Adjusted Neutrophil Count (ANC) single marker association in PGRNseq data

Here we run a set of single marker tests on the raw phenotype. You can see its distribution in the phenotype processing report. We look at three models: simple regression on genotype on phenotype in european samples only; multiple regression of genotype + covariates in european samples only; and multiple regression of genotype on phenotype + covariates + pricipal components in ALL available samples.

In addition to summarizing the scan in a QQ plot, we happened to have captured a previously associated variant. I report the estimated effect and p-value of the variant for each model.

Known association signal

The following was pulled from the 1000 Genomes browser. UGT1A1*93: rs10929302
hg19 chr2:234,665,782 G/A
1000 Genomes allele frequencies:
A: 27%
G: 73%

The known association was referenced in PharmGKB.

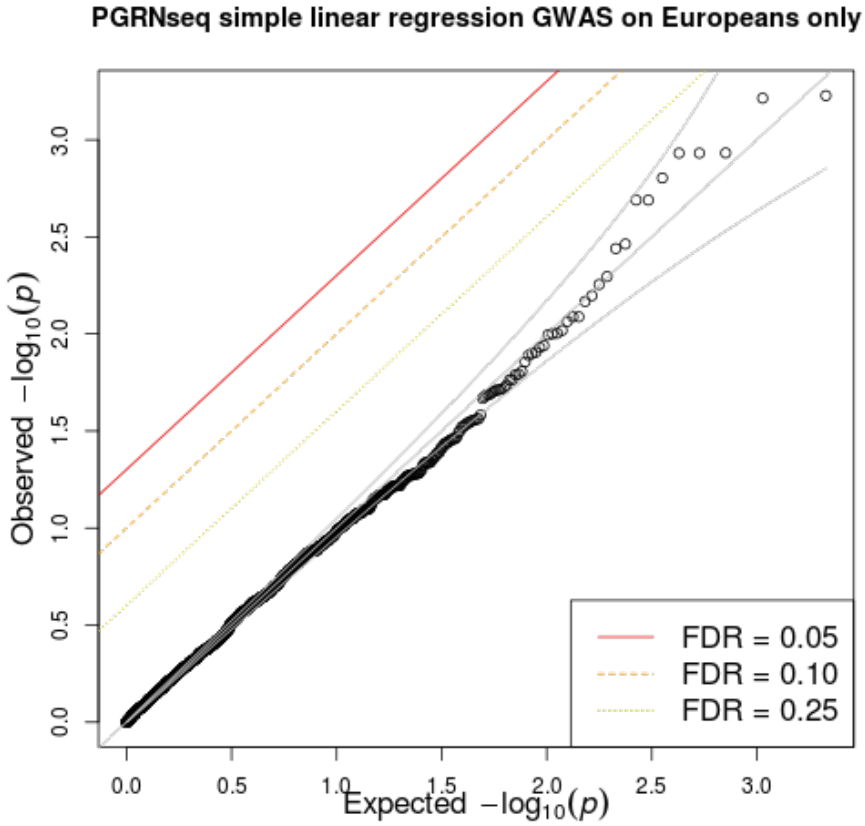
Genetic europeans

These samples were selected based on clustering analysis performed on the top two PCs of the exome chip data. This is a subset consisting of `dim(euro.seq.geno@phdata)[1]` samples. The subsetting is performed to remove any confouding of genetric ancestry, but ultimately may be too conservative for our sample size.

Simple regression

Regress phenotype on genotype.

```
seq.simplereg.results <- mlreg(basic.model, euro.seq.geno, trait="gaussian")
qqunif(seq.simplereg.results[,"P1df"])
title('PGRNseq simple linear regression GWAS on Europeans only')
```



Look at the association results of the previously observed signal:

```
rs10929302.res <- results(seq.simplereg.results)['chr2:234665782:G:A', c('A1', 'A2','N', 'ef
print(xtable(rs10929302.res, digits=6), include.rownames=FALSE)
```

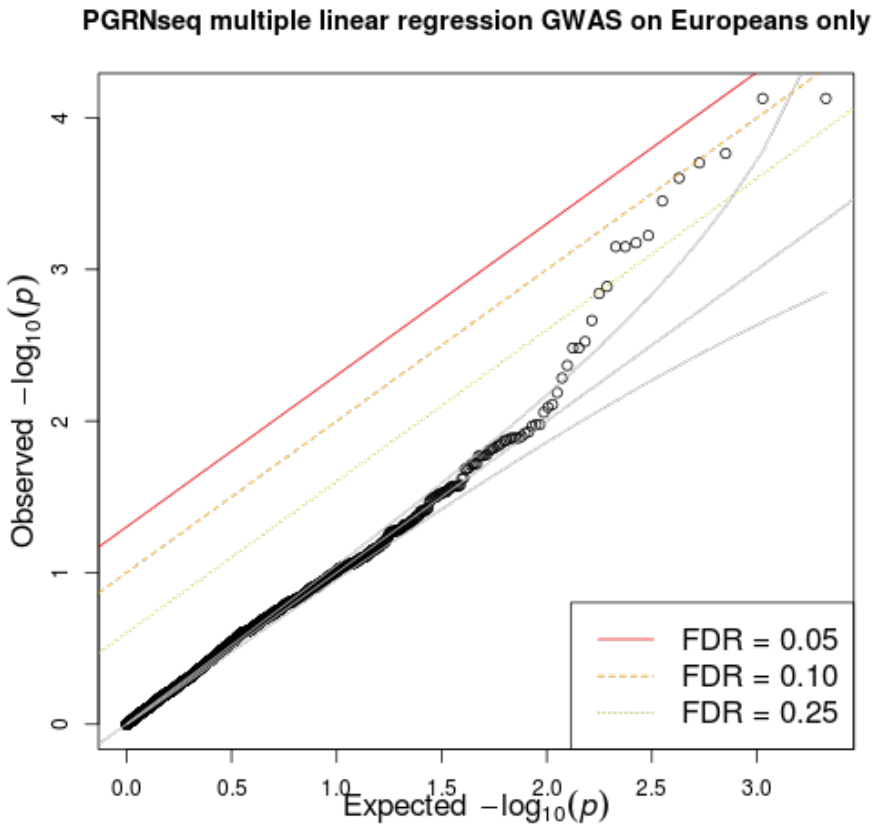
% latex table generated in R 3.1.1 by xtable 1.7-3 package % Wed Aug 27 10:57:20 2014

A1	A2	N	effB	se_effB	P1df
T	G	168.000000	-0.292835	0.122902	0.017188

Multiple regression

Regress phenotype on genotype and other sample covariates.

```
seq.multiplereg.results <- mlreg(full.model, euro.seq.geno, trait="gaussian")
qqunif(seq.multiplereg.results[,"P1df"])
title('PGRNseq multiple linear regression GWAS on Europeans only')
```



Look at the association results of the previously observed signal:

```
rs10929302.res <- results(seq.multiplereg.results)['chr2:234665782:G:A', c('A1', 'A2','N',
print(xtable(rs10929302.res, digits=6), include.rownames=FALSE)
```

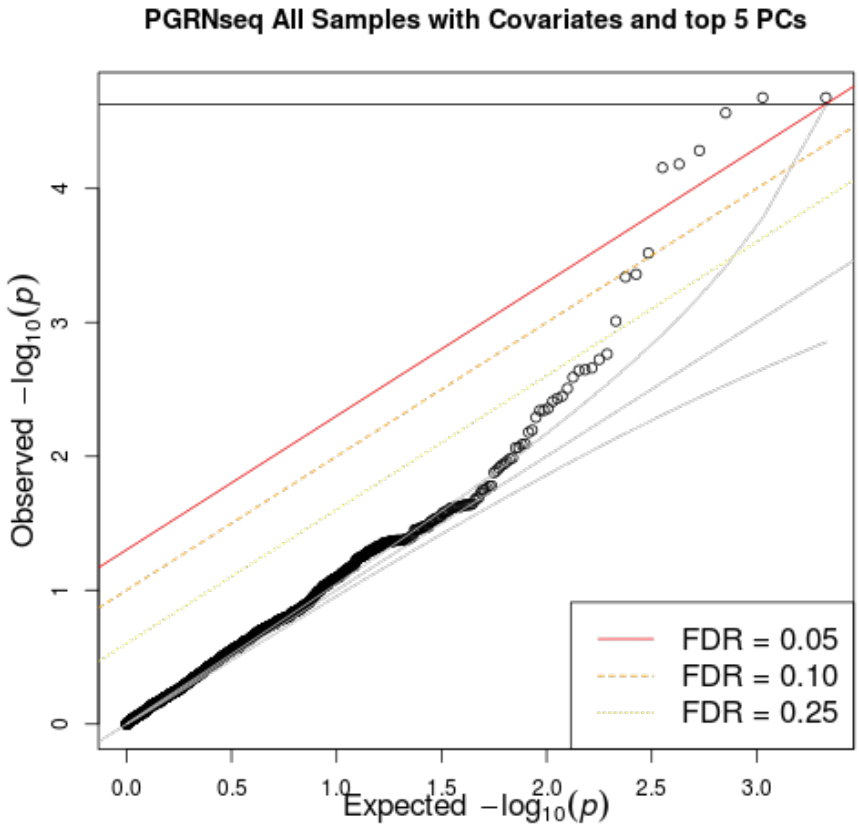
% latex table generated in R 3.1.1 by xtable 1.7-3 package % Wed Aug 27 10:57:20 2014

A1	A2	N	effB	se_effB	P1df
T	G	155.000000	-0.317849	0.124287	0.010546

Full Data with PC adjustment

Regress phenotype on genotype, sample covariates, and top five PCs using *all* samples.

```
seq.pcareg.results <- mlreg(pca.model, full.seq.geno, trait="gaussian")
qqunif(seq.pcareg.results[,"P1df"])
title('PGRNseq All Samples with Covariates and top 5 PCs')
```



Look at the association results of the previously observed signal:

```
rs10929302.res <- results(seq.pcareg.results)['chr2:234665782:G:A', c('A1', 'A2','N', 'effB
print(xtable(rs10929302.res, digits=6), include.rownames=FALSE)
```

% latex table generated in R 3.1.1 by xtable 1.7-3 package % Wed Aug 27 10:57:21 2014

A1	A2	N	effB	se_effB	P1df
T	G	206.000000	-0.336047	0.107192	0.001719