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.

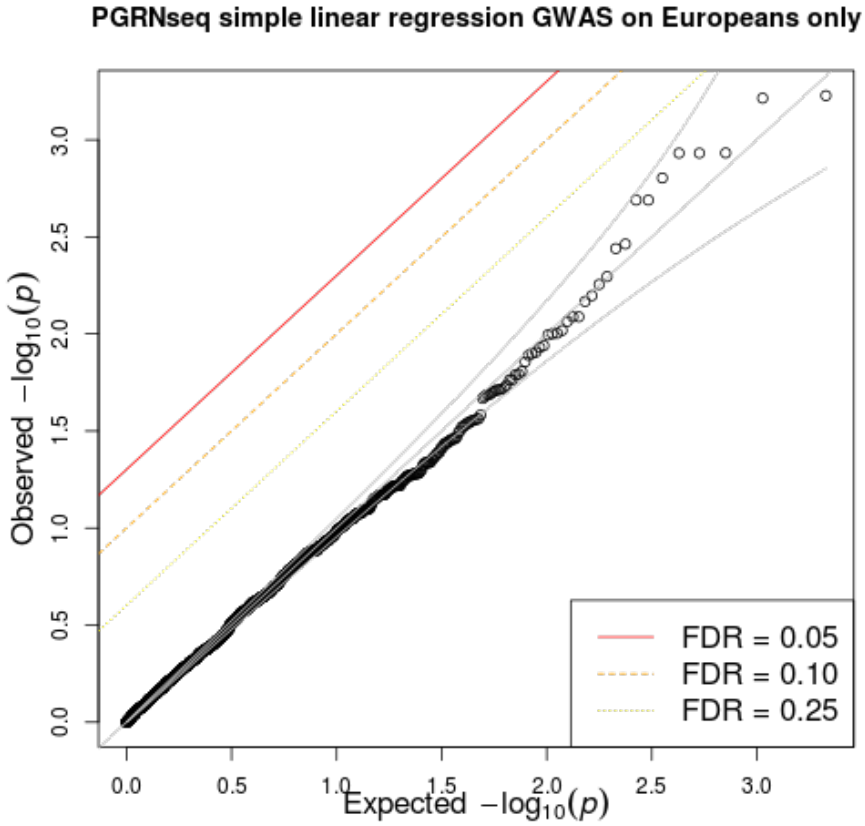
Genetic europeans

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 10:16:21 2014

	N	effB	se_effB	chi2.1df	P1df
chr11:74870884:A:G	168.000000	-0.465523	0.135501	11.803029	0.000591
chr10:96749666:T:A	168.000000	0.398204	0.116187	11.746233	0.000610
chr10:96535761:T:G	168.000000	0.368013	0.113355	10.540042	0.001168
chr2:234667496:C:A	168.000000	-0.405849	0.125034	10.535875	0.001171
chr2:234668828:T:C	168.000000	-0.405849	0.125034	10.535875	0.001171
chr1:47276726:C:T	168.000000	-0.436095	0.137959	9.992230	0.001572

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)
```

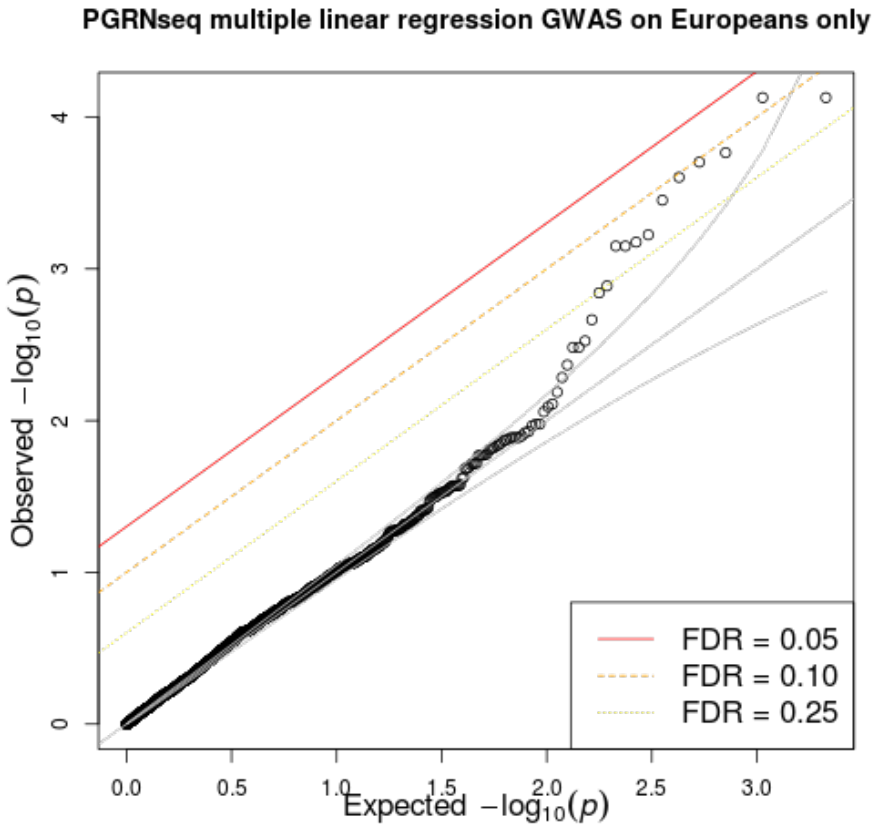
% latex table generated in R 3.1.1 by xtable 1.7-3 package % Mon Sep 15 10:16:21 2014

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')
```



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 10:16:22 2014

	N	effB	se_effB	chi2.1df	P1df
chr2:234667496:C:A	155.000000	-0.496287	0.125266	15.696408	0.000074
chr2:234668828:T:C	155.000000	-0.496287	0.125266	15.696408	0.000074
chr10:96749666:T:A	155.000000	0.449328	0.119579	14.119368	0.000172
chr10:96535761:T:G	155.000000	0.438448	0.117805	13.851828	0.000198
chr2:234602202:A:C	154.000000	-0.480939	0.131296	13.417752	0.000249
chr2:234665942:G:T	155.000000	-0.452930	0.126764	12.766495	0.000353

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)
```

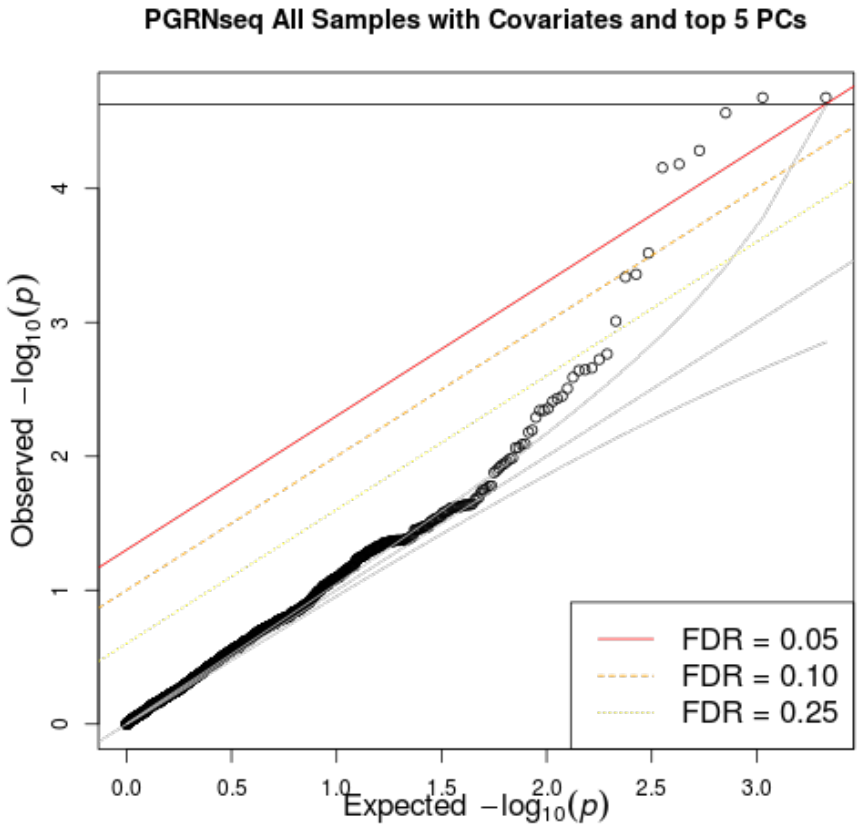
% latex table generated in R 3.1.1 by xtable 1.7-3 package % Mon Sep 15 10:16:22 2014

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')
```



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 10:16:22 2014

	N	effB	se_effB	chi2.1df	P1df
chr2:234667496:C:A	206.000000	-0.464753	0.109230	18.103430	0.000021
chr2:234668828:T:C	206.000000	-0.464753	0.109230	18.103430	0.000021
chr2:234602202:A:C	205.000000	-0.480647	0.114561	17.602576	0.000027
chr2:234602277:G:T	206.000000	-0.440514	0.108860	16.375179	0.000052
chr2:234665942:G:T	206.000000	-0.440676	0.110409	15.930567	0.000066
chr2:234602157:C:T	206.000000	-0.429797	0.108076	15.814937	0.000070

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 10:16:22 2014

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