Metadata for Supplemental Tables from Dissertation of Erin Bredeweg:

“Transcriptional Networks Controlled by the Fatty Acid Regulators, FAR-1 and FAR-2”

Identifications:

<http://hdl.handle.net/1957/44860>

<http://dx.doi.org/10.7267/N9BG2KWH>

1 of 5

Files: TableS(01 through 06)\_FAR(1 or 2)(S or B or O)\_ChIP.xlsx

Data origin: annotation of processed sequencing data

Data type: characters and integers

Instrument used: Illumina Hi-Seq, CASAVA18.1

Data acquisition details: Sequencing was done on Chipped DNA amplified by 21 cycles of PCR using FAR-1 or FAR-2 tagged with an epitope tag (FAR-1:GFP, FAR-2:V5)

File type: xlsx

Data processing methods: bowtie (mapping of 51-nucleotide single end reads), MACS2 (peak identification), IDR (R script) for FAR-1 in sucrose only—assessment of common and significant peaks from multiple replicates, unix (join function), Self-authored annotation script using MACS2 output to provide peak locations, and outputting nearby gene units in the *Neurospora* *crassa* genome, version 12.

*Example MACS2 command: macs2 callpeak -t* /Path/to/mapped/reads/file/sort\_v12\_far1\_120\_FC217.bam -f BAM -g 44000000 -n far1\_gfp-120m\_FC217\_v10

*Example IDR command (while in folder containing the script):*

Rscript batch-consistency-analysis.r /path/to/output/from/MACS2/far1\_gfp-0m\_FC217\_v10\_peaks.narrowPeak /path/to/output/from/MACS2/far1\_gfp-15m\_FC217\_v10\_peaks.narrowPeak -1 /path/to/new/file/created/by/this/program/IDR/chipSampleRep0m\_VS\_chipSampleRep15m 0 F p.value

Datatset parameters:

chr: chromosome or supercontig (sc) of the 7 identified in Neurospora crassa where the peak is found

pk\_start1: base coordinate of start of the peak called by MACS2 from replicate 1

pk\_stop1: base coordinate of stop of the peak called by MACS2 from replicate 1

sig\_value1: significance value calculated by MACS2 based upon local enrichment of

reads relative to background sequence coverage as well as the distribution of the reads

pk\_start2: base coordinate of start of the peak called by MACS2 from replicate 2

pk\_stop2: base coordinate of stop of the peak called by MACS2 from replicate 2

IDR-1: overall value supplied by R program ‘Irreproducible Discovery Rate’ which

compares two MACS2 peak files for overlap and reproducibility. Where a second

replicate was not available, the second file was a repeat of the replicate 1 MACS2 file.

bp-up-IDR1: Distance in nucleotides from the peak location identified for the peak

summit in replicate 1 to the start/end of the gene immediately upstream of the

peak.

chr: chromosome or supercontig (sc) of the 7 identified in Neurospora crassa where the

gene is found

NCU\_up: *Neurospora crassa* gene number of the gene upstream of the peak

up\_ori: orientation (+ for 5’ (left) to 3’ (right) or – for 3’ (left) to 5’ (right)) of upstream

gene

function1: annotated function of upstream gene.

RNA-sig: Total number of comparisons (of 27) with a significant p-value of change in

FPKM (fragments per kilobase per million reads) as assessed by CuffDiff (version 2.1.1, 4/11/2013) of strain to strain (varying carbon sources) or carbon source to carbon source (varying the strains). (sum of S, B, O, WT, dF1, dF2, dd categories)

S: Total number within the sucrose-to-sucrose strain comparisons found to be significant for this transcript of 6 possible

B: Same as for S, but for butyrate to butyrate comparisons

O: Same as for S, but for oleate to oleate comparisons

WT: Total number within the WT-to-WT condition comparisons found to be significant

for this transcript of 3 possible (WT sucrose to WT butyrate, WT sucrose to WT oleate, WT butyrate to WT oleate)

dF1: Same as for WT, but for the delta *far-1* (Δ*far-1*) mutant.

dF2: Same as for WT, but for the delta *far-2* (Δ*far-2*) mutant.

dd: Same as for WT, but for the double Δ*far-1;* Δ*far-2* mutant.

bp\_down-IDR1: Distance in nucleotides from the peak location identified for the peak

summit in replicate 1 to the start/end of the gene immediately downstream of the peak. Where applicable, the same definition can be applied for bp-down-IDR2 for other replicate comparisons.

NCU\_down: *Neurospora crassa* gene number of the gene downstream of the peak

down\_ori: orientation (+ for 5’ (left) to 3’ (right) or – for 3’ (left) to 5’ (right)) of

downstream gene.

function2: annotated function of downstream gene.

\*unique to FAR-1 Sucrose due to the number of replicates assessed by IDR.

pk3-1: base coordinate of start of the peak called by MACS2 from replicate 3

pk3-2: base coordinate of stop of the peak called by MACS2 from replicate 3

pk4-1: base coordinate of start of the peak called by MACS2 from replicate 4

pk4-2: base coordinate of stop of the peak called by MACS2 from replicate 4

2 of 5

Files: TableS07\_FAR1-GO\_ChIP.xlsx, TableS08\_FAR1-GO\_ChIP.xlsx

Data origin: annotation of processed sequencing data, MIPS database

Data type: characters and integers

Instrument used: Illumina Hi-Seq, CASAVA18.1

Data acquisition details: Sequencing was done on Chipped DNA amplified by 21 cycles of PCR using FAR-1 or FAR-2 tagged with an epitope tag (FAR-1:GFP, FAR-2:V5)

File type: xlsx

Data processing methods: bowtie (mapping of 51-nucleotide single end reads), MACS2 (peak identification), IDR (R script) for FAR-1 in sucrose only—assessment of common and significant peaks from multiple replicates, unix (join function), Self-authored annotation script using MACS2 output to provide peak locations, and outputting nearby gene units in the *Neurospora* *crassa* genome, version 12; the resulting NCU lists from annotation of MACS2/IDR output were ranked by significance and the top 1000 unique genes were entered by web submission to determine the functional distribution of the list: <http://mips.helmholtz-muenchen.de/proj/funcatDB/>. For tier 2, this was repeated with the next 1000 NCUs.

Dataset parameters:

Functional Category: A nested classification of functions organized into gene ontology or ‘GO’ categories.

abs SET: Total number of genes in this functional category in the provided NCU list

rel SET: Total proportion of genes in the provided NCU list in this category

genes SET: NCUs of all genes from ‘rel SET’

abs GENOME: Total number of genes in this functional category in *Neurospora crassa*

rel GENOME: Total proportion of genes in *Neurospora crassa* that fall into this category

rel SET/rel GENOME: Ratio of the relative enrichment of the functional category in the

list versus the entire genome.

P-VALUE: estimated significance of the enrichment of this category within the list

relative to the genome.

List of NCUs in each category

3 of 5

Files: TableS09\_MIPS-FAR1-RNAfiltered.xlsx, TableS10\_MIPS-FAR2-RNAfiltered.xlsx

Data origin: annotation of processed sequencing data, MIPS database

Data type: characters and integers

Instrument used: Illumina Hi-Seq, CASAVA18.1

Data acquisition details: Sequencing was done on Chipped DNA amplified by 21 cycles of PCR using FAR-1 or FAR-2 tagged with an epitope tag (FAR-1:GFP, FAR-2:V5)

File type: xlsx

Data processing methods: bowtie (mapping of 51-nucleotide single end reads), MACS2 (peak identification), IDR (R script) for FAR-1 in sucrose only—assessment of common and significant peaks from multiple replicates, unix (join function), Self-authored annotation script using MACS2 output to provide peak locations, and outputting nearby gene units in the *Neurospora* *crassa* genome, version 12. At this point, all NCUs were checked against CuffDiff output for significant change (non-zero RNA-sig columns in TablesS01-S06, TableS11), and retained if at least one of the 27 comparisons showed significantly different FPKM values in the two conditions/strains. The resulting NCU lists from annotation of MACS2/IDR output were ranked by significance and the top 1000 unique genes were entered by web submission to determine the functional distribution of the list: <http://mips.helmholtz-muenchen.de/proj/funcatDB/>. For tier 2, this was repeated with the next 1000 NCUs.

Dataset parameters: same as above (2 of 5)

4 of 5

File: TableS11\_gene\_differential-expression.txt

Data origin: Output of cuffdiff of the tuxedo suite of RNA-sequencing processing programs (<http://cufflinks.cbcb.umd.edu/manual.html>, Trapnell et al. *Nat. Biotechnol*. (2010) 28:511). Comparisons stem from sequencing of 12 different conditions, Neurospora strains (Wild type NM261, Δ*far-1* NMF641*,* Δ*far-2* NMF643*,* Δ*far-1;* Δ*far-2* NMF640*)* grown overnight for 16 hours on 1.5% sucrose Vogel’s media followed by transfer for 5-6 hours into a second carbon source (Vogel’s supplemented with 1.5% surcrose, 5 mM oleate, 5 mM butyrate) and tissue collection. RNA was made (Sokolovsky et al, *Fungal Genet. Newsl.* (1990) 37) and checked on a gel for integrity before mRNA enrichment and RNA-sequencing library preparation using Illumina procedures.

Data type: characters and integers

Instrument used: Illumina Hi-Seq, CASAVA18.1

Data acquisition details: Sequencing was done on cDNA libraries amplified by 13 cycles of PCR, in triplicate.

File type: txt

Data Processing Methods: RNA sequencing samples were mapped using cufflinks followed by differential expression quantitation using cuffdiff. Data for each of the 27 comparisons of strain to strain within each condition (e.g. WT-S to dF1-S) or condition to condition within each strain (e.g. dF1-S to dF1-O) is included, while comparisons with more than one variable were removed.

Dataset parameters:

test\_id and gene\_id: NCU number of the transcript being quantified by FPKM (fragments per kilobase per million reads)

locus: coordinates of the identified transcript

sample\_1: First strain or condition, WT=wild type, dF1= Δ*far-1* mutant, dF2= Δ*far-2* mutant, dd= Δ*far-1;* Δ*far-2* mutant (see above for strain numbers).

sample\_2: Second strain or condition

status: OK: a comparison was done; NOTEST: expression was low or difficult to quantify

value\_1: transcript level of sample\_1 quantified by normalized FPKM in each run and

averaged between replicates.

value\_2: transcript level of sample\_2 quantified by normalized FPKM in each run and

averaged between replicates.

log2(fold\_change): a value of log2(value\_2/value\_1) reflecting the degree of change

that has occurred in expression levels

test\_stat: The value of the test statistic used to compute significance of the observed

change in FPKM

p\_value: uncorrected *p*-value of the test statistic

q\_value: the false discovery rate (FDR) adjusted *p*-value of the test statistic

significant: a yes or no status whether the p-value is beyond the significance cutoff of

the FDR after Benjamini-Hochberg correction for multiple-testing.

5 of 5

File: TableS12\_genes\_fpkm\_tracking.txt

Data origin: Output of cuffdiff of the tuxedo suite of RNA-sequencing processing programs (<http://cufflinks.cbcb.umd.edu/manual.html>, Trapnell et al. *Nat. Biotechnol*. (2010) 28:511). Comparisons stem from sequencing of 12 different conditions, Neurospora strains (Wild type NM261, Δ*far-1* NMF641*,* Δ*far-2* NMF643*,* Δ*far-1;* Δ*far-2* NMF640*)* grown overnight for 16 hours on 1.5% sucrose Vogel’s media followed by transfer for 5-6 hours into a second carbon source (Vogel’s supplemented with 1.5% surcrose, 5 mM oleate, 5 mM butyrate) and tissue collection. RNA was made (Sokolovsky et al, *Fungal Genet. Newsl.* (1990) 37) and checked on a gel for integrity before mRNA enrichment and RNA-sequencing library preparation using Illumina procedures.

Data type: characters and integers

Instrument used: Illumina Hi-Seq, CASAVA18.1

Data acquisition details: Sequencing was done on cDNA libraries amplified by 13 cycles of PCR, in triplicate.

File type: txt

Data Processing Methods: RNA sequencing samples were mapped using cufflinks against version 12 of the *Neurospora crassa* genome, of which this is part of the output. Data for each of the 12 conditions (4 strains in 3 carbon sources) is included. These values are used in the cuffdiff comparisons (TableS11).

Dataset parameters:

tracking\_id: A unique identifier describing the object, the NCU of the transcript.

class\_code: N/A (-)

nearest\_ref\_id: N/A (-)

gene\_id: the gene\_id associated with the object

gene\_short\_name: N/A (-)

tss\_id: N/A (-)

locus: genomic coordinates of the object

length: N/A (-)

coverage: N/A (-)

(strain-condition)\_FPKM: FPKM of the object in sample listed

(strain-condition)\_FPKM\_lo: the lower bound of the 95% confidence interval on the

FPKM of the object in the sample listed

(strain-condition)\_FPKM\_hi: the upper bound of the 95% confidence interval on the

FPKM of the object in the sample listed

(strain-condition)\_status: Quantification status for the object in sample listed. Can be

OK (deconvolution successful), LOWDATA (too complex or shallowly sequenced), HIDATA (too many fragments in locus), or FAIL, when an ill-conditioned covariance matrix or other numerical exception prevents deconvolution.