

```

${phen}_${suf};
done
cp ~/data/tmp/polygen_results_${phen}*csv ~/data/heritability_change_rev/

```

```

phen=dti_JHUtracts_ADRDonly_OD0.95
cd ~/data/heritability_change_rev
for m in ad rd; do
  for t in {1..20}; do
    solar run_phen_var_OD_tracts_avgAgeForced ${phen}_moreBaseline
    ${m}_${t};
    solar run_phen_var_OD_tracts_SESandIQForced ${phen}_SESandIQ
    ${m}_${t};
  done;
done;

mv ${phen}_* ~/data/tmp/
for suf in moreBaseline_avgAgeForced SESandIQ_SESandIQForced; do
  cd ~/data/tmp/${phen}_${suf}
  for p in `ls`; do cp $p/polygenic.out ${p}_polygenic.out; done
  python ~/research_code/compile_solar_multivar_results.py
  ${phen}_${suf};
done
cp ~/data/tmp/polygen_results_${phen}*csv ~/data/heritability_change_rev/

```

2020-05-01 13:18:42

Some extra analysis to address comments. First:

"t tests comparing the included and excluded scans -bith fmri and dti. Maybe look at dx, sx, sex, age? Separately and may together?"

```

dti = read.csv('~data/heritability_change_rev/dti_all_scans_status.csv')
fmri =
read.csv('~data/heritability_change_rev/fmri_all_scans_status.csv')
a = dti[, c(1,2,4,9,36,37,34)]
b = fmri[, c(1,2,27,31, 38, 39, 36)]
colnames(a) = c('MRN', 'maskid', 'sex', 'age', 'inatt', 'hi', 'status')
colnames(b) = c('MRN', 'maskid', 'sex', 'age', 'inatt', 'hi', 'status')
data = rbind(a, b)
dused = dti$status=='final_set'
fused = fmri$status=='final_set'
used = data$status=='final_set'

```

```

> t.test(dti[dused,]$age_at_scan...Scan...Subjects,
dti[!dused,]$age_at_scan...Scan...Subjects)

```

Welch Two Sample t-test

```
data: dti[dused, ]$age_at_scan...Scan...Subjects and dti[!dused,
]$age_at_scan...Scan...Subjects
t = -0.62134, df = 876.72, p-value = 0.5345
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.4714639  0.2447305
sample estimates:
mean of x mean of y
 10.34417  10.45753

> t.test(fmri[fused,]$age_at_scan, fmri[!fused,]$age_at_scan)
```

Welch Two Sample t-test

```
data: fmri[fused, ]$age_at_scan and fmri[!fused, ]$age_at_scan
t = 1.1841, df = 521.32, p-value = 0.2369
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.1783233  0.7194785
sample estimates:
mean of x mean of y
 11.56605  11.29547

> t.test(data[used,]$age, data[!used,]$age)
```

Welch Two Sample t-test

```
data: data[used, ]$age and data[!used, ]$age
t = 1.0184, df = 1437.5, p-value = 0.3086
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.1363470  0.4307937
sample estimates:
mean of x mean of y
 10.92188  10.77465
```

Let's try some t-tests for SX as well, but running them as continuous might not be the best approach:

```
> t.test(dti[dused,]$SX_inatt, dti[!dused,]$SX_inatt)
```

Welch Two Sample t-test

```
data: dti[dused, ]$SX_inatt and dti[!dused, ]$SX_inatt
t = -1.252, df = 942.54, p-value = 0.2109
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.6911721  0.1527594
sample estimates:
mean of x mean of y
```

```

3.706349  3.975556

> t.test(fmri[fused,]$SX_inatt, fmri[!fused,]$SX_inatt)

Welch Two Sample t-test

data:  fmri[fused, ]$SX_inatt and fmri[!fused, ]$SX_inatt
t = 0.79338, df = 569.17, p-value = 0.4279
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.2959444  0.6970490
sample estimates:
mean of x mean of y
 3.696903  3.496350

> t.test(data[used,]$inatt, data[!used,]$inatt)

Welch Two Sample t-test

data:  data[used, ]$inatt and data[!used, ]$inatt
t = -0.56649, df = 1555.4, p-value = 0.5711
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.4119654  0.2273333
sample estimates:
mean of x mean of y
 3.701883  3.794199

```

All clear for inatt, so let's check HI:

```

> t.test(dti[dused,]$SX_HI, dti[!dused,]$SX_HI)

Welch Two Sample t-test

data:  dti[dused, ]$SX_HI and dti[!dused, ]$SX_HI
t = -0.011795, df = 942.87, p-value = 0.9906
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.3719632  0.3675188
sample estimates:
mean of x mean of y
 2.555556  2.557778

> t.test(fmri[fused,]$SX_HI, fmri[!fused,]$SX_HI)

Welch Two Sample t-test

data:  fmri[fused, ]$SX_HI and fmri[!fused, ]$SX_HI
t = 0.052315, df = 575.24, p-value = 0.9583
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.4018874  0.4238821

```

```

sample estimates:
mean of x mean of y
 2.219027  2.208029

> t.test(data[used,]$hi, data[!used,]$hi)

Welch Two Sample t-test

data:  data[used, ]$hi and data[!used, ]$hi
t = -0.20704, df = 1559.2, p-value = 0.836
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.3034410  0.2454993
sample estimates:
mean of x mean of y
 2.396444  2.425414

```

All clear as well. Finally, let's do the usual chi-squared test for sex and DX:

```

dti$used = F
dti[dused, 'used'] = T
fmri$used = F
fmri[fused, 'used'] = T
data$used = F
data[used, 'used'] = T

```

```

> t = table(dti$used, dti$Sex...Subjects)
> chisq.test(t)

```

Pearson's Chi-squared test with Yates' continuity correction

```

data:  t
X-squared = 2.5331, df = 1, p-value = 0.1115

```

```

> t = table(fmri$used, fmri$Sex...Subjects)
fmri$Sex
> t = table(fmri$used, fmri$Sex)
> chisq.test(t)

```

Pearson's Chi-squared test with Yates' continuity correction

```

data:  t
X-squared = 1.0907, df = 1, p-value = 0.2963

```

```

> t = table(data$used, data$sex)
> chisq.test(t)

```

Pearson's Chi-squared test with Yates' continuity correction

```
data:  t
X-squared = 4.1347, df = 1, p-value = 0.04201
```

We do see a difference in sex for the entire dataset, but I don't think it's something to worry about... let's check DX:

```
add_DX = function(mres, age_var) {
  mres$DX = NA
  for (r in 1:nrow(mres)) {
    if (mres[r, age_var] < 16) {
      if ((mres[r, 'SX_HI'] >= 6) || (mres[r, 'SX_inatt'] >= 6)) {
        mres[r, 'DX'] = 'ADHD'
      } else {
        mres[r, 'DX'] = 'NV'
      }
    } else {
      if ((mres[r, 'SX_HI'] >= 5) || (mres[r, 'SX_inatt'] >= 5)) {
        mres[r, 'DX'] = 'ADHD'
      } else {
        mres[r, 'DX'] = 'NV'
      }
    }
  }
  mres$DX = factor(mres$DX)
  return(mres)
}
dti2 = add_DX(dti, 'age_at_scan...Scan...Subjects')
fmri2 = add_DX(fmri, 'age_at_scan')
data$SX_HI = data$hi
data$SX_inatt = data$inatt
data2 = add_DX(data, 'age')
```

```
> t = table(dti2$used, dti2$DX)
> chisq.test(t)
```

Pearson's Chi-squared test with Yates' continuity correction

```
data:  t
X-squared = 1.5375, df = 1, p-value = 0.215
```

```
> t = table(fmri2$used, fmri2$DX)
> chisq.test(t)
```

Pearson's Chi-squared test with Yates' continuity correction

```
data:  t
X-squared = 0.9372, df = 1, p-value = 0.333
```

```
> t = table(data2$used, data2$DX)
```

```
> chisq.test(t)
```

Pearson's Chi-squared test with Yates' continuity correction

```
data: t
X-squared = 0.17603, df = 1, p-value = 0.6748
```

Also, no difference in exclusion of NVs and ADHDs.

T tests comparing a Ses and Iq by dx for the entire cohort (those with either rsfmri and/or dti).

```
dti =
read.csv('~data/heritability_change_rev/dti_JHUtracts_ADRDonly_0D0.95_SES
andIQ.csv')
fmri =
read.csv('~data/heritability_change_rev/rsfmri_7by7from100_4nets_p05SigSu
m_0D0.95_12052019_clean_SESandIQ.csv')
a = dti[, c(1, 93:96)]
b = fmri[, c(1, 70, 73, 71, 72)]
colnames(a) = c('ID', 'DX', 'DX2', 'SES', 'FSIQ')
colnames(b) = c('ID', 'DX', 'DX2', 'SES', 'FSIQ')
data = rbind(a,b)
data = data[!duplicated(data$ID), ]
```

```
> t.test(data[data$DX=='NV', 'SES'], data[data$DX=='ADHD', 'SES'])
```

Welch Two Sample t-test

```
data: data[data$DX == "NV", "SES"] and data[data$DX == "ADHD", "SES"]
t = -0.74389, df = 268.85, p-value = 0.4576
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -5.279908  2.384158
sample estimates:
mean of x mean of y
 32.20253  33.65041
```

```
> t.test(data[data$DX=='NV', 'FSIQ'], data[data$DX=='ADHD', 'FSIQ'])
```

Welch Two Sample t-test

```
data: data[data$DX == "NV", "FSIQ"] and data[data$DX == "ADHD", "FSIQ"]
t = 3.3496, df = 275.97, p-value = 0.0009221
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
  2.714006 10.451610
sample estimates:
```

```
mean of x mean of y
114.8205 108.2377
```

The IQ result is concerning, but not unexpected. The results seem to hold when we controlled for it, though.

```
> t.test(data[data$DX2=='NV', 'SES'], data[data$DX2=='ADHD', 'SES'])
```

Welch Two Sample t-test

```
data: data[data$DX2 == "NV", "SES"] and data[data$DX2 == "ADHD", "SES"]
t = -0.6808, df = 258.03, p-value = 0.4966
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
-5.208603 2.532363
sample estimates:
mean of x mean of y
32.07438 33.41250
```

```
> t.test(data[data$DX2=='NV', 'FSIQ'], data[data$DX2=='ADHD', 'FSIQ'])
```

Welch Two Sample t-test

```
data: data[data$DX2 == "NV", "FSIQ"] and data[data$DX2 == "ADHD", "FSIQ"]
t = 2.0305, df = 217.83, p-value = 0.04352
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
0.1264399 8.4873754
sample estimates:
mean of x mean of y
114.3950 110.0881
```

Difference is not as extreme if we use the categories we actually used in the paper, where ADHD is not DSM5 but ≥ 4 SX (DX2).

for the baseline sx and change (and the reverse). Could you add the beta std error t and p to the response

Adding tables to Word document.

Not needed... only needed changes to the text. Sent by Slack.

TODO
