PWM Enrichment of miRNA

Mark E. Pepin | MD/PhD Trainee

6/5/2018

Identifying Putative Nrf2 Binding Sites

In order to identify the response elements likely targeted by Nfe2l2 (Nrf2) we used the position-weighted matrix (PWM) generated by Brochley *et al.* We scanned the *mus musculus* genome (mm10) for all Nrf2 (Nfe2l2) response elements using the Nfe2l2 PWM via PWMScan (Ambrosini 2018); this process identified 80,756 Nrf2 response elements throughout the mouse genome. These genomic coordinates were then used to determine which differentially-expressed miRNA are likely under the control of Nrf2 based on the presents of Antioxidant Response Elements upstream of the miRNA coding region.

Nfe2l2 PWM (based on Brochley *et al.*) in Transfac format

AC MA0150.2
XX
ID Nfe2l2
XX DE MA0150.2 Nfe2l2 ; From JASPAR 2018
PO A C G T
01 278 23 103 6
02 0 2 0 408
03 0 0 400 10
04 410 0 0 0
05 14 287 74 35
06 52 11 11 336
07 26 315 30 39
08 329 2 47 32
09 3 3 404 0
10 3 398 6 3
11 346 13 15 36
XX
//

Nfe2l2 response Element: Genomic Annotation

Once the Nfe2l2 response elements were identified, we used Hypergeometric Optimization for Motif Enrichment (HOMER) to annotate these genomic regions according to the most proximal gene. The following command was used: `annotatePeaks pwmscan_mm10_40220_40766.bed mm10 > Nfe2l2_REs.Genome.txt`.

```
library(kableExtra)
Nfe2l2_REs_Genome <- read.delim("../1_Input/3_miRNA/Nfe2l2_REs.Genome.txt",
                                header = TRUE, sep = "\t")
Ex<-head(Nfe2l2_REs_Genome)
landscape(Ex %>% knitr::kable(
                          align="c",
                          longtable=T,
                          booktabs=T,
```

```
                          caption="Annotated Response Elements") %>%
    kable_styling(latex_options=c("striped", "repeat_header", "condensed"))
)
```

PeakID..cmd.annotatePeaks.pl.pwmscan_mm10_40220_40766.bed.mm10.	Chr	Start	End	Strand	Peak.Score	Focus.Ratio.Region.Size
ATGACTCAGCA-1912	chr6	82702040	82702050	+	1959	
ATGACTCAGCA-4868	chr17	64662442	64662452	-	1959	int
ATGACTCAGCA-1928	chr6	87853595	87853605	+	1959	
ATGACTCAGCA-405	chr1	191471513	191471523	-	1959	
ATGACTCAGCA-1617	chr5	100753629	100753639	+	1959	
ATGACTCAGCA-4388	chr15	36657700	36657710	-	1959	

miRNA with Proximal Nrf2 Response Elements

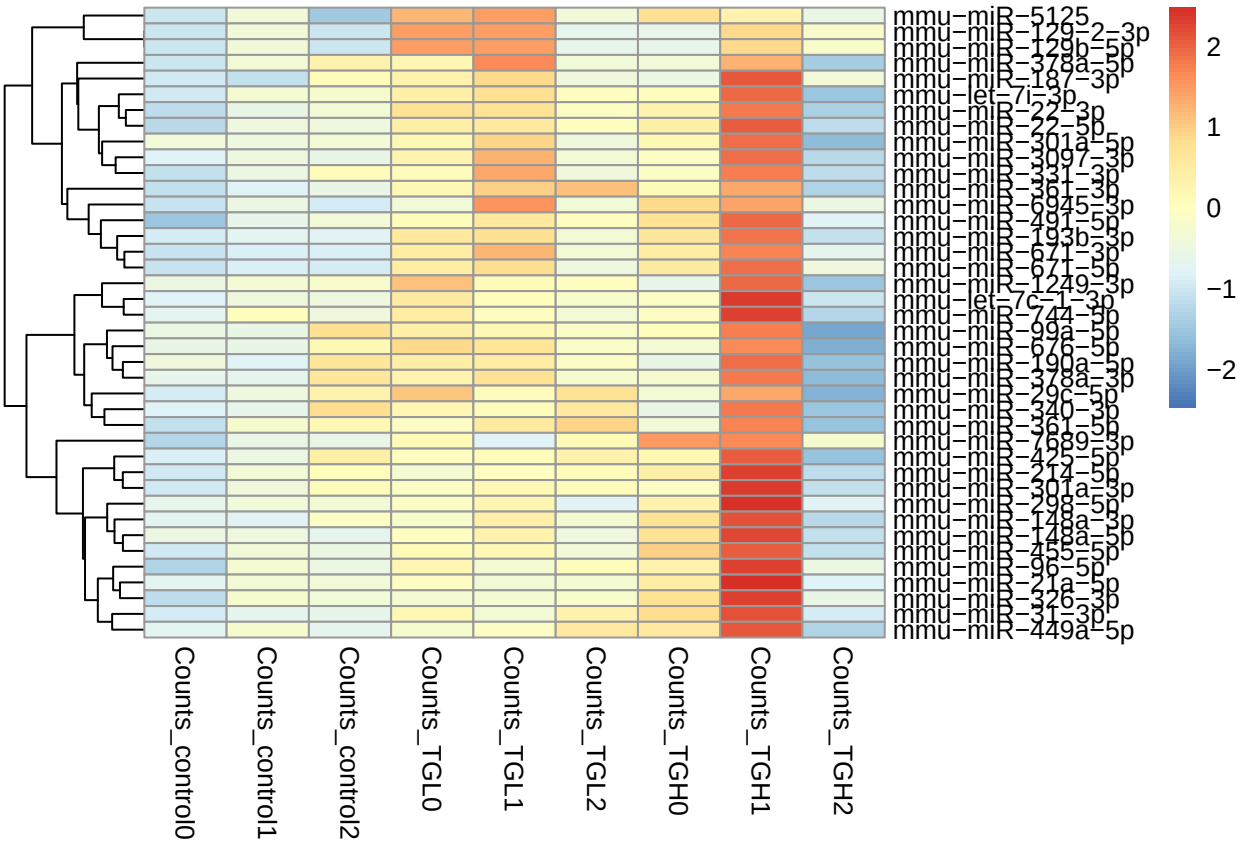
We then inspected the list of miRNA that were only differentially-expressed in the Dose Effect (TgH vs. TgL, excluding Tg vs. NTg). From this, we found that 20 of the 40 differentially-expressed miRNA contained at least 1 Nrf2 response element in the most-proximal promoter region.

```
##Annotated miRNAs (based on genomic coordinates)
miRNA_index_annotated <- read.delim("~/Box/Work/4_PhD/_Papers/Do/Nrf2 Soorappan/1_Input/3_miRNA/miRNA_i
                                    sep = "\t", header = TRUE)
##Annotated AREs (based on genomic coordinates)
Nfe2l2_REs_Genome <- read.delim("../1_Input/3_miRNA/Nfe2l2_REs.Genome.txt",
                                header = TRUE, sep = "\t")
##Differential miRNA expression (based on Tg vs. NTg)
miRNA_TG<-read.csv("../2_Output/miRNA/results_miRNA_TG.v.NTG.csv")
rownames(miRNA_TG)<-miRNA_TG$X
miRNA_TG.UP<-dplyr::filter(miRNA_TG, log2FoldChange>0, pvalue<0.05)

##Merge miRNA and Nfe2l2 based on proximal promoter location
miRNA_AREs<-merge(miRNA_index_annotated, Nfe2l2_REs_Genome, by = "Nearest.PromoterID")

##Filter to only miRNA that are differentially-induced in TG-Effect
Diff.EQ_miRNA.AREs<-merge(miRNA_AREs, miRNA_TG.UP, by.x = "PeakID..cmd.annotatePeaks.pl.miRNA_sorted.tx
                          by.y="X")
write.csv(Diff.EQ_miRNA.AREs, "../3_Results/DiffEq.miRNA containing proximal AREs.csv")
```

```
library(pheatmap)
library(RColorBrewer)
#select only heatmap data
hm.data<-dplyr::select(Diff.EQ_miRNA.AREs, Counts_control0:Counts_TGH2)
hm.data$miRNA.Names<-Diff.EQ_miRNA.AREs$PeakID..cmd.annotatePeaks.pl.miRNA_sorted.txt.mm10.
hm.data<-unique(hm.data)
rownames(hm.data)<-hm.data$miRNA.Names
hm.data<-dplyr::select(hm.data, -miRNA.Names)
#write.csv(hm.data, "../1_Input/3_miRNA/miRNA_Candidates.csv")
#Import heatmap data
data_hm<-read.csv("../1_Input/3_miRNA/miRNA_Candidates.csv")
rownames(data_hm)<-data_hm$X
data_hm<-dplyr::select(data_hm, -X)
data_hm<-data.matrix(data_hm)
#data matrix of counts
pheatmap(data_hm, scale = "row", cluster_cols = FALSE)
```




```
#plot the heatmap
paletteLength <- 100
myColor <- colorRampPalette(c("dodgerblue4", "white", "gold2"))(paletteLength)
pheatmap(data_hm,
         cluster_cols=T,
         border_color=NA,
         cluster_rows=T,
         scale = 'row',
         show_colnames = T,
         show_rownames = T,
         color = myColor,
         filename = "../3_Results/miRNA/miRNA.w.AREs_Heatmap.tiff")
```

```
library(multiMiR)
##create miRNA list
miRNA_list<-rownames(data_hm)
miRNA_targets.all<-get_multimir(org = "mmu", mirna = miRNA_list, table = "predicted", summary = TRUE)
```

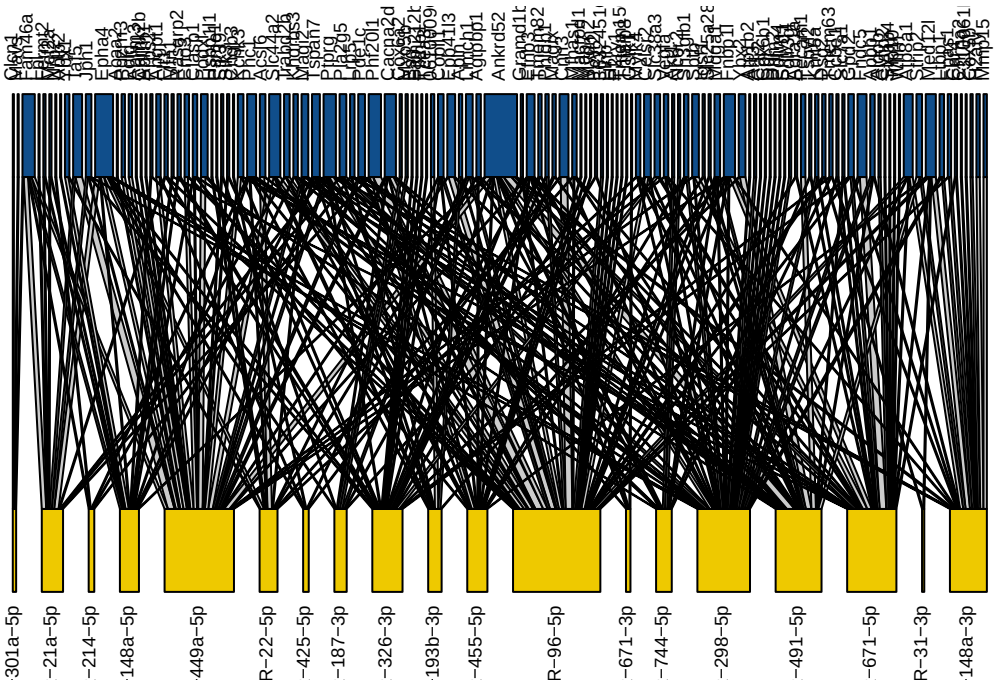
```
## Searching diana_microt ...
## Searching elmmo ...
## Searching microcosm ...
## Searching miranda ...
## Searching mirdb ...
## Searching pictar ...
## Searching pita ...
## Searching targetscan ...
```

```
miRNA_targets.all_table<-miRNA_targets.all@data
#Load the mRNA differentially suppressed in Tg and dose-dependent on Nrf2
library(readxl)
mRNA.both.up_p01 <- read_excel("../3_Results/mRNA/mRNA.Dose_Venn.p01.xlsx",
    sheet = "BOTH_DOWN")
miRNA_Targets.annotated<-dplyr::inner_join(miRNA_targets.all_table,
                                           mRNA.both.up_p01,
                                           by=c("target_ensembl"="ensembl_gene_id"))
write.csv(miRNA_Targets.annotated,
          "../3_Results/miRNA/Predicted miRNA Targets.csv")
miR_heatmap<-read_xlsx("../3_Results/miRNA/miRNA_Targets.annotated.xlsx",
                       sheet="mRNA Targets - Dose.ALL")
rownames(miR_heatmap)<-miR_heatmap$Gene.Symbol
hm_miR.targets<-dplyr::select(miR_heatmap, -Gene.Symbol)
hm_miR.targets<-data.matrix(hm_miR.targets)
#plot the heatmap
paletteLength <- 100
myColor <- colorRampPalette(c("dodgerblue4", "white", "gold2"))(paletteLength)
pheatmap(hm_miR.targets,
         cluster_cols=T,
         border_color=NA,
         cluster_rows=T,
         scale = 'row',
         show_colnames = T,
         show_rownames = T,
         color = myColor,
         filename = "../3_Results/miRNA/Suppressed mRNA Targets of Dose-Dependent miRNA.tiff")
```

miRNA-mRNA Network Analysis

The next component of this analysis was to determine whether upstream miRNA have overlapping predicted downstream mRNA targets. To accomplish this aim, a bipartite network analysis was created to depict (1) the differentially-expressed miRNA containing a proximal upstream Nrf2 response element, and (2) their downstream predicted mRNA targets identified via multiMiR.

```
library(bipartite)
#Create the occurence matrix from multiMiR output
library(readxl)
table<-read_xlsx("../3_Results/miRNA/miRNA_Targets.annotated.xlsx",
                 sheet="Dose.Dependent miRNA")
table_interactions<-dplyr::select(table, Gene.Symbol, mature_mirna_id)
matrix<-as.data.frame(table(table_interactions))
occurrence.matrix<-tidyr::spread(matrix, Gene.Symbol, Freq)
rownames(occurrence.matrix)<-occurrence.matrix$mature_mirna_id
occurrence.matrix<-dplyr::select(occurrence.matrix, -mature_mirna_id)
bipartite_in<-data.matrix(occurrence.matrix)
plotweb(bipartite_in, col.low = "gold2", col.high = "dodgerblue4", text.rot = 90)
```



Supplemental Table: R Session Information

All packages and setting are acquired using the following command:

```
library(kableExtra)
sinfo<-devtools::session_info()
sinfo$platform
```

```
##  setting  value
##  version  R version 3.4.2 (2017-09-28)
##  system   x86_64, darwin15.6.0
##  ui       X11
##  language (EN)
##  collate  en_US.UTF-8
##  tz       <NA>
##  date     2018-06-08
```

```
sinfo$packages %>% kable(
                          align="c",
                          longtable=T,
                          booktabs=T,
                          caption="Packages and Required Dependencies") %>%
    kable_styling(latex_options=c("striped", "repeat_header", "condensed"))
```

Table 2: Packages and Required Dependencies

package	*	version	date	source
AnnotationDbi		1.40.0	2017-10-31	Bioconductor
assertthat		0.2.0	2017-04-11	CRAN (R 3.4.0)
backports		1.1.2	2017-12-13	CRAN (R 3.4.3)

Table 2: Packages and Required Dependencies *(continued)*

package	*	version	date	source
base	*	3.4.2	2017-10-04	local
bindr		0.1.1	2018-03-13	CRAN (R 3.4.2)
bindrcpp	*	0.2.2	2018-03-29	CRAN (R 3.4.4)
Biobase		2.38.0	2017-10-31	Bioconductor
BiocGenerics		0.24.0	2017-10-31	Bioconductor
bipartite	*	2.08	2017-03-31	CRAN (R 3.4.0)
bit		1.1-14	2018-05-29	CRAN (R 3.4.4)
bit64		0.9-7	2017-05-08	CRAN (R 3.4.0)
bitops		1.0-6	2013-08-17	CRAN (R 3.4.0)
blob		1.1.1	2018-03-25	CRAN (R 3.4.4)
cellranger		1.1.0	2016-07-27	CRAN (R 3.4.0)
cluster		2.0.7-1	2018-04-09	CRAN (R 3.4.4)
coda		0.19-1	2016-12-08	CRAN (R 3.4.0)
codetools		0.2-15	2016-10-05	CRAN (R 3.4.2)
colorspace		1.3-2	2016-12-14	CRAN (R 3.4.0)
compiler		3.4.2	2017-10-04	local
datasets	*	3.4.2	2017-10-04	local
DBI		1.0.0	2018-05-02	CRAN (R 3.4.2)
devtools		1.13.5	2018-02-18	CRAN (R 3.4.3)
digest		0.6.15	2018-01-28	CRAN (R 3.4.3)
dotCall64		0.9-5.2	2018-01-11	CRAN (R 3.4.3)
dplyr		0.7.5	2018-05-19	CRAN (R 3.4.2)
evaluate		0.10.1	2017-06-24	CRAN (R 3.4.1)
fields		9.6	2018-01-29	CRAN (R 3.4.3)
glue		1.2.0	2017-10-29	CRAN (R 3.4.2)
graphics	*	3.4.2	2017-10-04	local
grDevices	*	3.4.2	2017-10-04	local
grid		3.4.2	2017-10-04	local
gtable		0.2.0	2016-02-26	CRAN (R 3.4.0)
hms		0.4.2	2018-03-10	CRAN (R 3.4.2)
htmltools		0.3.6	2017-04-28	CRAN (R 3.4.0)
httr		1.3.1	2017-08-20	CRAN (R 3.4.1)
igraph		1.2.1	2018-03-10	CRAN (R 3.4.4)
IRanges		2.12.0	2017-10-31	Bioconductor
kableExtra	*	0.9.0	2018-05-21	CRAN (R 3.4.2)
knitr	*	1.20	2018-02-20	CRAN (R 3.4.3)
lattice	*	0.20-35	2017-03-25	CRAN (R 3.4.2)
magrittr		1.5	2014-11-22	CRAN (R 3.4.0)
maps		3.3.0	2018-04-03	CRAN (R 3.4.4)
MASS		7.3-50	2018-04-30	CRAN (R 3.4.2)
Matrix		1.2-14	2018-04-09	CRAN (R 3.4.4)
memoise		1.1.0	2017-04-21	CRAN (R 3.4.0)
methods	*	3.4.2	2017-10-04	local
mgcv		1.8-23	2018-01-15	CRAN (R 3.4.3)
multiMiR	*	1.1.0	2018-02-10	Bioconductor
munsell		0.4.3	2016-02-13	CRAN (R 3.4.0)

Table 2: Packages and Required Dependencies *(continued)*

package	*	version	date	source
network	*	1.13.0.1	2018-04-02	CRAN (R 3.4.4)
nlme		3.1-137	2018-04-07	CRAN (R 3.4.4)
parallel		3.4.2	2017-10-04	local
permute	*	0.9-4	2016-09-09	CRAN (R 3.4.0)
pheatmap	*	1.0.10	2018-05-19	CRAN (R 3.4.2)
pillar		1.2.3	2018-05-25	CRAN (R 3.4.4)
pkgconfig		2.0.1	2017-03-21	CRAN (R 3.4.0)
plyr		1.8.4	2016-06-08	CRAN (R 3.4.0)
purrr		0.2.5	2018-05-29	CRAN (R 3.4.4)
R6		2.2.2	2017-06-17	CRAN (R 3.4.0)
RColorBrewer	*	1.1-2	2014-12-07	CRAN (R 3.4.0)
Rcpp		0.12.17	2018-05-18	CRAN (R 3.4.2)
RCurl		1.95-4.10	2018-01-04	CRAN (R 3.4.3)
readr		1.1.1	2017-05-16	CRAN (R 3.4.0)
readxl	*	1.1.0	2018-04-20	CRAN (R 3.4.4)
rlang		0.2.1	2018-05-30	CRAN (R 3.4.4)
rmarkdown		1.9	2018-03-01	CRAN (R 3.4.3)
rprojroot		1.3-2	2018-01-03	CRAN (R 3.4.3)
RSQLite		2.1.1	2018-05-06	CRAN (R 3.4.2)
rstudioapi		0.7	2017-09-07	CRAN (R 3.4.1)
rvest		0.3.2	2016-06-17	CRAN (R 3.4.0)
S4Vectors		0.16.0	2017-10-31	Bioconductor
scales		0.5.0	2017-08-24	CRAN (R 3.4.1)
sna	*	2.4	2016-08-08	CRAN (R 3.4.0)
spam		2.1-4	2018-04-12	CRAN (R 3.4.4)
statnet.common	*	4.1.2	2018-06-05	CRAN (R 3.4.2)
stats	*	3.4.2	2017-10-04	local
stats4		3.4.2	2017-10-04	local
stringi		1.2.2	2018-05-02	CRAN (R 3.4.2)
stringr		1.3.1	2018-05-10	CRAN (R 3.4.2)
tibble		1.4.2	2018-01-22	CRAN (R 3.4.3)
tidyr		0.8.1	2018-05-18	CRAN (R 3.4.2)
tidyselect		0.2.4	2018-02-26	CRAN (R 3.4.3)
tools		3.4.2	2017-10-04	local
utils	*	3.4.2	2017-10-04	local
vegan	*	2.5-2	2018-05-17	CRAN (R 3.4.4)
viridisLite		0.3.0	2018-02-01	CRAN (R 3.4.2)
withr		2.1.2	2018-03-15	CRAN (R 3.4.4)
XML		3.98-1.11	2018-04-16	CRAN (R 3.4.4)
xml2		1.2.0	2018-01-24	CRAN (R 3.4.3)
yaml		2.1.19	2018-05-01	CRAN (R 3.4.2)