```r
        fit = lm(as.formula(fm_str), data=data)
        res = summary(fit)$coefficients
        # filtering variables
        sig_vars = c()
        for (v in 1:length(test_vars)) {
            # rows in results table that correspond to the screened variable
            var_rows = which(grepl(rownames(res),
                            pattern=sprintf('^%s', test_vars[v])))
            for (r in var_rows) {
                if (res[r, 'Pr(>|t|)'] < pthresh) {
                    sig_vars = c(sig_vars, test_vars[v])
                }
            }
        }
        # factors might get added several times, so here we clean it up
        sig_vars = unique(sig_vars)
        if (length(sig_vars) > 0) {
            clean_fm_str = paste(dep_var, ' ~ ', paste(keep_vars,
collapse='+'), ' + ',
                            paste(sig_vars, collapse='+'), sep="")
        } else {
            clean_fm_str = paste(dep_var, ' ~ ', paste(keep_vars,
collapse='+'), sep="")
        }
        # new model
        clean_fit = lm(as.formula(clean_fm_str), data=data)
        res = data.frame(summary(clean_fit)$coefficients)
        # remove intercept
        res = res[2:nrow(res),]
        res$dep_var = dep_var
        res$formula = clean_fm_str
        res$orig_formula = fm_str
        res$predictor = rownames(res)
        hold = rbind(hold, res)
}
write.csv(hold, file=out_fname, row.names=F)
```

2020-03-18 19:38:05

Let's gather the results again. I had to create a script because it was getting too cumbersome to cut and paste code. So, now we run:

```
Rscript ~/research_code/compile_rnaseq_results.R
resWNH_ACC_pLT0.10_DiagnosisAge.csv
```

```
(base) HG-02035307-LM3:rnaseq_derek sudregp$ Rscript
~/research_code/compile_rnaseq_results.R
resPOP_ACC_pLT0.10_DiagnosisAge.csv
```

```
[1] "resPOP_ACC_pLT0.10_DiagnosisAge.csv"
[1] "Tests with DiagnosisControl p < .05: 2172"
[1] "Tests with DiagnosisControl p < .01: 459"
[1] "Tests with DiagnosisControl p < Meff: 64"
[1] "Tests with DiagnosisControl q < .05: 0"
[1] "Tests with DiagnosisControl q < .1: 2"
[1] "Tests with DiagnosisControl:Age p < .05: 2219"
[1] "Tests with DiagnosisControl:Age p < .01: 473"
[1] "Tests with DiagnosisControl:Age p < Meff: 66"
[1] "Tests with DiagnosisControl:Age q < .05: 1"
[1] "Tests with DiagnosisControl:Age q < .1: 2"
[1] "Tests with DiagnosisControl and DiagnosisControl:Age p < 0.05: 1437"
[1] "Tests with DiagnosisControl and DiagnosisControl:Age p < 0.01: 280"
(base) HG-02035307-LM3:rnaseq_derek sudregp$ Rscript
~/research_code/compile_rnaseq_results.R
resPOP_Caudate_pLT0.10_DiagnosisAge.csv
[1] "resPOP_Caudate_pLT0.10_DiagnosisAge.csv"
[1] "Tests with DiagnosisControl p < .05: 2709"
[1] "Tests with DiagnosisControl p < .01: 609"
[1] "Tests with DiagnosisControl p < Meff: 89"
[1] "Tests with DiagnosisControl q < .05: 0"
[1] "Tests with DiagnosisControl q < .1: 0"
[1] "Tests with DiagnosisControl:Age p < .05: 3309"
[1] "Tests with DiagnosisControl:Age p < .01: 895"
[1] "Tests with DiagnosisControl:Age p < Meff: 116"
[1] "Tests with DiagnosisControl:Age q < .05: 0"
[1] "Tests with DiagnosisControl:Age q < .1: 0"
[1] "Tests with DiagnosisControl and DiagnosisControl:Age p < 0.05: 2119"
[1] "Tests with DiagnosisControl and DiagnosisControl:Age p < 0.01: 469"
(base) HG-02035307-LM3:rnaseq_derek sudregp$ Rscript
~/research_code/compile_rnaseq_results.R resPOP_ACC_pLT0.10_Diagnosis.csv
[1] "resPOP_ACC_pLT0.10_Diagnosis.csv"
[1] "Tests with DiagnosisControl p < .05: 2626"
[1] "Tests with DiagnosisControl p < .01: 726"
[1] "Tests with DiagnosisControl p < Meff: 132"
[1] "Tests with DiagnosisControl q < .05: 2"
[1] "Tests with DiagnosisControl q < .1: 3"
(base) HG-02035307-LM3:rnaseq_derek sudregp$ Rscript
~/research_code/compile_rnaseq_results.R
resPOP_Caudate_pLT0.10_Diagnosis.csv
[1] "resPOP_Caudate_pLT0.10_Diagnosis.csv"
[1] "Tests with DiagnosisControl p < .05: 2966"
[1] "Tests with DiagnosisControl p < .01: 795"
[1] "Tests with DiagnosisControl p < Meff: 125"
[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
resWNH_ACC_pLT0.10_DiagnosisAge.csv
[1] "resWNH_ACC_pLT0.10_DiagnosisAge.csv"
[1] "Tests with DiagnosisControl p < .05: 4181"
[1] "Tests with DiagnosisControl p < .01: 1231"
[1] "Tests with DiagnosisControl p < Meff: 258"
[1] "Tests with DiagnosisControl q < .05: 0"
```

```
[1] "Tests with DiagnosisControl q < .1: 74"
[1] "Tests with DiagnosisControl:Age p < .05: 4198"
[1] "Tests with DiagnosisControl:Age p < .01: 1354"
[1] "Tests with DiagnosisControl:Age p < Meff: 299"
[1] "Tests with DiagnosisControl:Age q < .05: 0"
[1] "Tests with DiagnosisControl:Age q < .1: 110"
[1] "Tests with DiagnosisControl and DiagnosisControl:Age p < 0.05: 3207"
[1] "Tests with DiagnosisControl and DiagnosisControl:Age p < 0.01: 948"
(base) HG-02035307-LM3:rnaseq_derek sudregp$ Rscript
~/research_code/compile_rnaseq_results.R
resWNH_Caudate_pLT0.10_DiagnosisAge.csv
[1] "resWNH_Caudate_pLT0.10_DiagnosisAge.csv"
[1] "Tests with DiagnosisControl p < .05: 2602"
[1] "Tests with DiagnosisControl p < .01: 612"
[1] "Tests with DiagnosisControl p < Meff: 85"
[1] "Tests with DiagnosisControl q < .05: 0"
[1] "Tests with DiagnosisControl q < .1: 0"
[1] "Tests with DiagnosisControl:Age p < .05: 2168"
[1] "Tests with DiagnosisControl:Age p < .01: 498"
[1] "Tests with DiagnosisControl:Age p < Meff: 90"
[1] "Tests with DiagnosisControl:Age q < .05: 1"
[1] "Tests with DiagnosisControl:Age q < .1: 1"
[1] "Tests with DiagnosisControl and DiagnosisControl:Age p < 0.05: 1586"
[1] "Tests with DiagnosisControl and DiagnosisControl:Age p < 0.01: 364"
(base) HG-02035307-LM3:rnaseq_derek sudregp$ Rscript
~/research_code/compile_rnaseq_results.R resWNH_ACC_pLT0.10_Diagnosis.csv
[1] "resWNH_ACC_pLT0.10_Diagnosis.csv"
[1] "Tests with DiagnosisControl p < .05: 3739"
[1] "Tests with DiagnosisControl p < .01: 1492"
[1] "Tests with DiagnosisControl p < Meff: 415"
[1] "Tests with DiagnosisControl q < .05: 110"
[1] "Tests with DiagnosisControl q < .1: 442"
(base) HG-02035307-LM3:rnaseq_derek sudregp$ Rscript
~/research_code/compile_rnaseq_results.R
resWNH_Caudate_pLT0.10_Diagnosis.csv
[1] "resWNH_Caudate_pLT0.10_Diagnosis.csv"
[1] "Tests with DiagnosisControl p < .05: 3357"
[1] "Tests with DiagnosisControl p < .01: 879"
[1] "Tests with DiagnosisControl p < Meff: 144"
[1] "Tests with DiagnosisControl q < .05: 5"
[1] "Tests with DiagnosisControl q < .1: 5"
(base) HG-02035307-LM3:rnaseq_derek sudregp$ Rscript
~/research_code/compile_rnaseq_results.R resPOP_pLT0.10_Diagnosis.csv
[1] "resPOP_pLT0.10_Diagnosis.csv"
[1] "Tests with DiagnosisControl p < .05: 2756"
[1] "Tests with DiagnosisControl p < .01: 740"
[1] "Tests with DiagnosisControl p < Meff: 72"
[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 resWNH_pLT0.10_Diagnosis.csv
[1] "resWNH_pLT0.10_Diagnosis.csv"
[1] "Tests with DiagnosisControl p < .05: 2901"
[1] "Tests with DiagnosisControl p < .01: 698"
```

```
[1] "Tests with DiagnosisControl p < Meff: 55"
[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 resPOP_pLT0.10_DiagnosisAge.csv
[1] "resPOP_pLT0.10_DiagnosisAge.csv"
[1] "Tests with DiagnosisControl p < .05: 2132"
[1] "Tests with DiagnosisControl p < .01: 406"
[1] "Tests with DiagnosisControl p < Meff: 22"
[1] "Tests with DiagnosisControl q < .05: 0"
[1] "Tests with DiagnosisControl q < .1: 0"
[1] "Tests with DiagnosisControl:Age p < .05: 2201"
[1] "Tests with DiagnosisControl:Age p < .01: 449"
[1] "Tests with DiagnosisControl:Age p < Meff: 30"
[1] "Tests with DiagnosisControl:Age q < .05: 0"
[1] "Tests with DiagnosisControl:Age q < .1: 0"
[1] "Tests with DiagnosisControl and DiagnosisControl:Age p < 0.05: 1414"
[1] "Tests with DiagnosisControl and DiagnosisControl:Age p < 0.01: 243"
(base) HG-02035307-LM3:rnaseq_derek sudregp$ Rscript
~/research_code/compile_rnaseq_results.R resWNH_pLT0.10_DiagnosisAge.csv
[1] "resWNH_pLT0.10_DiagnosisAge.csv"
[1] "Tests with DiagnosisControl p < .05: 2965"
[1] "Tests with DiagnosisControl p < .01: 579"
[1] "Tests with DiagnosisControl p < Meff: 35"
[1] "Tests with DiagnosisControl q < .05: 0"
[1] "Tests with DiagnosisControl q < .1: 0"
[1] "Tests with DiagnosisControl:Age p < .05: 2198"
[1] "Tests with DiagnosisControl:Age p < .01: 389"
[1] "Tests with DiagnosisControl:Age p < Meff: 28"
[1] "Tests with DiagnosisControl:Age q < .05: 0"
[1] "Tests with DiagnosisControl:Age q < .1: 0"
[1] "Tests with DiagnosisControl and DiagnosisControl:Age p < 0.05: 1732"
[1] "Tests with DiagnosisControl and DiagnosisControl:Age p < 0.01: 276"
(base) HG-02035307-LM3:rnaseq_derek sudregp$ Rscript
~/research_code/compile_rnaseq_results.R
resPOP_pLT0.10_DiagnosisRegion.csv
[1] "resPOP_pLT0.10_DiagnosisRegion.csv"
[1] "Tests with DiagnosisControl p < .05: 2014"
[1] "Tests with DiagnosisControl p < .01: 503"
[1] "Tests with DiagnosisControl p < Meff: 45"
[1] "Tests with DiagnosisControl q < .05: 0"
[1] "Tests with DiagnosisControl q < .1: 0"
[1] "Tests with DiagnosisControl:RegionCaudate p < .05: 1427"
[1] "Tests with DiagnosisControl:RegionCaudate p < .01: 260"
[1] "Tests with DiagnosisControl:RegionCaudate p < Meff: 16"
[1] "Tests with DiagnosisControl:RegionCaudate q < .05: 0"
[1] "Tests with DiagnosisControl:RegionCaudate q < .1: 0"
[1] "Tests with DiagnosisControl and DiagnosisControl:RegionCaudate p <
0.05: 276"
[1] "Tests with DiagnosisControl and DiagnosisControl:RegionCaudate p <
0.01: 33"
(base) HG-02035307-LM3:rnaseq_derek sudregp$ Rscript
~/research_code/compile_rnaseq_results.R
resWNH_pLT0.10_DiagnosisRegion.csv
```

```
[1] "resWNH_pLT0.10_DiagnosisRegion.csv"
[1] "Tests with DiagnosisControl p < .05: 2217"
[1] "Tests with DiagnosisControl p < .01: 457"
[1] "Tests with DiagnosisControl p < Meff: 32"
[1] "Tests with DiagnosisControl q < .05: 0"
[1] "Tests with DiagnosisControl q < .1: 0"
[1] "Tests with DiagnosisControl:RegionCaudate p < .05: 1243"
[1] "Tests with DiagnosisControl:RegionCaudate p < .01: 232"
[1] "Tests with DiagnosisControl:RegionCaudate p < Meff: 18"
[1] "Tests with DiagnosisControl:RegionCaudate q < .05: 0"
[1] "Tests with DiagnosisControl:RegionCaudate q < .1: 0"
[1] "Tests with DiagnosisControl and DiagnosisControl:RegionCaudate p <
0.05: 287"
[1] "Tests with DiagnosisControl and DiagnosisControl:RegionCaudate p <
0.01: 51"
(base) HG-02035307-LM3:rnaseq_derek sudregp$ Rscript
~/research_code/compile_rnaseq_results.R
resPOP_pLT0.10_DiagnosisRegionAge.csv
[1] "resPOP_pLT0.10_DiagnosisRegionAge.csv"
[1] "Tests with DiagnosisControl p < .05: 2143"
[1] "Tests with DiagnosisControl p < .01: 418"
[1] "Tests with DiagnosisControl p < Meff: 22"
[1] "Tests with DiagnosisControl q < .05: 1"
[1] "Tests with DiagnosisControl q < .1: 1"
[1] "Tests with DiagnosisControl:RegionCaudate p < .05: 2564"
[1] "Tests with DiagnosisControl:RegionCaudate p < .01: 638"
[1] "Tests with DiagnosisControl:RegionCaudate p < Meff: 59"
[1] "Tests with DiagnosisControl:RegionCaudate q < .05: 1"
[1] "Tests with DiagnosisControl:RegionCaudate q < .1: 1"
[1] "Tests with DiagnosisControl:Age p < .05: 2148"
[1] "Tests with DiagnosisControl:Age p < .01: 421"
[1] "Tests with DiagnosisControl:Age p < Meff: 22"
[1] "Tests with DiagnosisControl:Age q < .05: 0"
[1] "Tests with DiagnosisControl:Age q < .1: 0"
[1] "Tests with DiagnosisControl:RegionCaudate:Age p < .05: 2407"
[1] "Tests with DiagnosisControl:RegionCaudate:Age p < .01: 546"
[1] "Tests with DiagnosisControl:RegionCaudate:Age p < Meff: 43"
[1] "Tests with DiagnosisControl:RegionCaudate:Age q < .05: 0"
[1] "Tests with DiagnosisControl:RegionCaudate:Age q < .1: 0"
[1] "Tests with DiagnosisControl and DiagnosisControl:RegionCaudate and
DiagnosisControl:Age and DiagnosisControl:RegionCaudate:Age p < 0.05: 391"
[1] "Tests with DiagnosisControl and DiagnosisControl:RegionCaudate and
DiagnosisControl:Age and DiagnosisControl:RegionCaudate:Age p < 0.01: 43"
(base) HG-02035307-LM3:rnaseq_derek sudregp$ Rscript
~/research_code/compile_rnaseq_results.R
resWNH_pLT0.10_DiagnosisAgeRegion.csv
[1] "resWNH_pLT0.10_DiagnosisAgeRegion.csv"
[1] "Tests with DiagnosisControl p < .05: 3659"
[1] "Tests with DiagnosisControl p < .01: 699"
[1] "Tests with DiagnosisControl p < Meff: 39"
[1] "Tests with DiagnosisControl q < .05: 0"
[1] "Tests with DiagnosisControl q < .1: 0"
[1] "Tests with DiagnosisControl:Age p < .05: 2757"
[1] "Tests with DiagnosisControl:Age p < .01: 483"
```

```
[1] "Tests with DiagnosisControl:Age p < Meff: 35"
[1] "Tests with DiagnosisControl:Age q < .05: 0"
[1] "Tests with DiagnosisControl:Age q < .1: 0"
[1] "Tests with DiagnosisControl:RegionCaudate p < .05: 3150"
[1] "Tests with DiagnosisControl:RegionCaudate p < .01: 821"
[1] "Tests with DiagnosisControl:RegionCaudate p < Meff: 59"
[1] "Tests with DiagnosisControl:RegionCaudate q < .05: 0"
[1] "Tests with DiagnosisControl:RegionCaudate q < .1: 4"
[1] "Tests with DiagnosisControl:Age:RegionCaudate p < .05: 2618"
[1] "Tests with DiagnosisControl:Age:RegionCaudate p < .01: 584"
[1] "Tests with DiagnosisControl:Age:RegionCaudate p < Meff: 38"
[1] "Tests with DiagnosisControl:Age:RegionCaudate q < .05: 0"
[1] "Tests with DiagnosisControl:Age:RegionCaudate q < .1: 2"
[1] "Tests with DiagnosisControl and DiagnosisControl:Age and
DiagnosisControl:RegionCaudate and DiagnosisControl:Age:RegionCaudate p <
0.05: 541"
[1] "Tests with DiagnosisControl and DiagnosisControl:Age and
DiagnosisControl:RegionCaudate and DiagnosisControl:Age:RegionCaudate p <
0.01: 62"
```

Then we can look for some intersections in the lists:

```
sort file1 file2 | uniq -d > out8
```

For example, the intersection of the main result and WNH only:

```
(base) HG-02035307-LM3:rnaseq_derek sudregp$ sort
grexlist_resPOP_pLT0.10_DiagnosisRegionAge_DiagnosisControl:RegionCaudate:
Age_Meff.txt
grexlist_resWNH_pLT0.10_DiagnosisAgeRegion_DiagnosisControl:Age:RegionCaud
ate_Meff.txt | uniq -d
grex13678
grex26766
grex4894
grex9272
(base) HG-02035307-LM3:rnaseq_derek sudregp$ sort
grexlist_resPOP_pLT0.10_Diagnosis_DiagnosisControl_Meff.txt
grexlist_resWNH_pLT0.10_Diagnosis_DiagnosisControl_Meff.txt | uniq -d
grex13047
grex1325
grex21764
grex2308
grex3593
grex6827
```

- play with adding the different covariate domains sequentially
- can Meff be used here safely?