Dr.deCL's questions

Peter Wu

3/23/2018

Goal: Answer Dr.deCL's questions

```
# part I: rediscover pathways

## queuosine biosynthesis pwy fis, queA, tgt, yajc are in Nichols' secD, secF
## -- not in Nichols'
ECK_1st_table[ECK_1st_table$associated_gene_names %in% c("fis", "queA", "tgt",
    "yajC"), ]
```

```
##         ids Original Name sorted_ECK_missing_gene_names_added
## 525     525   ECK3248-FIS                         ECK3248-FIS
## 1052   1052  ECK0401-YAJC                        ECK0401-YAJC
## 2024   2024  ECK0399-QUEA                        ECK0399-QUEA
## 2117   2117   ECK0400-TGT                         ECK0400-TGT
##      associated_gene_names EcoCycID other.synonyms
## 525                    fis  EG10317            nbp
## 1052                  yajC  EG11096
## 2024                  queA  EG10812           tsaA
## 2117                   tgt  EG10996      vacC(S.f.)
```

```
## id for those genes: 525,1052,2024,2117 This table shows the correlations
sort.pcc.sql.NoIdent[(sort.pcc.sql.NoIdent$strain1 %in% c(525, 1052, 2024, 2117)) &
    (sort.pcc.sql.NoIdent$strain2 %in% c(525, 1052, 2024, 2117)), ]
```

```
##         strain1 strain2 Pearson.Correlation.Coefficient
## 4184046    1052    2117                      0.156486820
## 2087020     525    2024                      0.132543722
## 2086048     525    1052                      0.129431623
## 8051634    2024    2117                      0.114491923
## 2087113     525    2117                      0.008260818
## 4183953    1052    2024                     -0.222288784
```

```
## sulfoquinovose degradation I pwy yihS,yihT,yihU,yihV,yihW are all in
## Nichols' (Note: yihW is colored differently than the other genes in this
## pwy. In ecoliWiki yihW is a predicted DNA-binding transcriptional
## regulator. Maybe that's the reason?)
ECK_1st_table[ECK_1st_table$associated_gene_names %in% c("yihS", "yihT", "yihU",
    "yihV", "yihW"), ]
```

```
##         ids Original Name sorted_ECK_missing_gene_names_added
## 1436   1436  ECK3874-YIHT                        ECK3874-YIHT
## 1525   1525  ECK3875-YIHU                        ECK3875-YIHU
## 2690   2690  ECK3873-YIHS                        ECK3873-YIHS
## 3196   3196  ECK3877-YIHW                        ECK3877-YIHW
## 3270   3270  ECK3876-YIHV                        ECK3876-YIHV
##      associated_gene_names EcoCycID other.synonyms
## 1436                  yihT  EG11846           squT
## 1525                  yihU  EG11847           squU
```

```
## 2690                     yihS  EG11845                squS
## 3196                     yihW  EG11849
## 3270                     yihV  EG11848
```

id for those genes: 1436, 1525, 2690, 3196, 3270 This table shows the
correlations

```
sort.pcc.sql.NoIdent[(sort.pcc.sql.NoIdent$strain1 %in% c(1436, 1525, 2690,
    3196, 3270)) & (sort.pcc.sql.NoIdent$strain2 %in% c(1436, 1525, 2690,
    3270)), ]
```

```
##          strain1 strain2 Pearson.Correlation.Coefficient
## 5711390     1436    1525                     0.448443958
## 5713061     1436    3196                     0.246341504
## 5713135     1436    3270                     0.233833012
## 12716175    3196    3270                     0.188260978
## 6067266     1525    3270                     0.154173577
## 6066686     1525    2690                     0.107678647
## 6067192     1525    3196                     0.066208803
## 10702801    2690    3270                    -0.004373082
## 5712555     1436    2690                    -0.011264246
## 10702727    2690    3196                    -0.034306011
```

ygbJ,ygbK,ygbL (Their functions are all unknown) ygbJ,ygbK,ygbL are all in
Nichols' (Note: yihW is colored differently than the other genes in this
pwy. In ecoliWiki yihW is a predicted DNA-binding transcriptional
regulator. Maybe that's the reason?)

```
ECK_1st_table[ECK_1st_table$associated_gene_names %in% c("ygbJ", "ygbK", "ygbL"),
    ]
```

```
##       ids Original Name sorted_ECK_missing_gene_names_added
## 798   798   ECK2731-YGBJ                        ECK2731-YGBJ
## 919   919   ECK2733-YGBL                        ECK2733-YGBL
## 1125 1125   ECK2732-YGBK                        ECK2732-YGBK
##      associated_gene_names EcoCycID other.synonyms
## 798                   ygbJ    G7417
## 919                   ygbL    G7419
## 1125                  ygbK    G7418
```

id for those genes: 798, 919, 1125 This table shows the correlations

```
sort.pcc.sql.NoIdent[(sort.pcc.sql.NoIdent$strain1 %in% c(798, 919, 1125)) &
    (sort.pcc.sql.NoIdent$strain2 %in% c(798, 919, 1125)), ]
```

```
##         strain1 strain2 Pearson.Correlation.Coefficient
## 3172182     798     919                      0.05707873
## 3172388     798    1125                      0.04424527
## 3653847     919    1125                     -0.03914287
```

Conclusion: genes in the 2 pathways and the other 3 genes are not that
correlated