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.

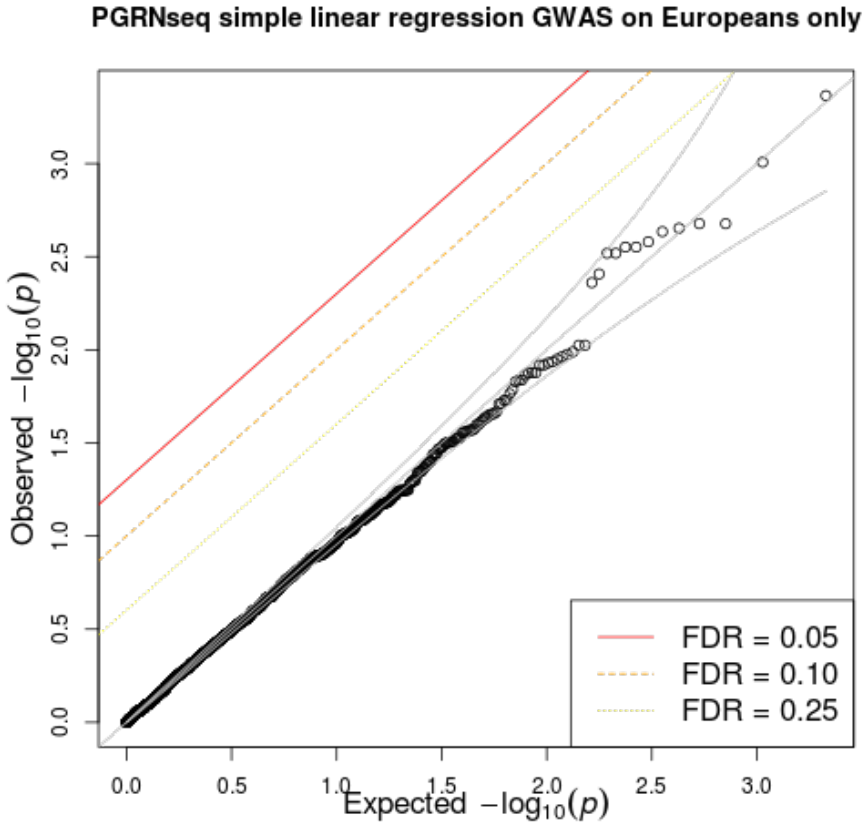
Genetic europeans

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')
```



Take a look at the top hits:

```
top.hits <- seq.simplereg.results[order(seq.simplereg.results[,'P1df'], decreasing=F),]
print(xtable(head(top.hits[ c("N", "effB", "se_effB", "chi2.1df", "P1df") ]), digits=6), inc
```

% latex table generated in R 3.1.1 by xtable 1.7-3 package % Mon Sep 15 10:23:29 2014

	N	effB	se_effB	chi2.1df	P1df
chr10:96749666:T:A	168.000000	0.384511	0.109239	12.389682	0.000432
chr10:96535761:T:G	168.000000	0.351524	0.106673	10.859172	0.000983
chr2:234667496:C:A	168.000000	-0.363291	0.118132	9.457504	0.002103
chr2:234668828:T:C	168.000000	-0.363291	0.118132	9.457504	0.002103
chr1:110256162:C:G	167.000000	-0.505400	0.165236	9.355331	0.002223
chr10:96521681:G:A	168.000000	-0.408833	0.134224	9.277560	0.002320

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)
```

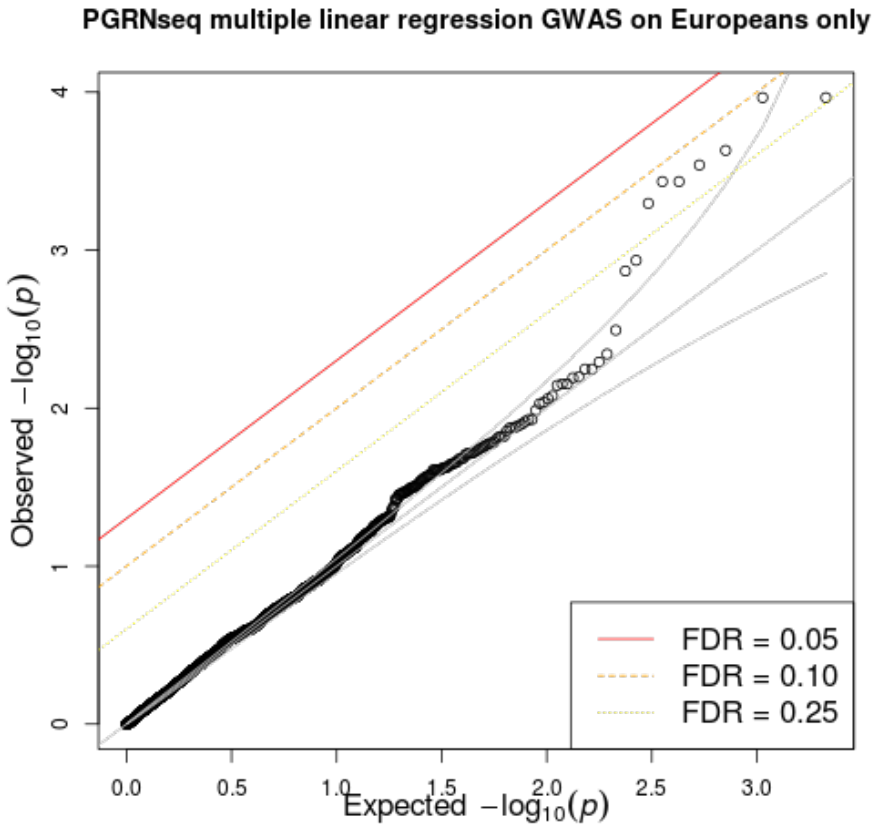
% latex table generated in R 3.1.1 by xtable 1.7-3 package % Mon Sep 15 10:23:29 2014

A1	A2	N	effB	se_effB	P1df
T	G	168.000000	-0.249350	0.116123	0.031771

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')
```



Take a look at top hits:

```
top.hits <- seq.multiplereg.results[order(seq.multiplereg.results[,'P1df'], decreasing=F),]
print(xtable(head(top.hits[ c("N", "effB", "se_effB", "chi2.1df", "P1df") ]), digits=6), inc
```

% latex table generated in R 3.1.1 by xtable 1.7-3 package % Mon Sep 15 10:23:29 2014

	N	effB	se_effB	chi2.1df	P1df
chr2:234667496:C:A	155.000000	-0.457865	0.118268	14.987907	0.000108
chr2:234668828:T:C	155.000000	-0.457865	0.118268	14.987907	0.000108
chr10:96749666:T:A	155.000000	0.415277	0.112856	13.540119	0.000234
chr2:234602202:A:C	154.000000	-0.448412	0.123708	13.138945	0.000289
chr2:234602157:C:T	155.000000	-0.416845	0.117022	12.688569	0.000368
chr2:234602277:G:T	155.000000	-0.416845	0.117022	12.688569	0.000368

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)
```

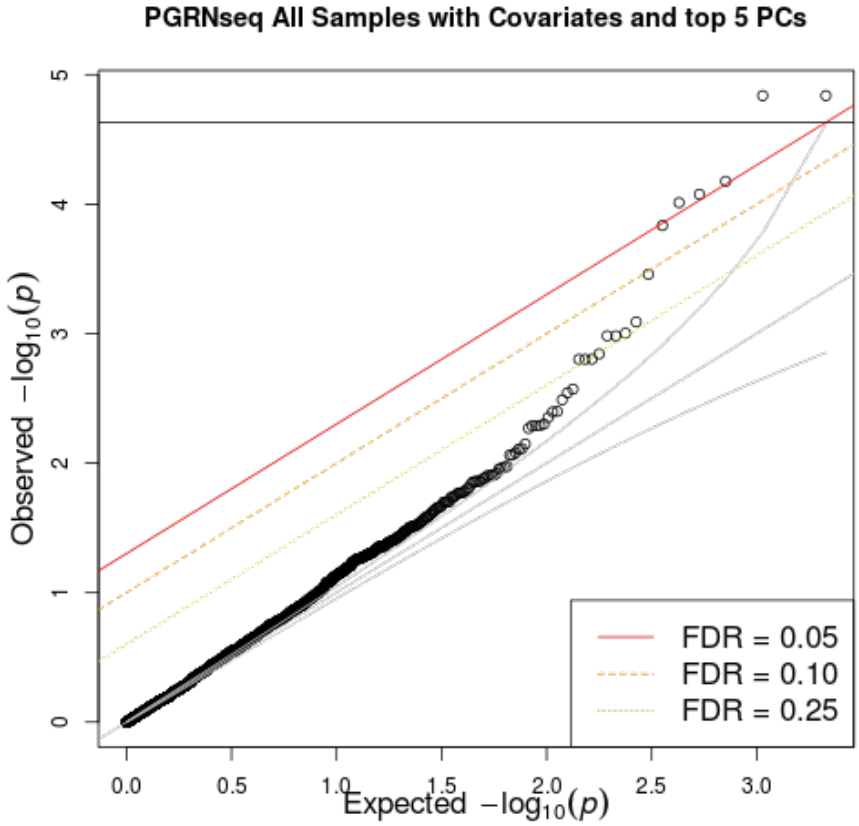
% latex table generated in R 3.1.1 by xtable 1.7-3 package % Mon Sep 15 10:23:29 2014

A1	A2	N	effB	se_effB	P1df
T	G	155.000000	-0.280978	0.117398	0.016694

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')
```



Take a look at top hits:

```
top.hits <- seq.pcareg.results[order(seq.pcareg.results[,'P1df'], decreasing=F),]
print(xtable(head(top.hits[ c("N", "effB", "se_effB", "chi2.1df", "P1df") ]), digits=6), inc
```

% latex table generated in R 3.1.1 by xtable 1.7-3 package % Mon Sep 15 10:23:30 2014

	N	effB	se_effB	chi2.1df	P1df
chr2:234667496:C:A	206.000000	-0.444287	0.102461	18.802278	0.000014
chr2:234668828:T:C	206.000000	-0.444287	0.102461	18.802278	0.000014
chr2:234602277:G:T	206.000000	-0.408373	0.102395	15.905963	0.000067
chr2:234602157:C:T	206.000000	-0.399695	0.101628	15.467755	0.000084
chr2:234665942:G:T	206.000000	-0.405008	0.103919	15.189186	0.000097
chr2:234602202:A:C	205.000000	-0.411600	0.108390	14.420145	0.000146

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 % Mon Sep 15 10:23:30 2014

A1	A2	N	effB	se_effB	P1df
T	G	206.000000	-0.291090	0.101094	0.003984