genome Analyzer. Specifically, bacterial transcriptomes were compared between stationary phase cells of *L. monocytogenes* 10403S and an otherwise isogenic DsigB mutant, which does not express the alternative sigma factor **sigma B**, a major regulator of genes contributing to stress response.

Keywords Transcriptome and differential expression analyses

Overall design a laboratory strain, 10403S and its otherwise isogenic mutant lacking **sigB** were analyzed. Two replicates of each strain were analyzed for a total of 4 runs

The four sample files have been renamed on the DNASTAR web site. The names on the GEO site are labeled as:

File Name	Replicate name
GSM391674	10403S_replicate1
GSM391675	DsigB_replicate1
GSM391676	10403S_replicate2
GSM391677	DsigB_replicate2

6.2 Appendix B: SigmaB

SigmaB definition (Raengpradub, Wiedmann, and Boor 2008) A sigma factor is a dissociable protein subunit that directs bacterial RNA polymerase holoenzyme to recognize a promoter sequence upstream of a gene prior to transcription initiation. New associations between alternative sigma factors and core RNA polymerase essentially reprogram promoter recognition specificities of the enzyme in response to changing environmental conditions, thus allowing expression of new sets of target genes appropriate for the conditions.

Sigma B modulates the stress response (Schaik and Abee 2005) The alternative sigma factor sigmaB modulates the stress response of several Gram-positive bacteria, including *Bacillus subtilis* and the food-borne human pathogens *Bacillus cereus*, *Listeria monocytogenes* and *Staphylococcus aureus*. In all these bacteria, sigmaB is responsible for the transcription of genes that can confer stress resistance to the vegetative cell.

The question as to what extent and under which conditions sigmaB is responsible for survival during

⁶<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE15651>

stress has been addressed by phenotypic characterization of *sigB* deletion mutants. These studies revealed that sigmaB is involved in the resistance to a variety of stresses including heat, high osmolarity, high ethanol concentrations, high and low pH, and oxidizing agents [...]. In *L. monocytogenes* and *B. subtilis*, sigmaB was shown to have a role in growth and survival under low temperatures [...].

6.3 Appendix C: Sigma B Operon

In *L. monocytogenes*, 168 genes were positively regulated by sigmaB; 145 of these genes were preceded by a putative sigmaB consensus promoter (Raengpradub, Wiedmann, and Boor 2008.)

The genes positively regulated by sigmaB were classified into nine functional categories:

1. Stress
2. Virulence and virulence associated
3. Transcriptional regulation
4. Transport and transport systems
5. Metabolism
6. DNA metabolism and transport
7. Protein synthesis and modification
8. Cell envelope and cellular processes
9. Unknown and hypothetical

6.4 Appendix D: Listeria monocytogenes

L. monocytogenes is a non-spore-forming facultative intracellular pathogen that causes listeriosis, a serious invasive disease in both animals and humans. To establish a food-borne bacterial infection, *L. monocytogenes* must have the ability to survive under a variety of stress conditions, including those encountered in a wide range of nonhost environments and food matrices, as well as under rapidly changing conditions encountered during gastrointestinal passage (exposure to organic acids, bile salts, and osmotic gradients) and subsequent stages of infection (*e.g.*, in the intracellular environment). *L. monocytogenes* sigmaB is activated following exposure to a number of environmental stress conditions [...] and contributes to bacterial survival under acid and oxidative stresses and during carbon starvation [...] (Raengpradub, Wiedmann, and Boor 2008.)

7 Resources

- A survey of best practices for RNA-seq data analysis⁷ Conesa et al. (2016a), Conesa et al. (2016b)
- RNA-seq Analysis Workshop Course Materials⁸
- RNA-seqlopedia⁹
- Standards, Guidelines and Best Practices for RNA-Seq V1.0 (June 2011) The ENCODE Consortium¹⁰
- Guidelines for RNA-Seq data analysis (prot 67)¹¹

⁷<https://genomebiology.biomedcentral.com/articles/10.1186/s13059-016-0881-8>

⁸<http://chagall.med.cornell.edu/RNASEQcourse/>

⁹<http://rnaseq.uoregon.edu/>

¹⁰https://genome.ucsc.edu/ENCODE/protocols/dataStandards/ENCODE_RNAseq_Standards_V1.0.pdf

¹¹<http://www.epigenesys.eu/en/protocols/bio-informatics/1283-guidelines-for-rna-seq-data-analysis>

REFERENCES

- Conesa, A., P. Madrigal, S. Tarazona, D. Gomez-Cabrero, A. Cervera, A. McPherson, M. W. Szczepaniak, et al. 2016a. "A survey of best practices for RNA-seq data analysis." *Genome Biol.* 17 (January): 13. <https://www.ncbi.nlm.nih.gov/pubmed/26813401>.
- . 2016b. "Erratum to: A survey of best practices for RNA-seq data analysis." *Genome Biol.* 17 (1): 181. <https://www.ncbi.nlm.nih.gov/pubmed/27565134>.
- Dworkin M., Falkow S., Rosenberg E., Schleifer K.-E., Stackebrandt E., ed. 2006. *The Prokaryotes A Handbook on the Biology of Bacteria*. 3rd ed. Vol. 4. Bacteria: Firmicutes, Cyanobacteria. New York NY: Springer. <http://link.springer.com/referencework/10.1007%2F0-387-30744-3>.
- Fraser, K. R., D. Harvie, P. J. Coote, and C. P. O'Byrne. 2000. "Identification and characterization of an ATP binding cassette L-carnitine transporter in *Listeria monocytogenes*." *Appl. Environ. Microbiol.* 66 (11): 4696–4704. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC92368/>.
- Oliver, H. F., R. H. Orsi, L. Ponnala, U. Keich, W. Wang, Q. Sun, S. W. Cartinhour, M. J. Filiatrault, M. Wiedmann, and K. J. Boor. 2009. "Deep RNA sequencing of *L. monocytogenes* reveals overlapping and extensive stationary phase and sigma B-dependent transcriptomes, including multiple highly transcribed noncoding RNAs." *BMC Genomics* 10 (December): 641. <http://www.ncbi.nlm.nih.gov/pubmed/11581570>.
- Raengpradub, S., M. Wiedmann, and K. J. Boor. 2008. "Comparative analysis of the sigma B-dependent stress responses in *Listeria monocytogenes* and *Listeria innocua* strains exposed to selected stress conditions." *Appl. Environ. Microbiol.* 74 (1): 158–71. <https://www.ncbi.nlm.nih.gov/pubmed/18024685>.
- Schaik, W. van, and T. Abee. 2005. "The role of sigmaB in the stress response of Gram-positive bacteria – targets for food preservation and safety." *Curr. Opin. Biotechnol.* 16 (2): 218–24. <https://www.ncbi.nlm.nih.gov/pubmed/15831390>.