

```

data = readRDS('~/.data/rnaseq_derek/data_from_philip.rds')
grex_names = sapply(colnames(data)[34:ncol(data)],
                    function(x) sprintf('grex%s', x))
colnames(data)[34:ncol(data)] = grex_names
data = data[data$Region=='ACC', grex_names]
# some variables have zero sd after removing one of the regions
sds = apply(data, 2, sd)
keep_me = which(sds>0)
data = data[, keep_me]
cc = cor(data)
M = nrow(cc)
cnt = 0
for (j in 1:M) {
  print(j)
  for (k in 1:M) {
    cnt = cnt + (1 - cc[j, k])**2)
  }
}
meff = 1 + cnt / M
cat(sprintf('Galwey Meff = %.2f\n', meff))

```

OK, now our Meff is around 34K for all 3 cases, so it's definitely more realistic. Also, it doesn't help anything.

But I do have to re-run the summary scripts without the pH variable, so here it is:

```

(base) HG-02035307-LM3:rnaseq_derek sudregp$ Rscript
~/research_code/compile_rnaseq_results.R
resPOPnoPH_ACC_pLT0.10_Diagnosis.csv
[1] "resPOPnoPH_ACC_pLT0.10_Diagnosis.csv"
[1] "Tests with DiagnosisControl p < .05: 3014"
[1] "Tests with DiagnosisControl p < .01: 829"
[1] "Tests with DiagnosisControl q < .05: 3"
[1] "Tests with DiagnosisControl q < .1: 3"
(base) HG-02035307-LM3:rnaseq_derek sudregp$ Rscript
~/research_code/compile_rnaseq_results.R
resWNHnoPH_ACC_pLT0.10_Diagnosis.csv
[1] "resWNHnoPH_ACC_pLT0.10_Diagnosis.csv"
[1] "Tests with DiagnosisControl p < .05: 3369"
[1] "Tests with DiagnosisControl p < .01: 905"
[1] "Tests with DiagnosisControl q < .05: 1"
[1] "Tests with DiagnosisControl q < .1: 1"
(base) HG-02035307-LM3:rnaseq_derek sudregp$ Rscript
~/research_code/compile_rnaseq_results.R
resPOPnoPH_Caudate_pLT0.10_Diagnosis.csv
[1] "resPOPnoPH_Caudate_pLT0.10_Diagnosis.csv"
[1] "Tests with DiagnosisControl p < .05: 3600"
[1] "Tests with DiagnosisControl p < .01: 974"
[1] "Tests with DiagnosisControl q < .05: 1"
[1] "Tests with DiagnosisControl q < .1: 1"
(base) HG-02035307-LM3:rnaseq_derek sudregp$ Rscript

```

```

~/research_code/compile_rnaseq_results.R
resWNHnoPH_Caudate_pLT0.10_Diagnosis.csv
[1] "resWNHnoPH_Caudate_pLT0.10_Diagnosis.csv"
[1] "Tests with DiagnosisControl p < .05: 3319"
[1] "Tests with DiagnosisControl p < .01: 874"
[1] "Tests with DiagnosisControl q < .05: 5"
[1] "Tests with DiagnosisControl q < .1: 10"
(base) HG-02035307-LM3:rnaseq_derek sudregp$ Rscript
~/research_code/compile_rnaseq_results.R
resPOPnoPH_ACC_pLT0.10_DiagnosisAge.csv
[1] "resPOPnoPH_ACC_pLT0.10_DiagnosisAge.csv"
[1] "Tests with DiagnosisControl p < .05: 2492"
[1] "Tests with DiagnosisControl p < .01: 526"
[1] "Tests with DiagnosisControl q < .05: 0"
[1] "Tests with DiagnosisControl q < .1: 1"
[1] "Tests with DiagnosisControl:Age p < .05: 2714"
[1] "Tests with DiagnosisControl:Age p < .01: 661"
[1] "Tests with DiagnosisControl:Age q < .05: 1"
[1] "Tests with DiagnosisControl:Age q < .1: 1"
(base) HG-02035307-LM3:rnaseq_derek sudregp$ Rscript
~/research_code/compile_rnaseq_results.R
resWNHnoPH_ACC_pLT0.10_DiagnosisAge.csv
[1] "resWNHnoPH_ACC_pLT0.10_DiagnosisAge.csv"
[1] "Tests with DiagnosisControl p < .05: 3349"
[1] "Tests with DiagnosisControl p < .01: 709"
[1] "Tests with DiagnosisControl q < .05: 1"
[1] "Tests with DiagnosisControl q < .1: 2"
[1] "Tests with DiagnosisControl:Age p < .05: 3345"
[1] "Tests with DiagnosisControl:Age p < .01: 838"
[1] "Tests with DiagnosisControl:Age q < .05: 2"
[1] "Tests with DiagnosisControl:Age q < .1: 2"
(base) HG-02035307-LM3:rnaseq_derek sudregp$ Rscript
~/research_code/compile_rnaseq_results.R
resPOPnoPH_Caudate_pLT0.10_DiagnosisAge.csv
[1] "resPOPnoPH_Caudate_pLT0.10_DiagnosisAge.csv"
[1] "Tests with DiagnosisControl p < .05: 2642"
[1] "Tests with DiagnosisControl p < .01: 575"
[1] "Tests with DiagnosisControl q < .05: 0"
[1] "Tests with DiagnosisControl q < .1: 0"
[1] "Tests with DiagnosisControl:Age p < .05: 3270"
[1] "Tests with DiagnosisControl:Age p < .01: 798"
[1] "Tests with DiagnosisControl:Age q < .05: 1"
[1] "Tests with DiagnosisControl:Age q < .1: 1"
(base) HG-02035307-LM3:rnaseq_derek sudregp$ Rscript
~/research_code/compile_rnaseq_results.R
resWNHnoPH_Caudate_pLT0.10_DiagnosisAge.csv
[1] "resWNHnoPH_Caudate_pLT0.10_DiagnosisAge.csv"
[1] "Tests with DiagnosisControl p < .05: 2811"
[1] "Tests with DiagnosisControl p < .01: 622"
[1] "Tests with DiagnosisControl q < .05: 0"
[1] "Tests with DiagnosisControl q < .1: 2"
[1] "Tests with DiagnosisControl:Age p < .05: 2596"
[1] "Tests with DiagnosisControl:Age p < .01: 586"
[1] "Tests with DiagnosisControl:Age q < .05: 2"

```

```

[1] "Tests with DiagnosisControl:Age q < .1: 3"
(base) HG-02035307-LM3:rnaseq_derek sudregp$ Rscript
~/research_code/compile_rnaseq_results.R resPOPnoPH_pLT0.10_Diagnosis.csv
[1] "resPOPnoPH_pLT0.10_Diagnosis.csv"
[1] "Tests with DiagnosisControl p < .05: 3092"
[1] "Tests with DiagnosisControl p < .01: 755"
[1] "Tests with DiagnosisControl q < .05: 0"
[1] "Tests with DiagnosisControl q < .1: 0"
(base) HG-02035307-LM3:rnaseq_derek sudregp$ Rscript
~/research_code/compile_rnaseq_results.R resWNHnoPH_pLT0.10_Diagnosis.csv
[1] "resWNHnoPH_pLT0.10_Diagnosis.csv"
[1] "Tests with DiagnosisControl p < .05: 2924"
[1] "Tests with DiagnosisControl p < .01: 667"
[1] "Tests with DiagnosisControl q < .05: 0"
[1] "Tests with DiagnosisControl q < .1: 0"
(base) HG-02035307-LM3:rnaseq_derek sudregp$ Rscript
~/research_code/compile_rnaseq_results.R
resPOPnoPH_pLT0.10_DiagnosisAge.csv
[1] "resPOPnoPH_pLT0.10_DiagnosisAge.csv"
[1] "Tests with DiagnosisControl p < .05: 1988"
[1] "Tests with DiagnosisControl p < .01: 393"
[1] "Tests with DiagnosisControl q < .05: 0"
[1] "Tests with DiagnosisControl q < .1: 0"
[1] "Tests with DiagnosisControl:Age p < .05: 2387"
[1] "Tests with DiagnosisControl:Age p < .01: 474"
[1] "Tests with DiagnosisControl:Age q < .05: 0"
[1] "Tests with DiagnosisControl:Age q < .1: 1"
(base) HG-02035307-LM3:rnaseq_derek sudregp$ Rscript
~/research_code/compile_rnaseq_results.R
resWNHnoPH_pLT0.10_DiagnosisAge.csv
[1] "resWNHnoPH_pLT0.10_DiagnosisAge.csv"
[1] "Tests with DiagnosisControl p < .05: 2865"
[1] "Tests with DiagnosisControl p < .01: 595"
[1] "Tests with DiagnosisControl q < .05: 0"
[1] "Tests with DiagnosisControl q < .1: 1"
[1] "Tests with DiagnosisControl:Age p < .05: 2629"
[1] "Tests with DiagnosisControl:Age p < .01: 597"
[1] "Tests with DiagnosisControl:Age q < .05: 1"
[1] "Tests with DiagnosisControl:Age q < .1: 1"
(base) HG-02035307-LM3:rnaseq_derek sudregp$ Rscript
~/research_code/compile_rnaseq_results.R
resPOPnoPH_pLT0.10_DiagnosisRegion.csv
[1] "resPOPnoPH_pLT0.10_DiagnosisRegion.csv"
[1] "Tests with DiagnosisControl p < .05: 2139"
[1] "Tests with DiagnosisControl p < .01: 473"
[1] "Tests with DiagnosisControl q < .05: 0"
[1] "Tests with DiagnosisControl q < .1: 0"
[1] "Tests with DiagnosisControl:RegionCaudate p < .05: 1578"
[1] "Tests with DiagnosisControl:RegionCaudate p < .01: 296"
[1] "Tests with DiagnosisControl:RegionCaudate q < .05: 0"
[1] "Tests with DiagnosisControl:RegionCaudate q < .1: 0"
(base) HG-02035307-LM3:rnaseq_derek sudregp$ Rscript
~/research_code/compile_rnaseq_results.R
resWNHnoPH_pLT0.10_DiagnosisRegion.csv

```

```

[1] "resWNHnoPH_pLT0.10_DiagnosisRegion.csv"
[1] "Tests with DiagnosisControl p < .05: 2219"
[1] "Tests with DiagnosisControl p < .01: 428"
[1] "Tests with DiagnosisControl q < .05: 0"
[1] "Tests with DiagnosisControl q < .1: 0"
[1] "Tests with DiagnosisControl:RegionCaudate p < .05: 1438"
[1] "Tests with DiagnosisControl:RegionCaudate p < .01: 270"
[1] "Tests with DiagnosisControl:RegionCaudate q < .05: 0"
[1] "Tests with DiagnosisControl:RegionCaudate q < .1: 0"
(base) HG-02035307-LM3:rnaseq_derek sudregp$ Rscript
~/research_code/compile_rnaseq_results.R
resPOPnoPH_pLT0.10_DiagnosisRegionAge.csv
[1] "resPOPnoPH_pLT0.10_DiagnosisRegionAge.csv"
[1] "Tests with DiagnosisControl p < .05: 2167"
[1] "Tests with DiagnosisControl p < .01: 415"
[1] "Tests with DiagnosisControl q < .05: 1"
[1] "Tests with DiagnosisControl q < .1: 1"
[1] "Tests with DiagnosisControl:RegionCaudate p < .05: 2961"
[1] "Tests with DiagnosisControl:RegionCaudate p < .01: 778"
[1] "Tests with DiagnosisControl:RegionCaudate q < .05: 1"
[1] "Tests with DiagnosisControl:RegionCaudate q < .1: 1"
[1] "Tests with DiagnosisControl:Age p < .05: 2216"
[1] "Tests with DiagnosisControl:Age p < .01: 426"
[1] "Tests with DiagnosisControl:Age q < .05: 0"
[1] "Tests with DiagnosisControl:Age q < .1: 0"
[1] "Tests with DiagnosisControl:RegionCaudate:Age p < .05: 2589"
[1] "Tests with DiagnosisControl:RegionCaudate:Age p < .01: 623"
[1] "Tests with DiagnosisControl:RegionCaudate:Age q < .05: 0"
[1] "Tests with DiagnosisControl:RegionCaudate:Age q < .1: 0"
(base) HG-02035307-LM3:rnaseq_derek sudregp$ Rscript
~/research_code/compile_rnaseq_results.R
resWNHnoPH_pLT0.10_DiagnosisRegionAge.csv
[1] "resWNHnoPH_pLT0.10_DiagnosisRegionAge.csv"
[1] "Tests with DiagnosisControl p < .05: 3389"
[1] "Tests with DiagnosisControl p < .01: 600"
[1] "Tests with DiagnosisControl q < .05: 0"
[1] "Tests with DiagnosisControl q < .1: 0"
[1] "Tests with DiagnosisControl:RegionCaudate p < .05: 3463"
[1] "Tests with DiagnosisControl:RegionCaudate p < .01: 891"
[1] "Tests with DiagnosisControl:RegionCaudate q < .05: 0"
[1] "Tests with DiagnosisControl:RegionCaudate q < .1: 5"
[1] "Tests with DiagnosisControl:Age p < .05: 2891"
[1] "Tests with DiagnosisControl:Age p < .01: 572"
[1] "Tests with DiagnosisControl:Age q < .05: 1"
[1] "Tests with DiagnosisControl:Age q < .1: 2"
[1] "Tests with DiagnosisControl:RegionCaudate:Age p < .05: 2743"
[1] "Tests with DiagnosisControl:RegionCaudate:Age p < .01: 605"
[1] "Tests with DiagnosisControl:RegionCaudate:Age q < .05: 2"
[1] "Tests with DiagnosisControl:RegionCaudate:Age q < .1: 2"

```

- play with adding the different covariate domains sequentially
- maybe add PRS?

- add population covariates?