

hw1: metagenomics

Maria Firulyova

Load libraries

```
library(data.table)
library(dplyr)
```

Read data

```
g_d <- fread('mystery_bins-GENE-DETECTION.txt') %>% tibble::column_to_rownames(var = 'key')
colnames(g_d) <- gsub('_', '.', colnames(g_d))
g_c <- fread('mystery_bins-GENE-COVERAGES.txt')
colnames(g_c) <- gsub('_', '.', colnames(g_c))
```

Get genes which are not detected in some individuals

```
target_genes <- g_d[apply(g_d, 1, function(x) sum(x) < 4),]
knitr::kable(target_genes)
```

	ANDREA	EMILY	EVAN	ISAAC	KAREN
3161	0.0370370	1	1	1	0.0518519
3162	0.0041667	1	1	1	0.0114583

Get coverage values for chosen genes

```
knitr::kable(g_c %>% filter(grepl('3161|3162', key)))
```

key	ANDREA	EMILY	EVAN	ISAAC	KAREN
3161	0.0962963	26.63333	18.03704	4.962963	0.2740741
3162	0.0041667	28.35833	17.37917	4.080208	0.0531250