05_PQM1_Statistics

March 23, 2022

```python
[1]: import pandas as pd
     import numpy as np
     import scipy
     import matplotlib as mpl
     import matplotlib.pyplot as plt
     import seaborn as sns
     from matplotlib import rc

     rc('text', usetex=False)
     rc('text.latex', preamble=r'\usepackage{cmbright}')
     rc('font', **{'family': 'sans-serif', 'sans-serif': ['Helvetica']})

     %matplotlib inline

     # This enables SVG graphics inline.
     %config InlineBackend.figure_formats = {'png', 'retina'}

     rc = {'lines.linewidth': 2,
           'axes.labelsize': 25,
           'axes.titlesize': 25,
           'axes.facecolor': 'DFDFE5'}
     sns.set_context('notebook', rc=rc)
     sns.set_style("dark")

     mpl.rcParams['xtick.labelsize'] = 18
     mpl.rcParams['ytick.labelsize'] = 18
     mpl.rcParams['legend.fontsize'] = 20
```

```python
[2]: res = pd.read_csv('../data/master_table.tsv', sep='\t', index_col=0)
     res.rename(columns={'logq-pqm1': 'neglogq-pqm1'}, inplace=True)
     cat_type = pd.CategoricalDtype(categories=['I', 'II', 'III', 'IV', 'V', 'X'],␣
      ↪ordered=True)
     res['chromosome'] = res.chromosome.astype(cat_type)

     cat_type = pd.CategoricalDtype(categories=['Not DE in pqm-1', 'DE in pqm-1',␣
      ↪'DE in all'], ordered=True)
     res['Significance-pqm1'] = res['Significance-pqm1'].astype(cat_type)
```

```python
res.sort_values(['Significance-pqm1', 'padj-50'], inplace=True)
```

```
[4]: (res['padj-pqm1'] < 0.05).sum()
```

```
[4]: 8835
```

```
[10]: res[(res['padj-pqm1'] < 0.05)].sort_values('log2FoldChange-pqm1').head()
```

```
[10]:                 baseMean-50  log2FoldChange-50  lfcSE-50     stat-50  \
      WBGene00002026    84.368376           0.754205  0.215818   3.494642
      WBGene00009692   205.617351          -0.128245  0.628330  -0.204105
      WBGene00002017   339.744829           0.706777  0.098047   7.208543
      WBGene00002016    47.832731           0.447647  0.260663   1.717341
      WBGene00002018    57.945524           0.540141  0.239997   2.250615

                          pvalue-50        padj-50 genename-50  neglogq-50  \
      WBGene00002026   4.746988e-04   5.958007e-03      hsp-70    2.224899
      WBGene00009692   8.382715e-01   9.425291e-01     F44E5.5    0.025705
      WBGene00002017   5.655362e-13   3.743467e-11   hsp-16.11   10.426726
      WBGene00002016   8.591686e-02   2.978437e-01    hsp-16.2    0.526012
      WBGene00002018   2.440993e-02   1.305097e-01   hsp-16.41    0.884357

                      baseMean-58  log2FoldChange-58  …    Sign-58  Sign-pqm1  \
      WBGene00002026   747.271397          -0.149884  …   Negative  Different
      WBGene00009692  3095.211953          -0.335570  …   Negative       Same
      WBGene00002017  2544.936484          -0.012697  …   Negative  Different
      WBGene00002016   502.686510          -0.242398  …   Negative  Different
      WBGene00002018   610.672549          -0.084258  …   Negative  Different

                        Sign-WT  Significance-WT Significance-pqm1      Ratio  Above  \
      WBGene00002026  Different        DE at 50hrs        DE in pqm-1 -0.198731  False
      WBGene00009692       Same        DE at 58hrs        DE in pqm-1  2.616625  False
      WBGene00002017  Different        DE at 50hrs        DE in pqm-1 -0.017964  False
      WBGene00002016  Different        DE at 58hrs        DE in pqm-1 -0.541495  False
      WBGene00002018  Different            Not DE        DE in pqm-1 -0.155992  False

                      MinCountsDetected  logBM-50  logBM-58
      WBGene00002026           1.700802  1.926180  2.873478
      WBGene00009692           1.920426  2.313060  3.490690
      WBGene00002017           0.000000  2.531153  3.405677
      WBGene00002016           1.275804  1.679725  2.701297
      WBGene00002018           1.488438  1.763020  2.785808

      [5 rows x 57 columns]
```

```python
[3]: res[res.externalgenename == 'pqm-1'][[c for c in res.columns if ('log2' in c)
      ↪or ('padj' in c)]]
```

```
[3]:                log2FoldChange-50   padj-50  log2FoldChange-58   padj-58  \
     WBGene00004096              0.16615  0.042771           -0.037965  0.614881

                    log2FoldChange-pqm1     padj-pqm1
     WBGene00004096           -0.467671  2.218613e-10
```

```
[5]: sns.scatterplot('log2FoldChange-pqm1', 'neglogq-pqm1', hue='Significance-pqm1',␣
     ↪data=res)
```

/Users/davidangeles/opt/anaconda3/lib/python3.7/site-
packages/seaborn/_decorators.py:43: FutureWarning: Pass the following variables
as keyword args: x, y. From version 0.12, the only valid positional argument
will be `data`, and passing other arguments without an explicit keyword will
result in an error or misinterpretation.
  FutureWarning

```
[5]: <AxesSubplot:xlabel='log2FoldChange-pqm1', ylabel='neglogq-pqm1'>
```
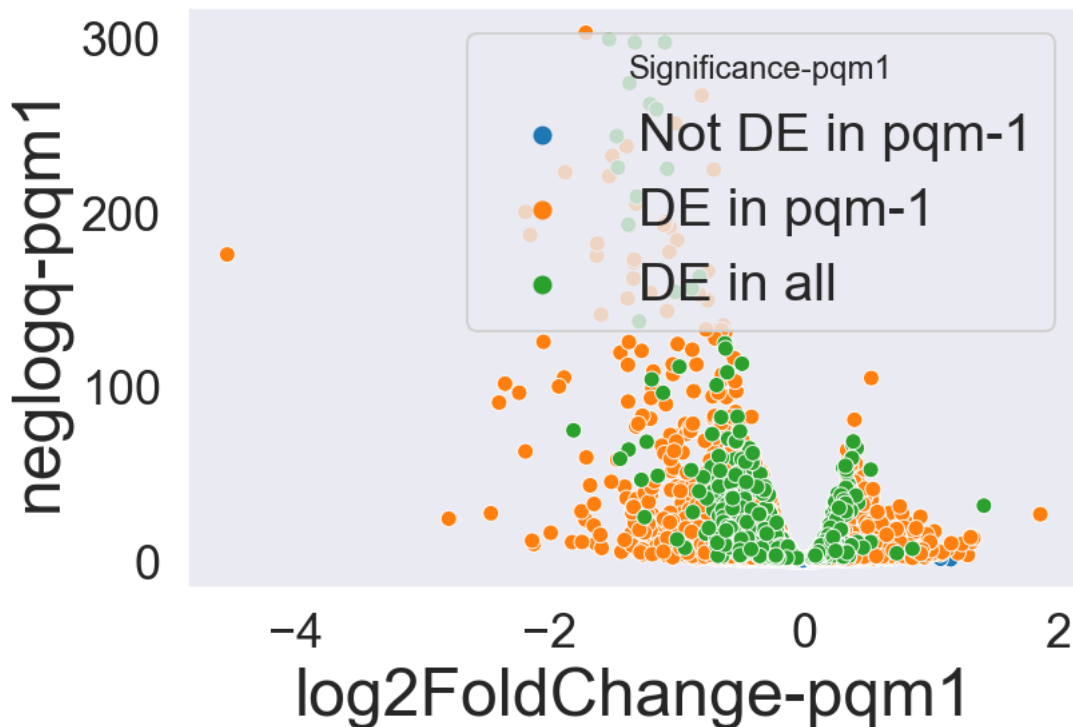


```
[6]: main_cond = (res['padj-50'] < 0.05) & (res['padj-pqm1'] < 0.05 ) &␣
     ↪(res['Sign-WT'] == "Same")
     pqm1_cond = (res.pqm1 == 'Y') & (res['log2FoldChange-50'] > 0)
     g = sns.jointplot(x='log2FoldChange-50', y='log2FoldChange-pqm1', color='black',
                     data=res[main_cond & pqm1_cond], alpha=0.4)
```

```
g.ax_joint.plot([0, .4], [0, .4], color='black', ls='--')
g.ax_joint.axhline(0, color='black')
g.ax_joint.set_xlabel('$\log_2(FC)$, 50hrs')
g.ax_joint.set_ylabel('$\log_2(FC)$, pqm-1')
g.ax_joint.set_ylim(-2.5, .4)
plt.savefig('../figs/pqm1-targets-comparison-50vspqm1.svg', bbox_inches='tight')
```
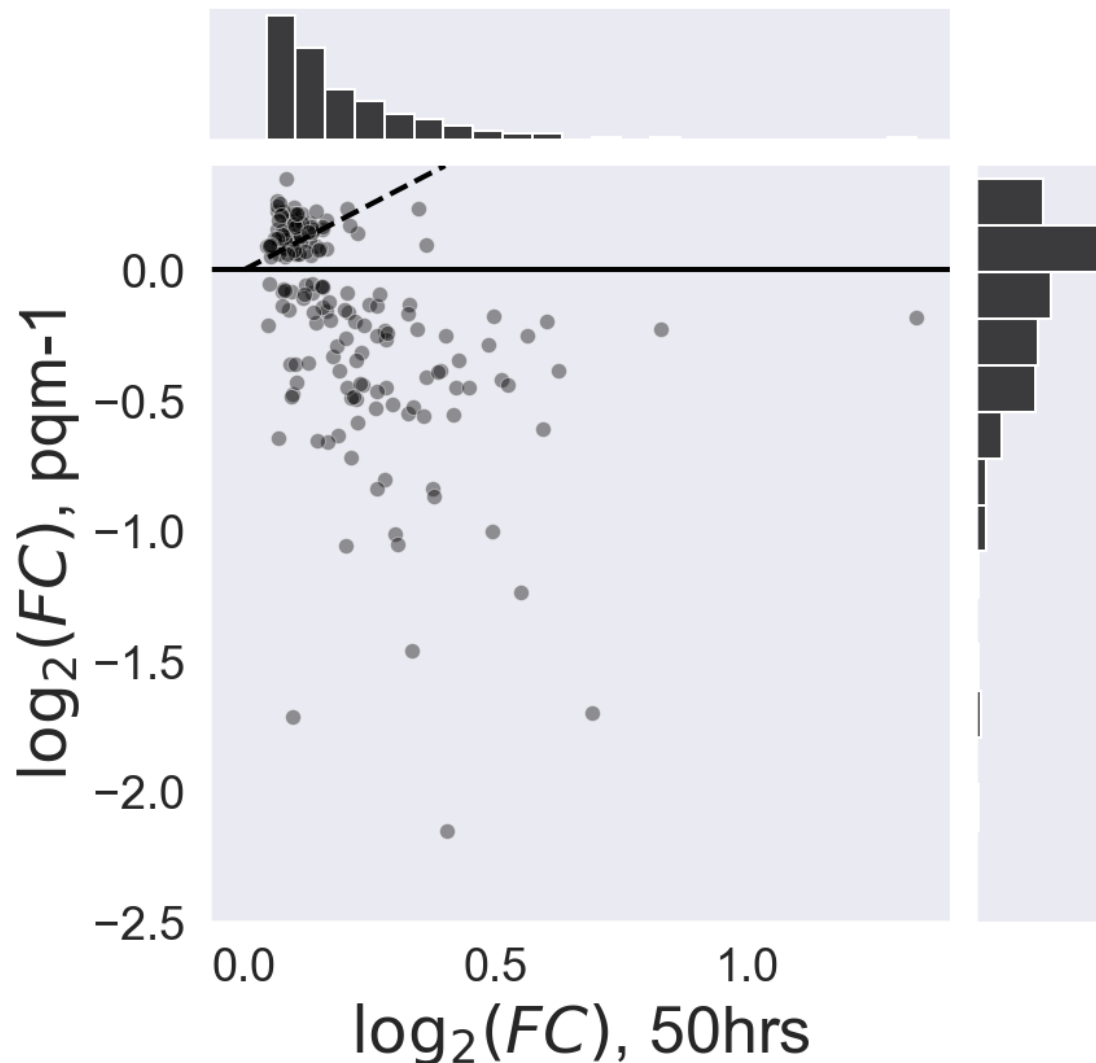


```
[7]: fig, ax = plt.subplots(nrows=2, ncols=2, figsize=(10, 10), sharey=True)

pqm1_cond = (res.pqm1 == 'Y') & (res['log2FoldChange-50'] > 0)
sns.scatterplot(x='log2FoldChange-50', y='log2FoldChange-pqm1', color='red',
                data=res[main_cond & pqm1_cond], alpha=0.4, ax=ax[0, 1])
```

4

```python
pqm1_cond =  (res.pqm1 == 'Y') & (res['log2FoldChange-50'] < 0)
sns.scatterplot(x='log2FoldChange-50', y='log2FoldChange-pqm1', color='black',
                data=res[main_cond & pqm1_cond], alpha=0.4, ax=ax[0, 0])

pqm1_cond =  (res.pqm1 == 'N') & (res['log2FoldChange-50'] > 0)
sns.scatterplot(x='log2FoldChange-50', y='log2FoldChange-pqm1',  color='black',
                data=res[main_cond & pqm1_cond], alpha=0.4, ax=ax[1, 1])

pqm1_cond = (res.pqm1 == 'N') & (res['log2FoldChange-50'] < 0)
sns.scatterplot(x='log2FoldChange-50', y='log2FoldChange-pqm1',  color='black',
                data=res[main_cond & pqm1_cond], alpha=0.4, ax=ax[1, 0])

for axi in ax:
    for a in axi:
        a.axhline(0, color='black', ls='--')
        a.legend([])

ax[0, 0].set_xlim(-1, 0)
ax[1, 0].set_xlim(-1, 0)
ax[0, 1].set_xlim(0, 1)
ax[1, 1].set_xlim(0, 1)

ax[0, 0].set_ylim(-2, 1)
ax[0, 1].set_ylim(-2, 1)
ax[1, 0].set_ylim(-2, 1)
ax[1, 1].set_ylim(-2, 1)

ax[0, 0].set_xticks([])
ax[0, 1].set_xticks([])

ax[1, 0].set_xticks([-1, -.5, 0])
ax[1, 1].set_xticks([0, .5, 1])
ax[0, 0].set_yticks([-2, -1, 0, 1])
ax[0, 1].set_yticks([-2, -1, 0, 1])

ax[0, 0].set_xlabel('')
ax[0, 1].set_xlabel('')

ax[0, 0].set_ylabel('log2(FC), PQM1-bound,\npqm-1(-), 50hrs')
ax[1, 0].set_ylabel('log2(FC), non-PQM1-bound,\npqm-1(-), 50hrs')
ax[1, 0].set_xlabel('log2(FC), wt, 50hrs')
ax[1, 1].set_xlabel('log2(FC), wt, 50hrs')

ax[0, 1].set_xlabel('Putative PQM1-Promoted')

plt.tight_layout()
```
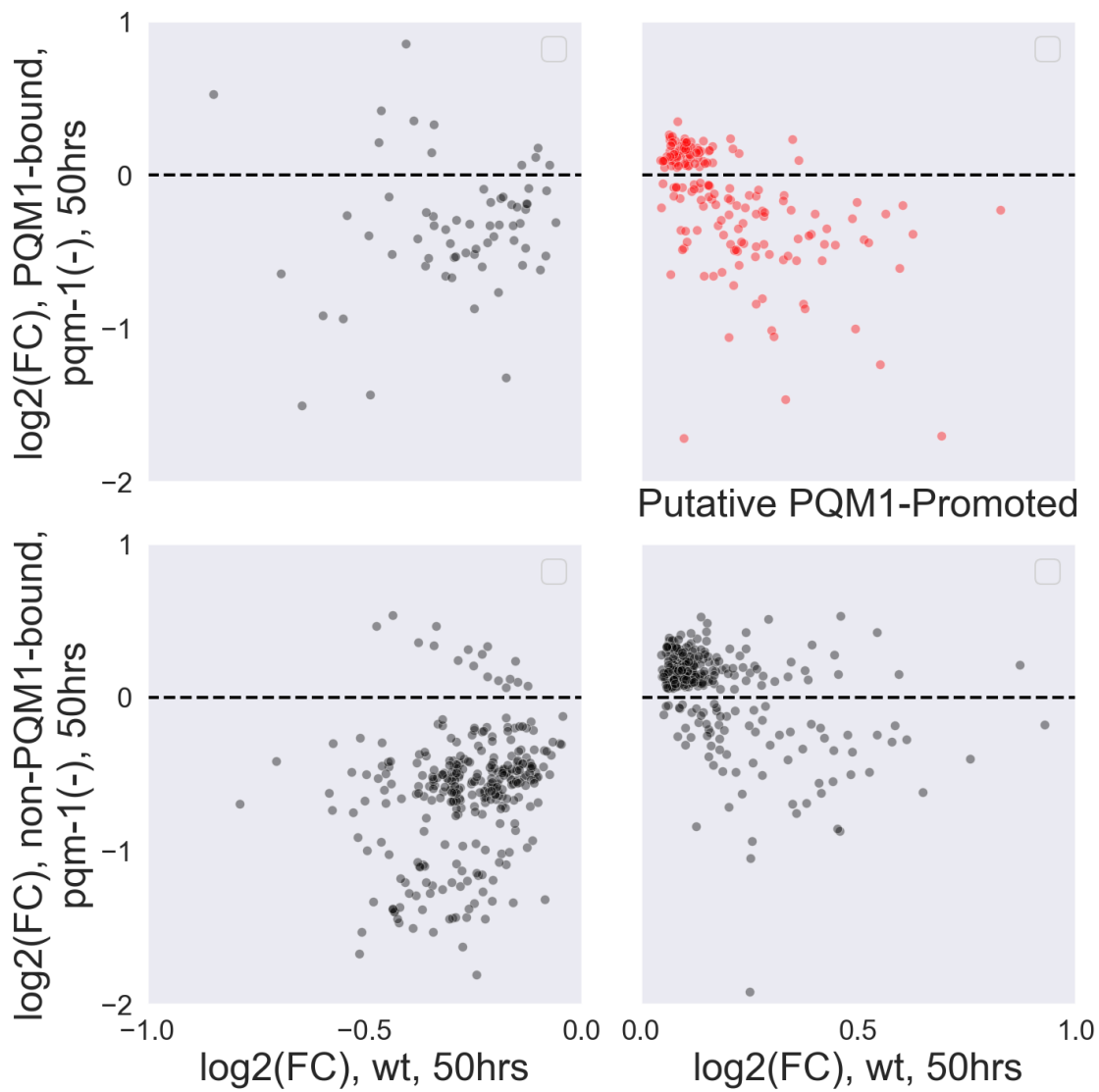
Putative PQM1-Promoted

`[ ]:`