Using Transcript Levels and Cell Culture Phenotypes to Extend Association Studies

Applications to Clinical Trials

Nancy J. Cox, Ph.D.

The University of Chicago

http://genemed.bsd.uchicago.edu

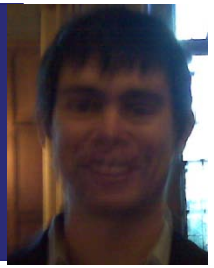
GWAS Works!

Every week
All top journals
Similar reports

The New England Journal of Medicine

GWAS success!

GWAS: Michigan



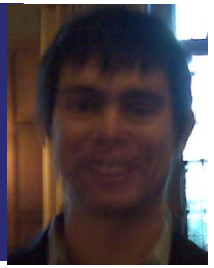
Sequencing Works!

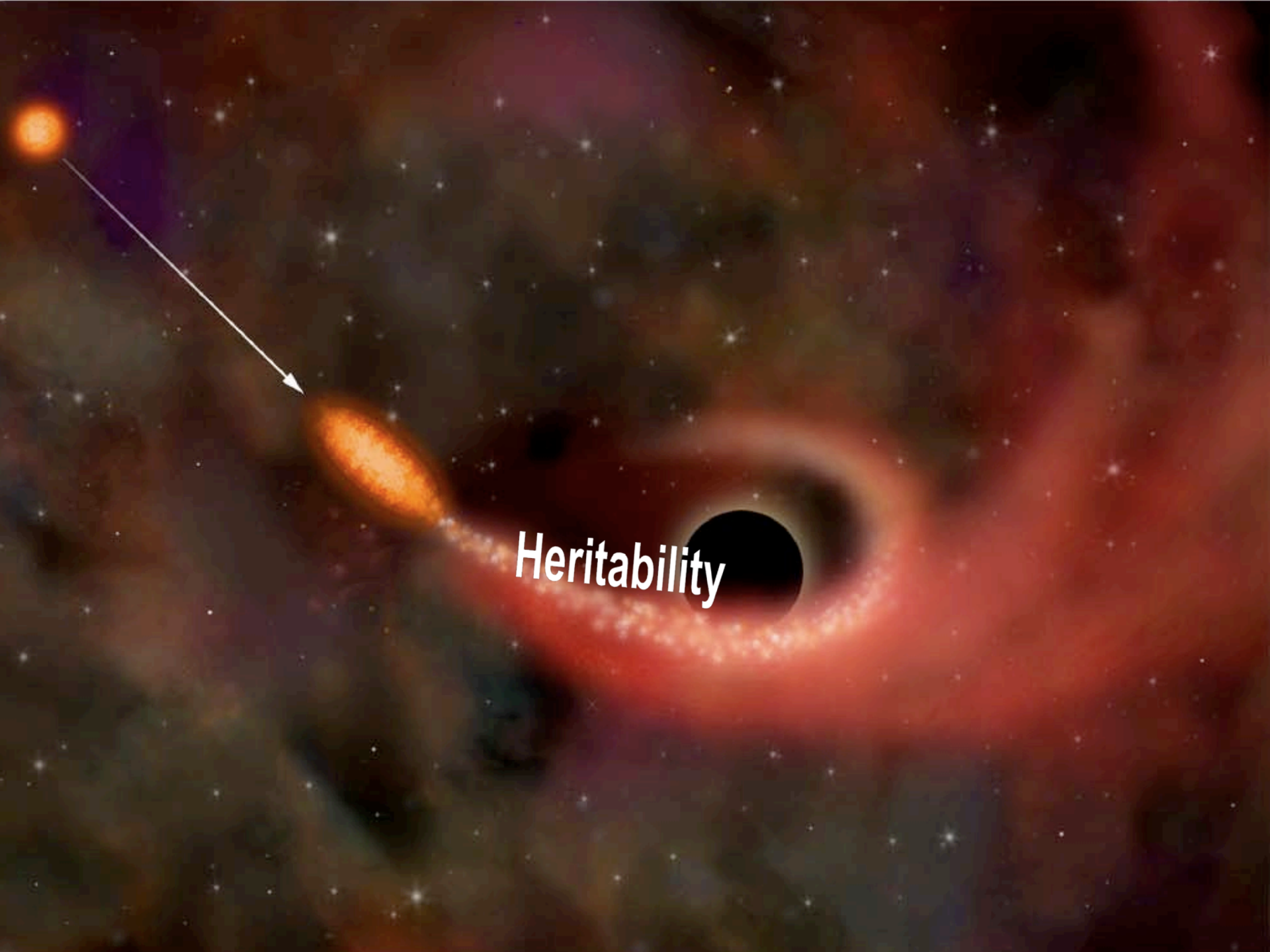
Every week
All top journals
Similar reports

The New England Journal of Medicine

Sequencing success!

Sequenc e Michiga




Heritability

We’ve Been Picking the Cherries

The case of the missing heritability

When scientists opened up the human genome, they expected to find the genetic components of common traits and diseases. But they were nowhere to be seen. **Brendan Maher** shines a light on six places where the missing loot could be stashed away.

If you want to predict how tall your children might one day be, a good bet would be to look in the mirror, and at your mate. Studies going back almost a century have estimated that height is 80–90% heritable. So if 29 centimetres separate the tallest 5% of a population from the shortest, then genetics would account for as many as 27 of them[1].

This year, three groups of researchers[2–4] scoured the genomes of huge populations (the largest study[4] looked at more than 30,000 people) for genetic variants associated with the height differences. More than 40 turned up.

But there was a problem: the variants had tiny effects. Altogether, they accounted for little more than 5% of height's heritability — just 6 centimetres by the calculations above. Even though these genome-wide association studies (GWAS) turned up dozens of variants, they did "very little of the prediction that you would do just by asking people how tall their parents are", says Joel Hirschhorn at the Broad Institute in Cambridge, Massachusetts, who led one of the studies[3].

Height isn't the only trait in which genes have gone missing, nor is it the most important. Studies looking at similarities between identical and fraternal twins estimate heritability at more than 90% for autism[5] and more than 80% for schizophrenia[6]. And genetics makes a major contribution to disorders such as obesity, diabetes and heart disease. GWAS, one of the most celebrated techniques of the past five years, promised to deliver many of the genes involved (see 'Where's the reward?', page 20). And to some extent they have, identifying more than 400 genetic variants that contribute to a variety of traits and common diseases. But even when dozens of genes have been linked to a trait, both the individual and cumulative effects are disappointingly small and nowhere near enough to explain earlier estimates of heritability. "It is the big topic in the genetics of common disease right now," says Francis Collins, former head of the National Human Genome Research Institute (NHGRI) in Bethesda, Maryland. The unexpected results left researchers at a point "where we all had to scratch our heads and say, 'Huh?'", he says.

Although flummoxed by this missing heritability, geneticists remain optimistic that they can find more of it. "These are very early days, and there are things that are doable in the next year or two that may well explain another sizeable chunk of heritability," says Hirschhorn. So where might it be hiding?

ILLUSTRATION BY D. PARKINS

© 2008 Macmillan Publishers Limited. All rights reserved

The case of the missing heritability

When scientists ...ned up the human genome, they expected to find the genet... components of common traits ...d diseases. But they were nowhere to be seen. **Brendan Maher** sh...s a light on six places ...ere the missing loot could be stashed away.

If you want to predict how tall your children might one day be, a good bet would be to look in the mirror, and at your mate. Studies going back almost a century have estimated that height is 80–90% heritable. So if 29 centimetres separate the tallest 5% of a population from the shortest, then genetics would account for as many as 27 of them[1].

This year, three groups of researchers[2–4] scoured the genomes of huge populations (the largest study[4] looked at more than 30,000 people) for genetic variants associated with the height differences. More than 40 turned up.

But there was a problem: the variants had tiny effects. Altogether, they accounted for little more than 5% of height's heritability — just 6 centimetres by the calculations above.

Even though these genome-wide association studies (GWAS) turned up dozens of variants, they did "very little of the prediction that you would do just by asking people how tall their parents are", says Joel Hirschhorn at the Broad Institute in Cambridge, Massachusetts, who led one of the studies[3].

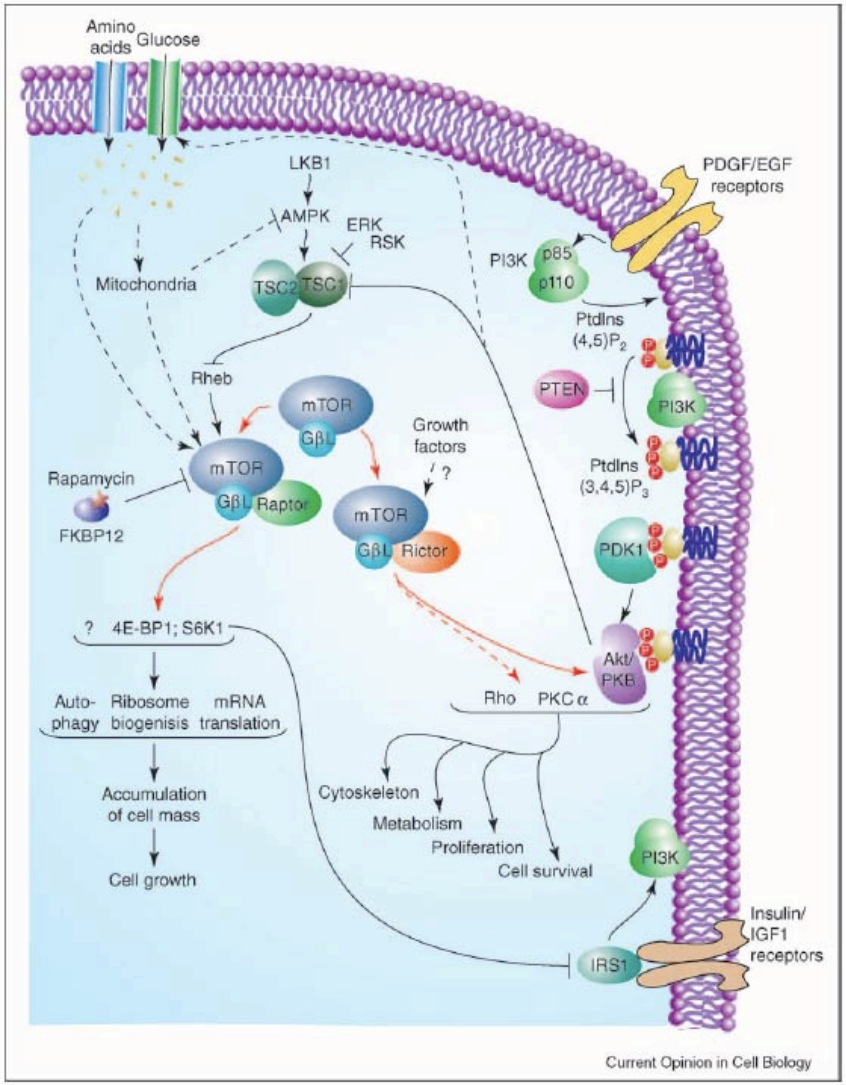
Height isn't the only trait in which genes have gone missing, nor is it the most important. Studies looking at similarities between identical and fraternal twins estimate heritability at more than 90% for autism[5] and more than 80% for schizophrenia[6]. And genetics makes a major contribution to disorders such as obesity, diabetes and heart disease. GWAS, one of the most celebrated techniques of the past five years, promised to deliver many of the genes involved (see 'Where's the reward?', page 20). And to some extent they have, identifying more than 400 genetic variants that contribute to a variety of traits and co...mon diseases. But even when dozens of genes ... been linked to a trait, both the individual and cumulative effects are disappointingly small and nowhere near enough to explain earlier estimates of heritability. "It is the big topic in the genetics of common disease right now," says Francis Collins, former head of the National Human Genome Research Institute (NHGRI) in Bethesda, Maryland. The unexpected results left researchers at a point "where we all had to scratch our heads and say, 'Huh?'", he says.

Although flummoxed by this missing heritability, geneticists remain optimistic that they can find more of it. "These are very early days, and there are things that are doable in the next year or two that may well explain another sizeable chunk of heritability," says Hirschhorn. So where might it be hiding?

ILLUSTRATION BY ...RRING

© 2008 Macmillan Publishers Limited. All rights reserved

The Real Problem?



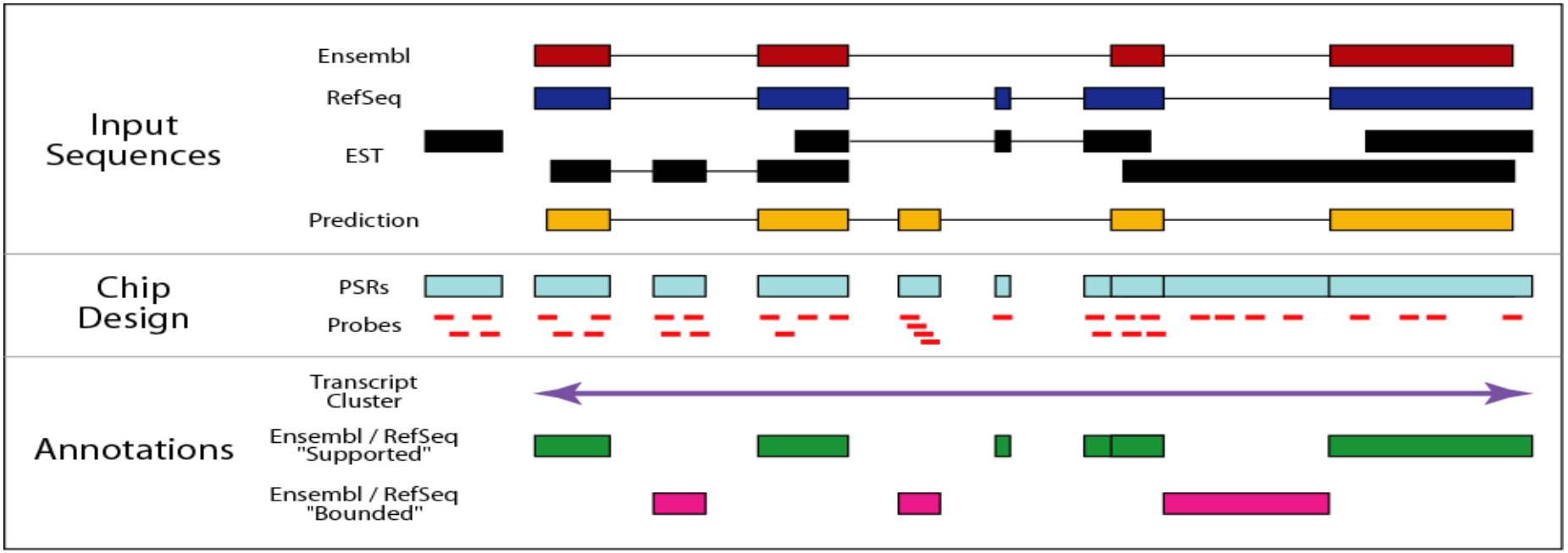
Missing Biology

How Are We Doing with Identifying Primary Biology?

- **Complement system in macular degeneration**
- **Autophagy in Crohn's disease**
- **?**
- **?**

How Can We Improve Our Ability to Identify Genes and Use Genetics to Understand Biology?

Transcriptome Studies



Affymetrix GeneChip® Human Exon 1.0ST Array in HapMap

- No 3' bias with this gene expression array
- Provide up to 4 probes per probeset (sub-exon)
- Core, extended, full probesets for flexible analyses

Results of eQTL Discovery Studies Served in SCAN

- SNP and Copy number ANnotation database
- http://www.scandb.org
- Query by SNP (rs number), gene, or physical region
- eQTL information, multi-locus LD information, physical and functional annotations

Are Trait-Associated SNPs More Likely to Be eQTLs?

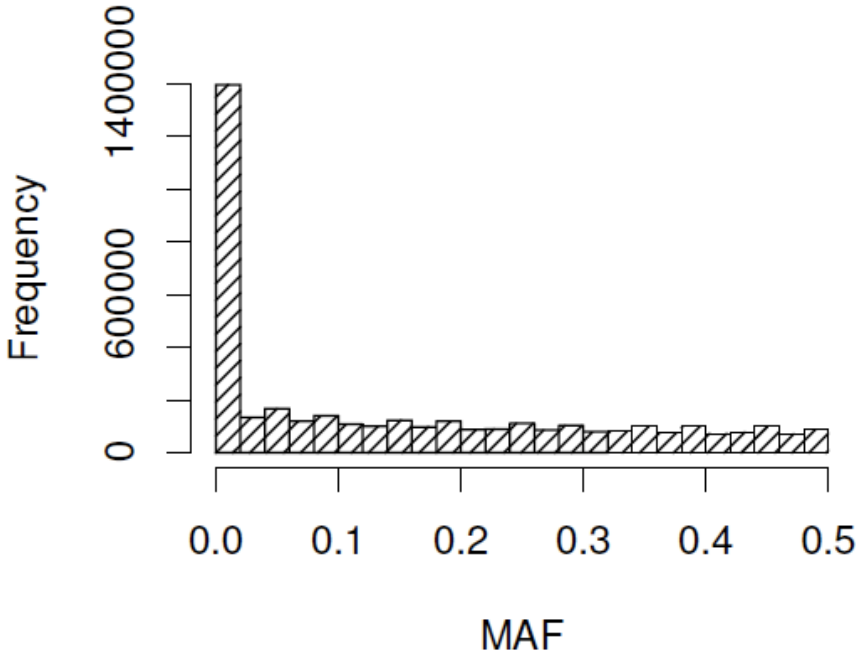
SCAN database

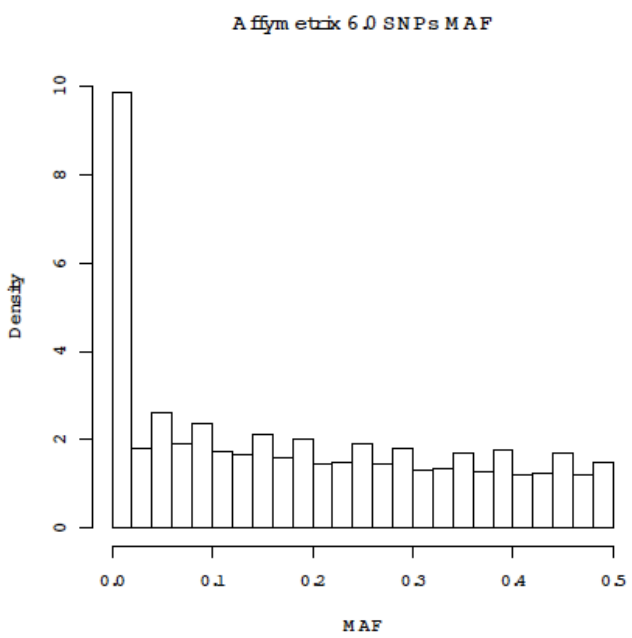
http://www.scandb.org

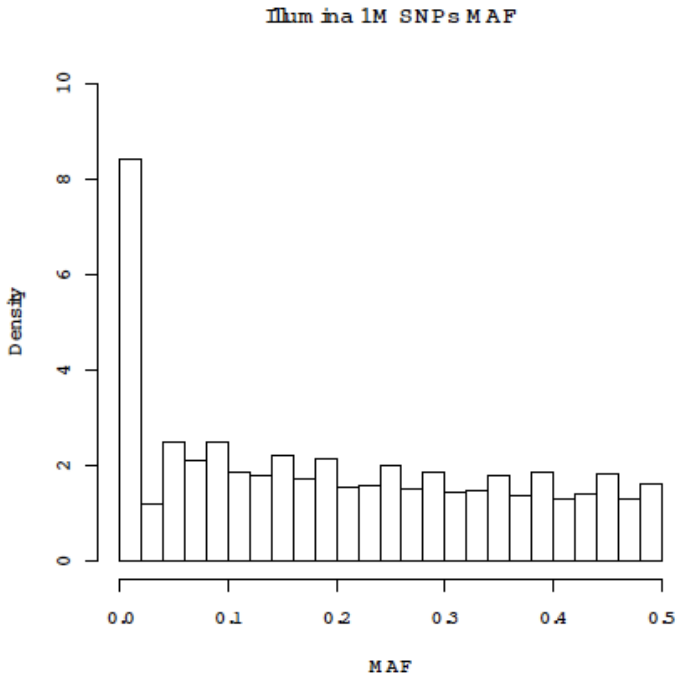
X

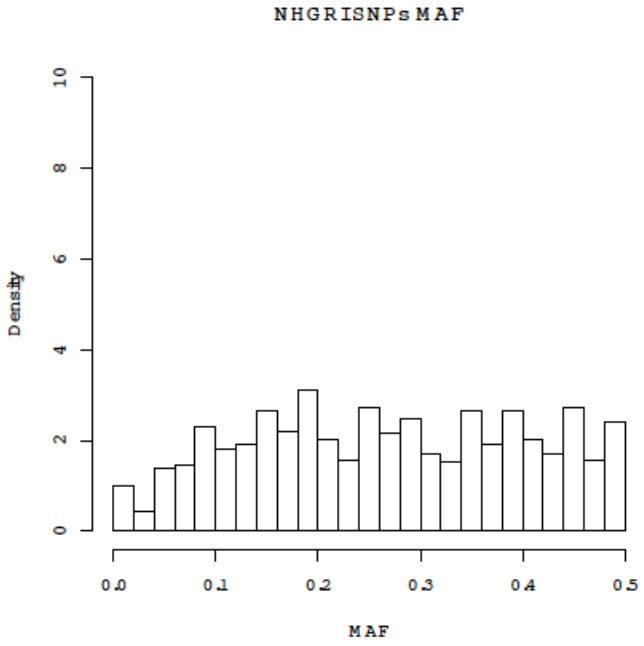
NHGRI GWAS Catalog

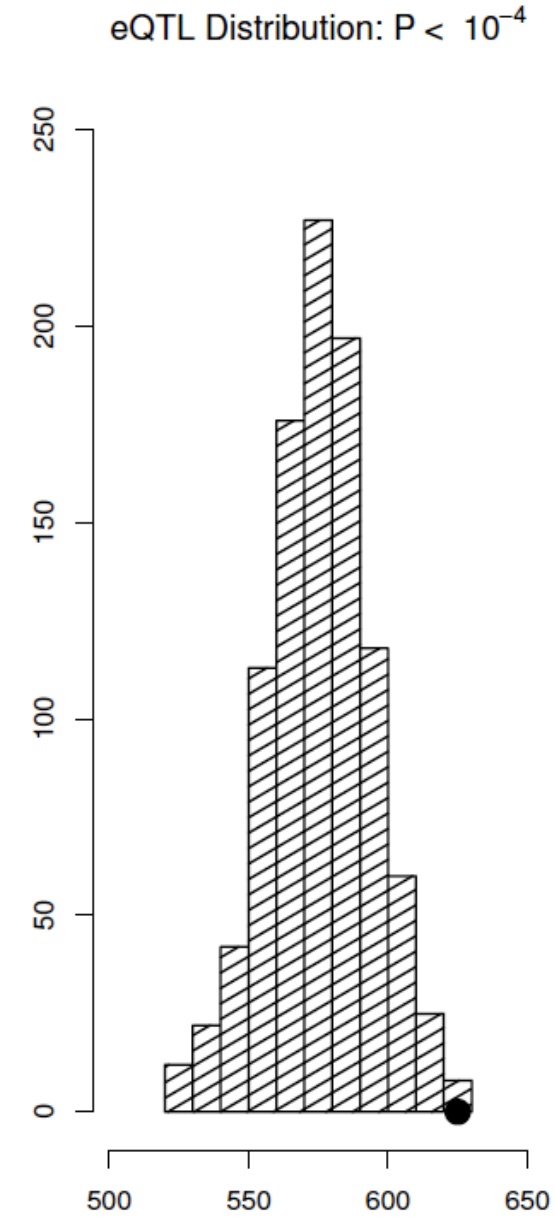
http://www.genome.gov/gwastudies/

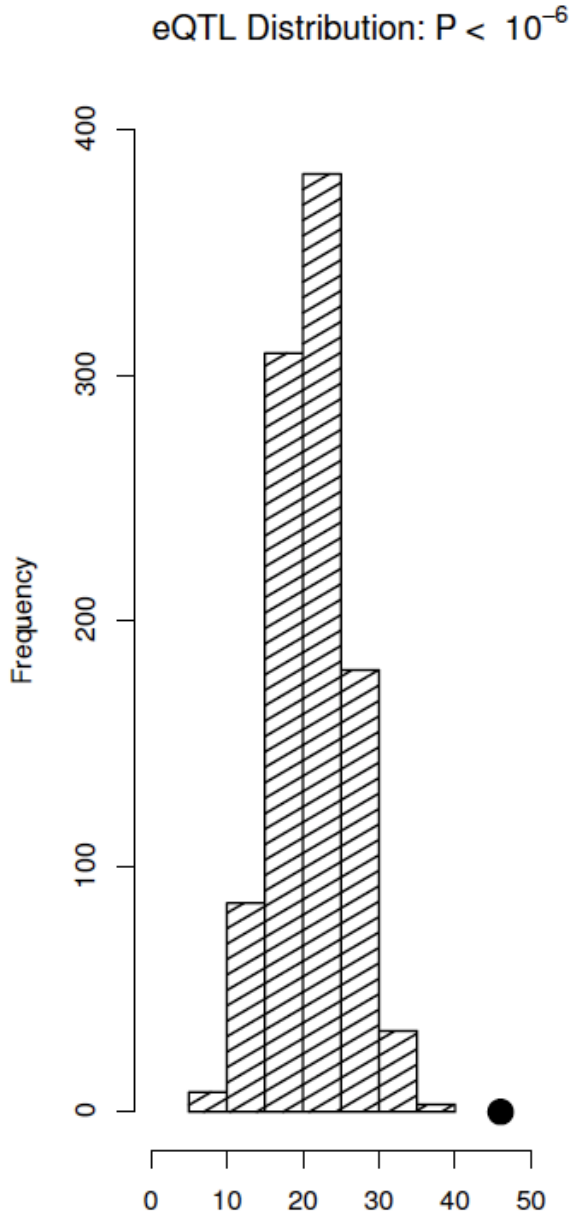

Frequency
0
600000
1400000
0.0
0.1
0.2
0.3
0.4
0.5
MAF

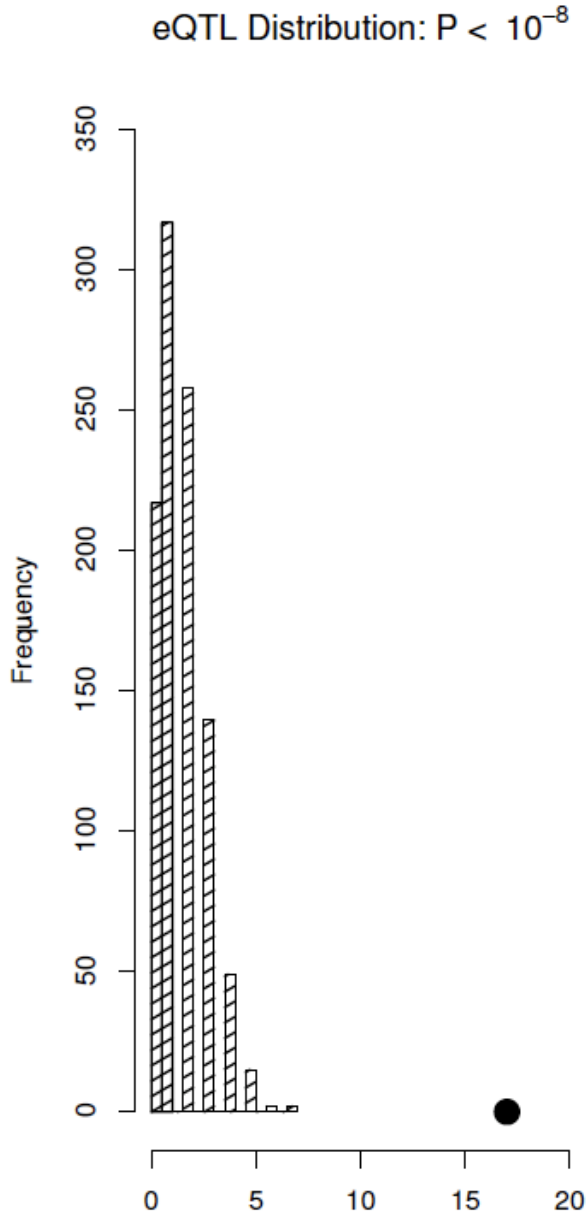

Affymetrix 6.0 SNPs MAF
Density
0
2
4
6
8
10
0.0
0.1
0.2
0.3
0.4
0.5
MAF

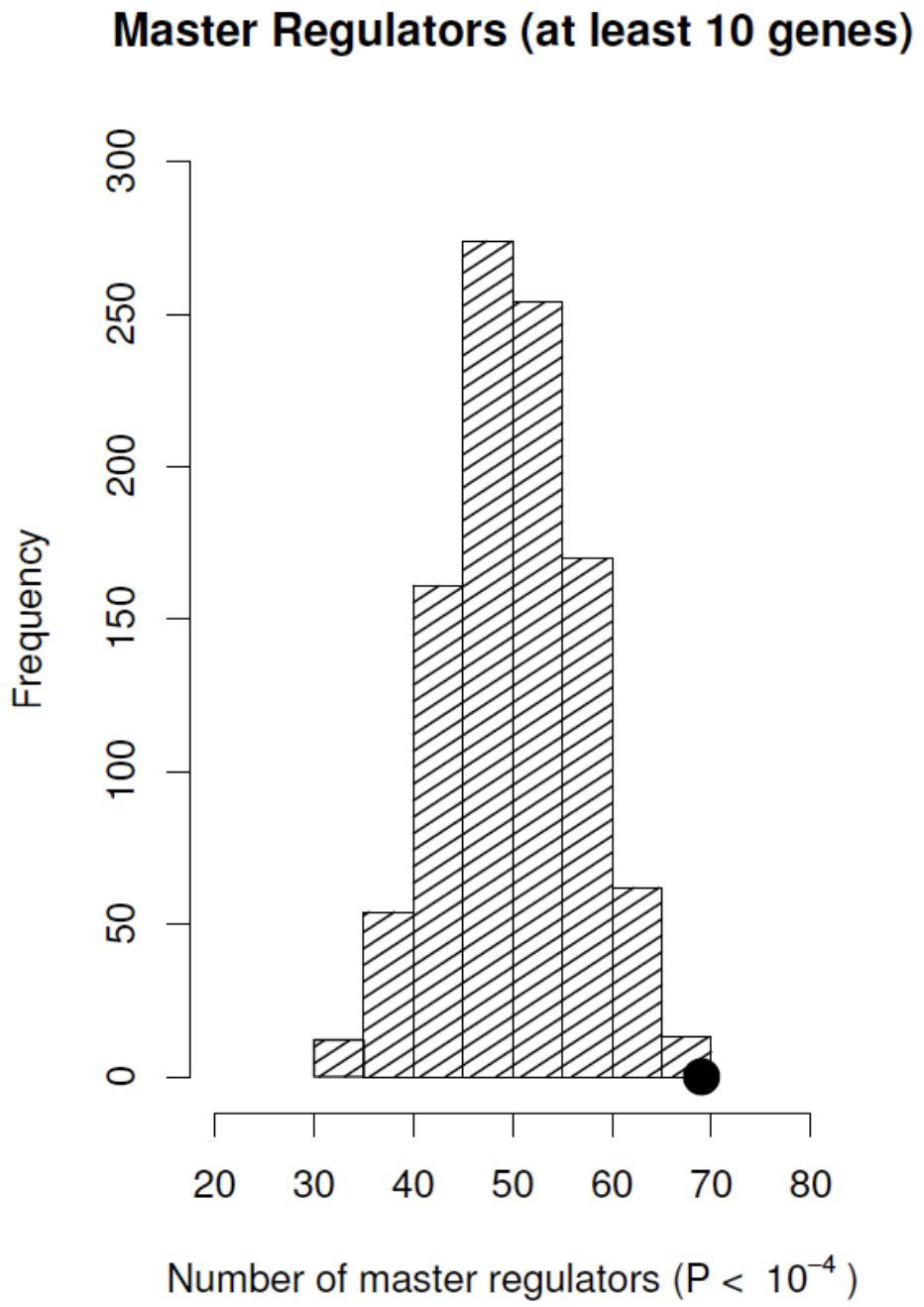

Illumina 1M SNPs MAF
Density
0
2
4
6
8
10
0.0
0.1
0.2
0.3
0.4
0.5
MAF

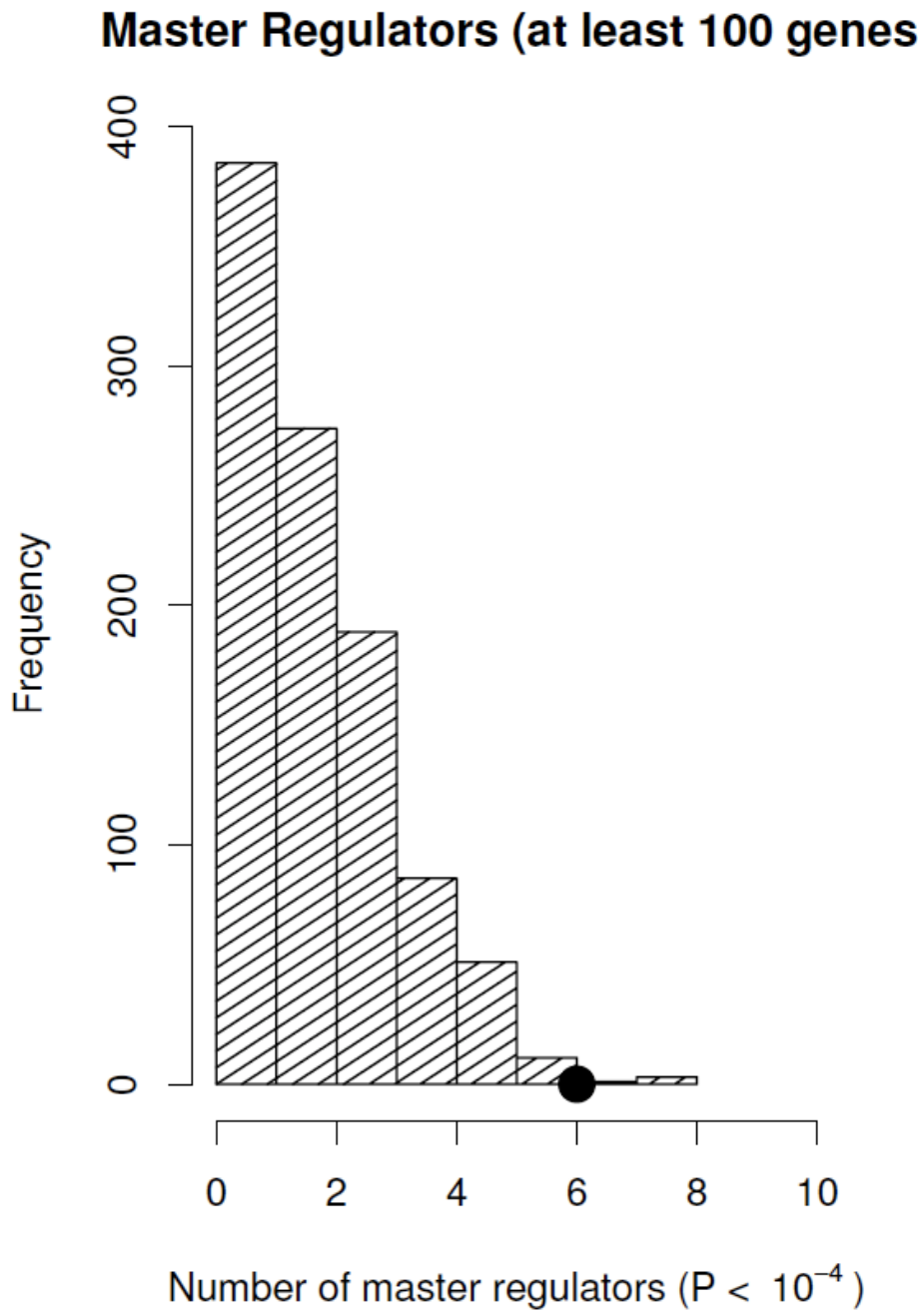

NHGRISNPs MAF
Density
0
2
4
6
8
10
0.0
0.1
0.2
0.3
0.4
0.5
MAF


eQTL Distribution: P < 10^{-4}
0
50
100
150
200
250
500
550
600
650

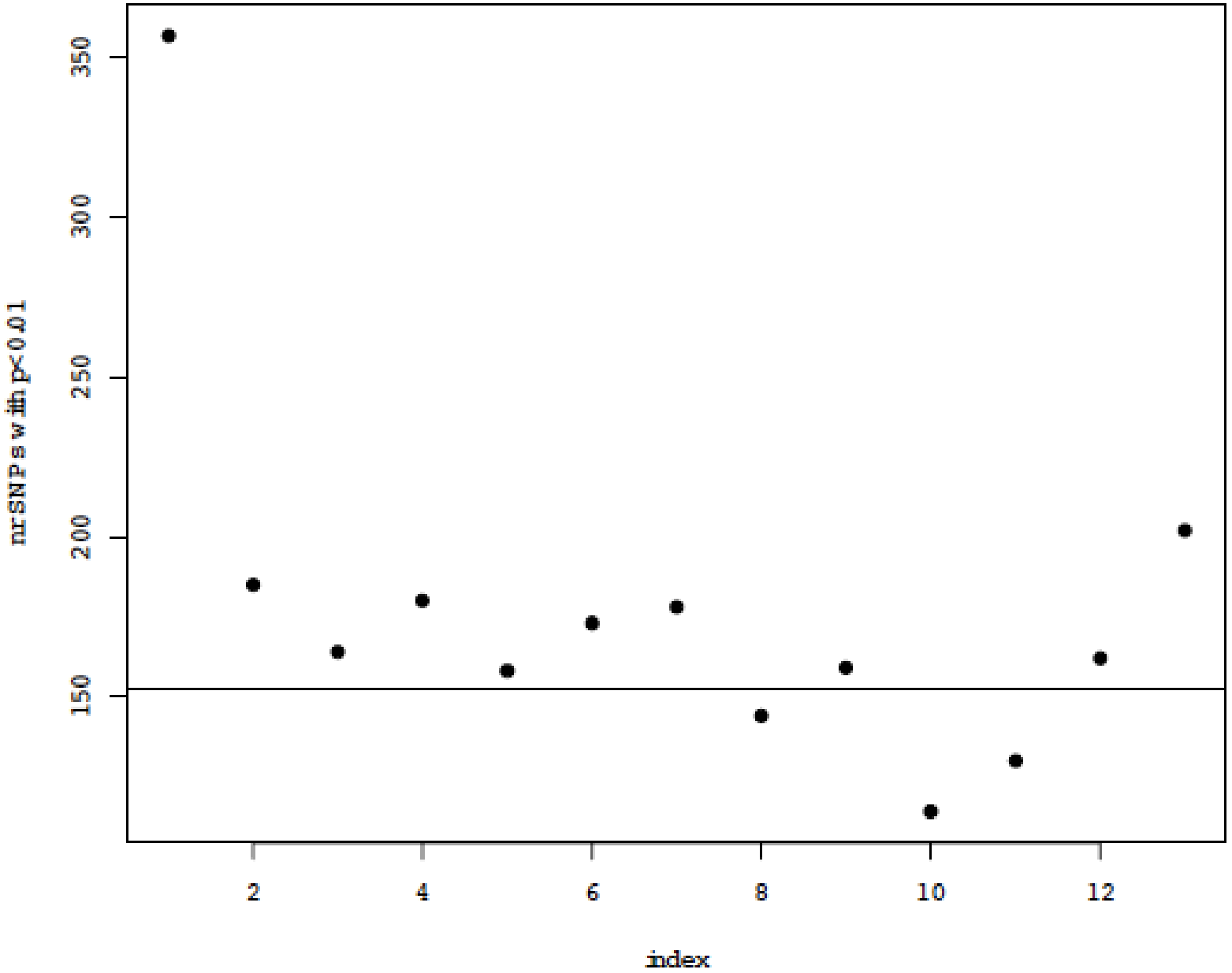

eQTL Distribution: P < 10^{-6}
Frequency
0
100
200
300
400
0
10
20
30
40
50


eQTL Distribution: P < 10^{-8}
Frequency
0
50
100
150
200
250
300
350
0
5
10
15
20


Master Regulators (at least 10 genes)
Frequency
0
50
100
150
200
250
300
20
30
40
50
60
70
80
Number of master regulators (P < 10^-4)

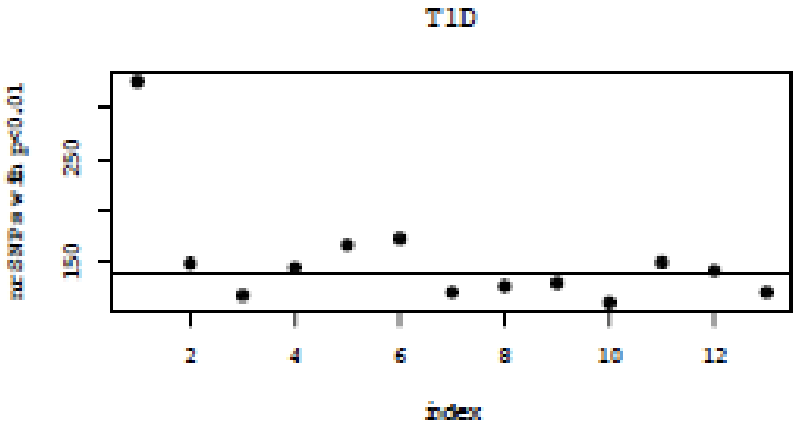

Master Regulators (at least 100 genes
Frequency
0
100
200
300
400
0
2
4
6
8
10
Number of master regulators (P < 10^-4)

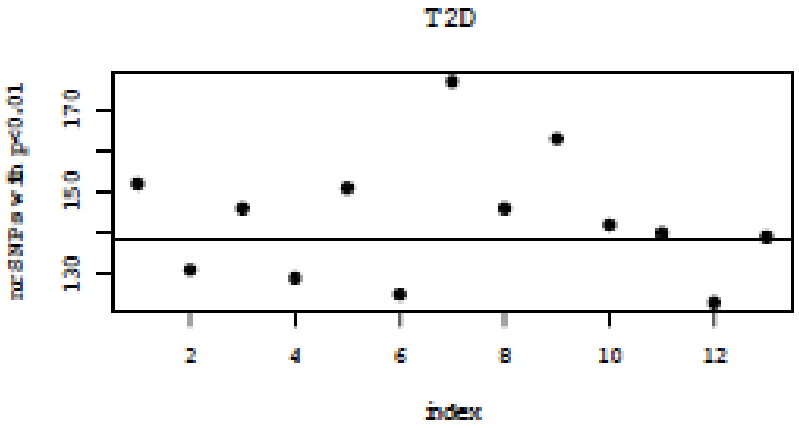
SNPs with the highest eQTL function scores are enriched for WTCCC Crohn's susceptibility loci

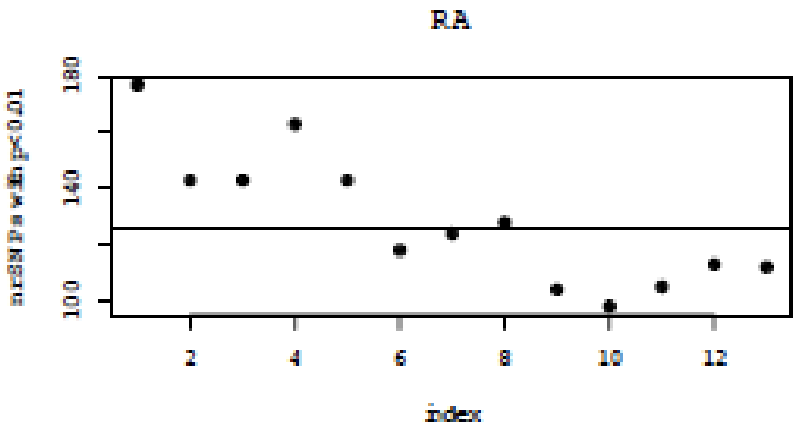


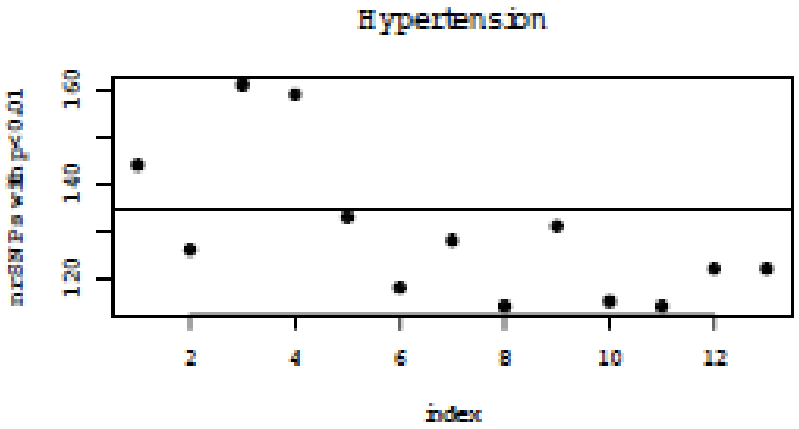
WTCCC Crohn's GWAS Has eQTL Enrichment, But …

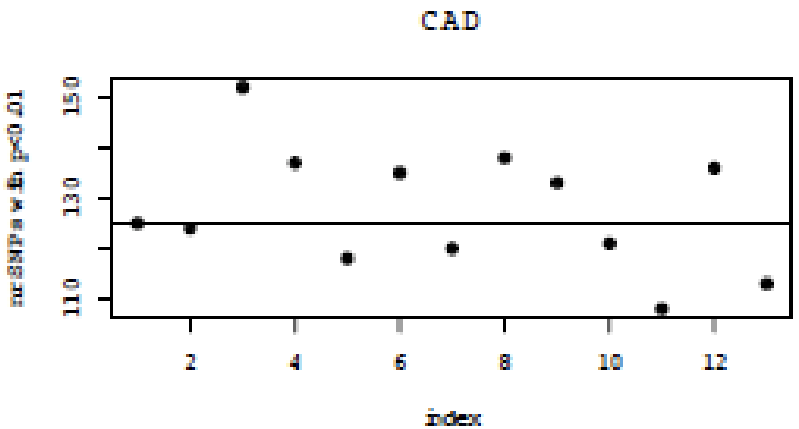
- NO Enrichment of non-synonymous SNPs
- NO enrichment of SNPs in introns
- NO enrichment of SNPs in 3' or 5' UTR

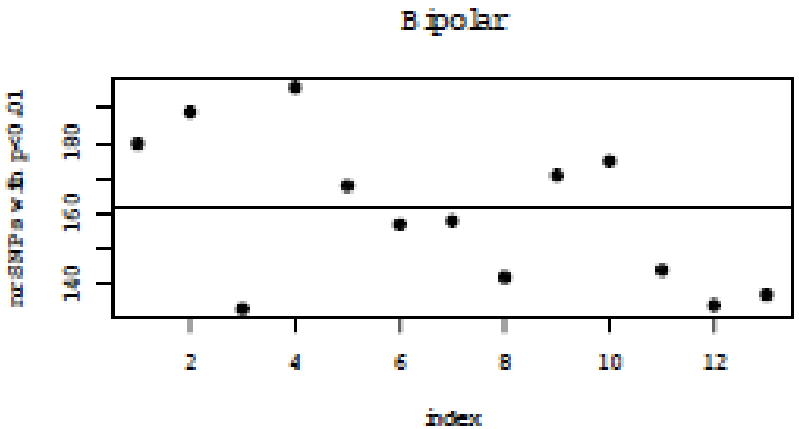

T1D
nrSNPs with p<0.01
index


T2D
nrSNPs with p<0.01
index

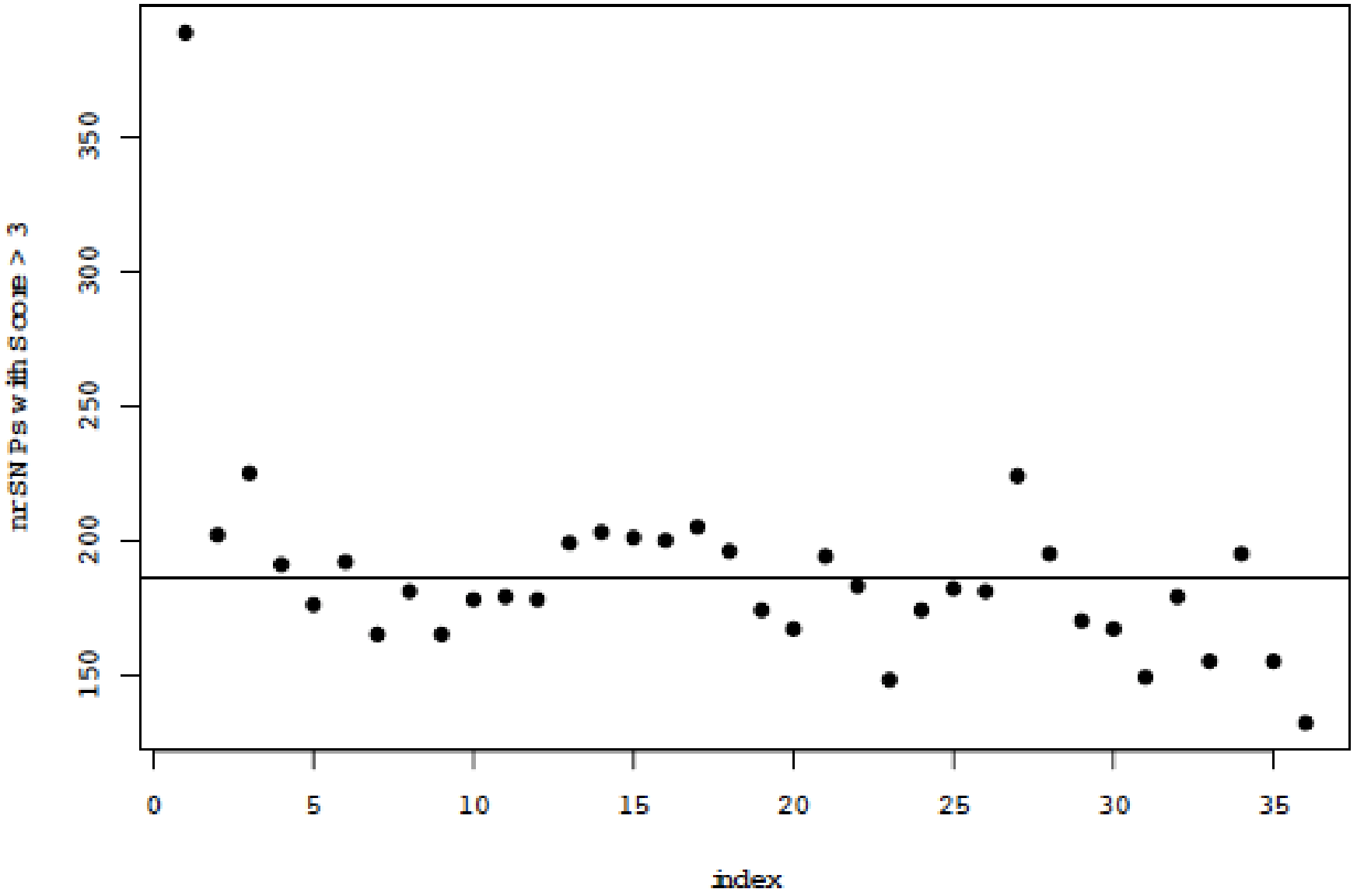

RA
nrSNPs with p<0.01
index

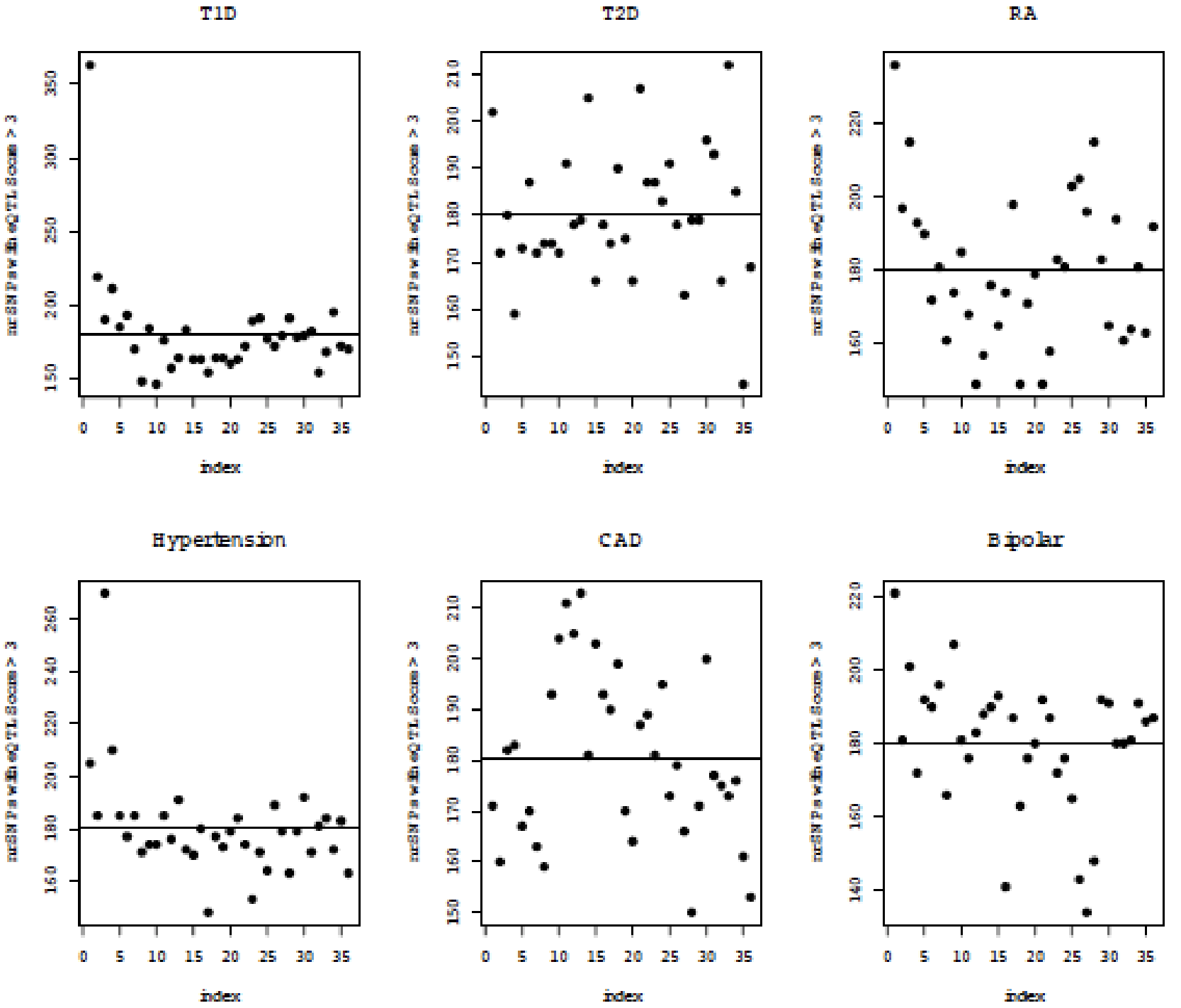

Hypertension
nrSNPs with p<0.01
index


CAD
nrSNPs with p<0.01
index


Bipolar
nrSNPs with p<0.01
index

SNPs with the strongest evidence for association in WTCCC Crohn's disease are more likely to have eQTL function scores > 3




T1D
T2D
RA
Hypertension
CAD
Bipolar
nrSNPswitheQTLScore > 3
index

Results Are Robust Across Wide Range of Phenotypes

Using eQTLs from CEU+YRI	# SNPs	10^{-4}	10^{-6}
Platform SNPs (MAF > .05)	1,213,906	595285 (.490)	29347 (.024)
NHGRI	1598	972 (0.608)	83 (0.052)
Autoimmune Disorders	259	165 (0.637)	21 (0.081)
Cancers	93	56 (0.602)	4 (0.043)
Neurological/Psychiatric Disorders	63	41 (0.651)	2 (0.032)

Why Is This Important?

- **Annotating SNPs with eQTL information (even from LCLs) will improve our ability to identify SNPs truly associated with disease**
- **Utilizing this information should enhance our understanding of the underlying biology of the disease/trait**
- **Magnitude of enrichment suggests that (for at least some disorders) we are no where near complete identification of associated common variants**

How Can We Make Use of the Information?

- Bayesian (and related frequentist approaches) that allow you to put higher priors on functional SNPs
- Suggests new ways of looking for gene-gene interactions
- Provides novel approaches to building pathways for disease
- New types of GWAS

Enhancements to SCAN

- **Results of eQTL studies on brain**
- **Results of eQTL studies on liver**
- **Results of eQTL studies on adipose tissue**
- **Results of eQTL studies on skeletal muscle**
- **New software for function-based GWAS**

Results Databases

- **SCAN (SNP and Copy number Anotation) http://www.scandb.org**
- **PACdb (Pharmacogenetics And Cell lines) http://www.pacdb.org**
- **PGScore (PharmacoGenomics Score) http://www.pgscore.org**

Phenotype / Genotype Association

» **Select Study:**

80303: A study on pancreatic cancer

» **Select phenotype:**

Hypertension grades 2-4

» **Select drug** Help**:**

gemcitabine + bevacizumab

» **Enter SNPs:**

or

Upload SNP list:

Choose File no file selected

» **Enter p-value threshold:**

.00001

Submit Reset

www.pgscore.org

Chrom	SNP	Position	P-value	OR	Minor Allele	MAF	HWE	CEU Minor Allele	CEU MAF	CEU HWE	YRI Minor Allele	YRI MAF	YRI HWE	ASN Minor Allele	ASN MAF	ASN HWE	Phenotype
2	rs2216322	56228414	3.88e-06	0.3058	G	0.217172	0.009597	G	0.279661	0.7539	G	0.190909	0.1803	G	0.136364	1	neutro34
1	rs4847428	93715030	7.71e-06	2.238	A	0.380471	0.5382	A	0.330508	0.2389	G	0.0438596	1	A	0.353933	1	neutro34
8	rs6473987	56525712	1.18e-05	2.453	A	0.210438	0.4884	T	0.2	0.4286	C	0.403509	0.02833	T	0.117978	1	neutro34
18	rs1950068	66642466	1.28e-05	0.3016	A	0.189831	0.0559	A	0.216667	0.4417	A	0.401786	0.7807	A	0.0786517	0.4252	neutro34
6	rs4118099	154656118	1.45e-05	0.4396	A	0.407095	0.2305	A	0.408333	1	A	0.0438596	1	A	0.252809	0.4148	neutro34
22	rs9614417	42311949	1.93e-05	2.302	C	0.254209	0.6457	C	0.241667	0.1527	C	0.377193	0.3955	C	0.0393258	1	neutro34
19	rs2691239	56243624	2.09e-05	2.147	A	0.398305	0.08878	A	0.333333	0.5595	G	0.27193	1	G	0.140449	0.06243	neutro34
4	rs12509540	131516658	2.09e-05	2.21	A	0.308081	0.07601	A	0.237288	0.7175	A	0.342105	0.5544	A	0.448864	0.8316	neutro34
4	rs1509434	131537020	2.09e-05	2.21	G	0.308081	0.07601	G	0.258333	1	G	0.236842	0.7119	G	0.448864	0.8316	neutro34
8	rs2016573	122677260	2.16e-05	2.441	G	0.217172	1	C	0.25	0.317	C	0.236842	0.4864	T	0.157303	1	neutro34
16	rs3748396	85918250	2.38e-05	2.243	A	0.271331	0.8831	T	0.233333	0.1554	T	0.245614	0.2849	T	0.359551	0.1104	neutro34
8	rs13256272	13652680	2.48e-05	5.277	C	0.0454545	1	G	0.059322	0.1718	G	0.210526	1	G	0.191011	0.7292	neutro34
8	rs10088270	56550383	3.13e-05	2.195	G	0.287879	0.2057	G	0.313559	1	A	0.377193	0.3955	G	0.117978	1	neutro34
19	rs7245440	5345056	3.17e-05	2.267	G	0.250842	0.3531	G	0.25	0.162	A	0.25	0.291	G	0.494382	0.2937	neutro34
6	rs857601	44913192	3.18e-05	2.177	C	0.293919	0.04902	G	0.245614	1	G	0.307018	0.2152	G	0.0393258	1	neutro34
5	rs3776751	171514329	3.47e-05	0.2561	A	0.149832	0.02086	T	0.183333	1	T	0.245614	0.2849	T	0.370787	0.1777	neutro34
18	rs10853217	12065501	3.59e-05	2.126	A	0.350168	0.007416	T	0.316667	0.1323	T	0.366071	0.7771	T	0.455056	0.1997	neutro34
2	rs2111469	56229812	3.81e-05	0.4126	A	0.29223	0.4819	T	0.362069	0.7824	T	0.245614	0.4784	T	0.157303	1	neutro34
12	rs4763331	13292110	3.92e-05	2.703	G	0.124579	0.7904	C	0.116667	1	C	0.145455	0.3018	C	0.0337079	1	neutro34
3	rs1461251	176427638	4.29e-05	0.2136	A	0.132143	4.29e-06	NA	0	0	NA	0	0	T	0.147059	0.01404	neutro34
6	rs1222955	102272022	4.46e-05	0.2314	G	0.133446	0.3228	G	0.125	0.5795	G	0.184211	0.6683	G	0.376404	0.6508	neutro34
18	rs1282559	66632074	4.95e-05	0.3482	G	0.198653	0.1424	G	0.233333	0.009965	A	0.342105	0.5544	G	0.224719	0.7623	neutro34
1	rs237425	93784875	5.08e-05	2.318	G	0.198653	1	G	0.125	0.5795	G	0.22807	0.4539	G	0	1	neutro34
2	rs11693165	59834499	5.19e-05	0.3342	A	0.181818	0.6955	A	0.161017	1	A	0.333333	0.7659	A	0.219101	1	neutro34
11	rs10834677	3611252	5.2e-05	2.074	C	0.375421	0.711	C	0.375	1	A	0.00877193	1	C	0.369318	1	neutro34
7	rs1055010	29517649	5.25e-05	2.211	G	0.260714	0.6418	G	0.333333	0.5595	G	0.263158	0.7351	G	0.365169	0.4951	neutro34
4	rs4333176	10084378	5.38e-05	0.432	C	0.326014	0.5093	C	0.358333	0.1692	A	0.22807	1	A	0.426966	0.2776	neutro34
6	rs12198308	45019864	5.47e-05	2.129	A	0.29223	0.03458	T	0.201923	1	T	0.0181818	1	T	0.0287356	1	neutro34

Example Application

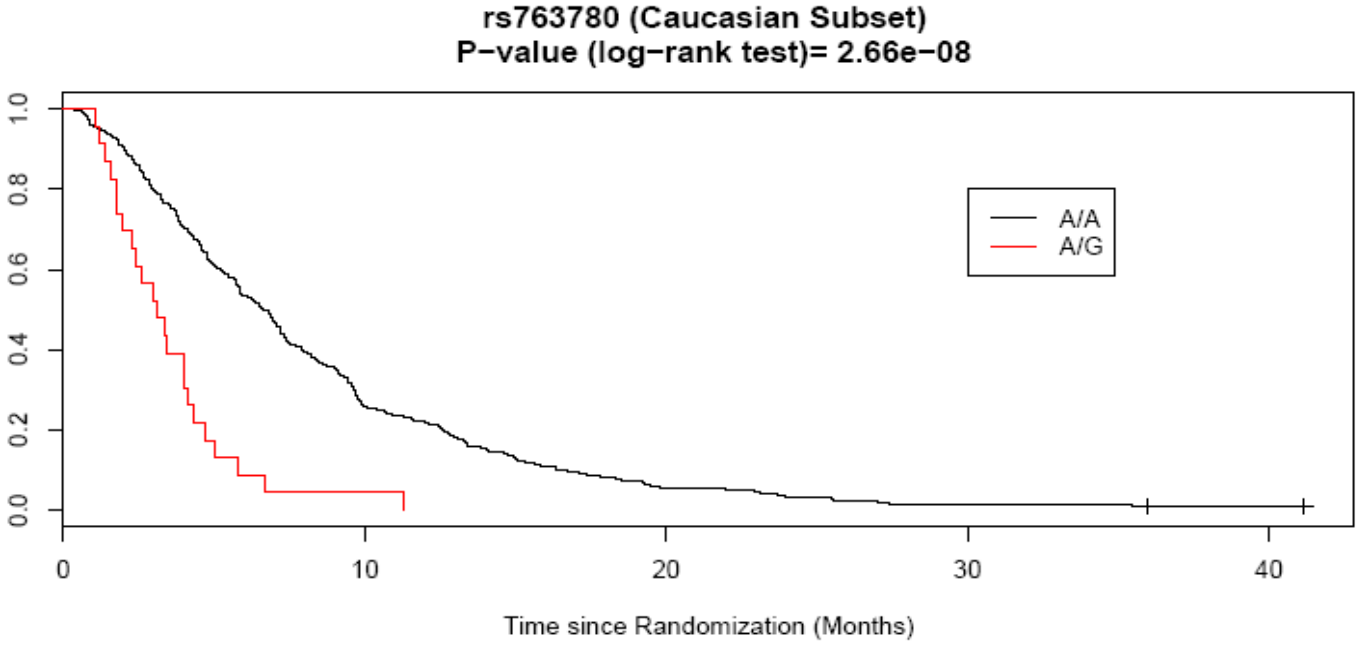
Top hits, neuropenia (3-4), gemcitabine, from PGScore

Chrom	SNP	Position	P-value	OR	Minor Allele	MAF	HWE	CEU Minor Allele	CEU MAF	CEU HWE	YRI Minor Allele	YRI MAF	YRI HWE	ASN Minor Allele	ASN MAF	ASN HWE	Phenotype
2	rs2216322	56228414	3.88e-06	0.3058	G	0.217172	0.009597	G	0.279661	0.7539	G	0.190909	0.1803	G	0.136364	1	neutro34
1	rs4847428	93715030	7.71e-06	2.238	A	0.380471	0.5382	A	0.330508	0.2389	G	0.0438596	1	A	0.353933	1	neutro34
8	rs6473987	56525712	1.18e-05	2.453	A	0.210438	0.4884	T	0.2	0.4286	C	0.403509	0.02833	T	0.117978	1	neutro34
[illegible]	[illegible]	[illegible]	1.28e-	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

Identify overlap with cytotoxicity phenotypes in PACdb

Population	Drug	Phenotype	SNP	Allele	Chrom	Position	Annotation (Host Gene, SNP Function)	P-value
CEU	ARAC	IC50	rs10088270	A/G	8	56550383	XKR4, intron[NM_052898.1]	0.0074
YRI	ARAC	AUC	rs10088270	A/G	8	56550383	XKR4, intron[NM_052898.1]	0.0093

IL17F, ch 1, 10^{-8}, q 0.04, HR 3.27 coding nonsynonymous

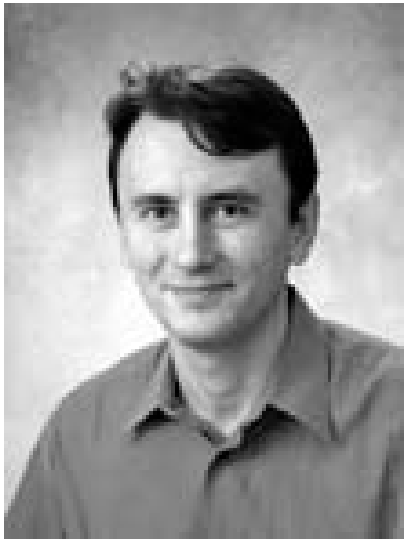



```
Call: survfit(formula = SURV[CAUC] ~ snpi[CAUC])

                      n events median 0.95LCL 0.95UCL
snpi[CAUC]=C/A   22      22   3.24    2.43    4.34
snpi[CAUC]=C/C 272     269   6.64    5.82    7.23
```

The Team

Nancy Cox



Dan Nicolae

M. Eileen Dolan

Eric Gamazon

Wei Zhang and Shiwei Duan

Supported by ENDGAMe and PGRN (PAAR) resources