Beyond GWAS

Erik Fransen

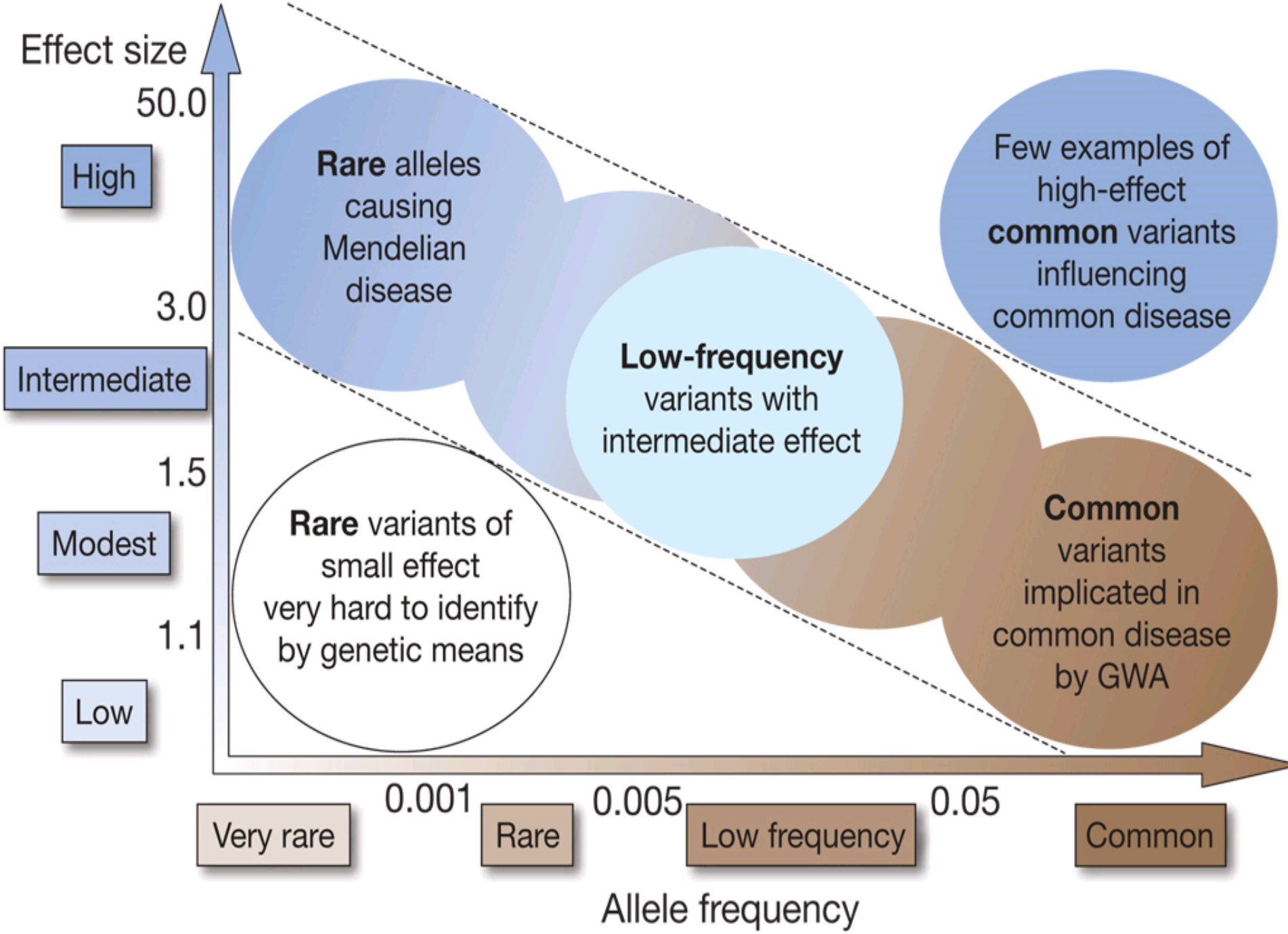
Universiteit Antwerpen

Missing heritability

- Rare variants
- Common variants of smaller effect
- Structural variants
- Gene-gene interactions
- Inflated heritability

Missing heritability

- Rare variants
- Common variants of smaller effect
- Structural variants
- Gene-gene interactions
- Inflated heritability


Effect size
50.0
High
3.0
Intermediate
1.5
Modest
1.1
Low
Rare alleles causing Mendelian disease
Few examples of high-effect common variants influencing common disease
Low-frequency variants with intermediate effect
Rare variants of small effect very hard to identify by genetic means
Common variants implicated in common disease by GWA
0.001
0.005
0.05
Very rare
Rare
Low frequency
Common
Allele frequency

Common and rare variants

- Traditional GWAS identify common variants
  - MAF > 0.05
  - Low-frequency (LF) variants :
    - poorly covered on chip
    - Low LD with SNPs on chip
- Next-generation sequencing (NGS) techniques to identify LF variants in large cohorts

Testing LF variants

- Do LF variants contribute to complex traits?
  - LF variants with large effect?
  - Multiple rare variants in 1 gene?
  - Missing heritability?
- Testing association
  - Classic testing
  - Gene-based testing

Classic association test

- Test association between single variant and disease
  - Chisquare test

	AA	Aa	aa	Total
Case				1000
Control				1000

Classic association test

- Test association between single variant and disease
  - Chisquare test : common variant

	AA	Aa	aa	Total
Case	560	380	60	1000
Control	640	320	40	1000

Classic association test

- Test association between single variant and disease
  - Chisquare test : rare variant

	AA	Aa	aa	Total
Case	987	12	1	1000
Control	991	9	0	1000

  - Test not valid, or low power, unless high N
  - GWAS : MAF<0.05 = problem

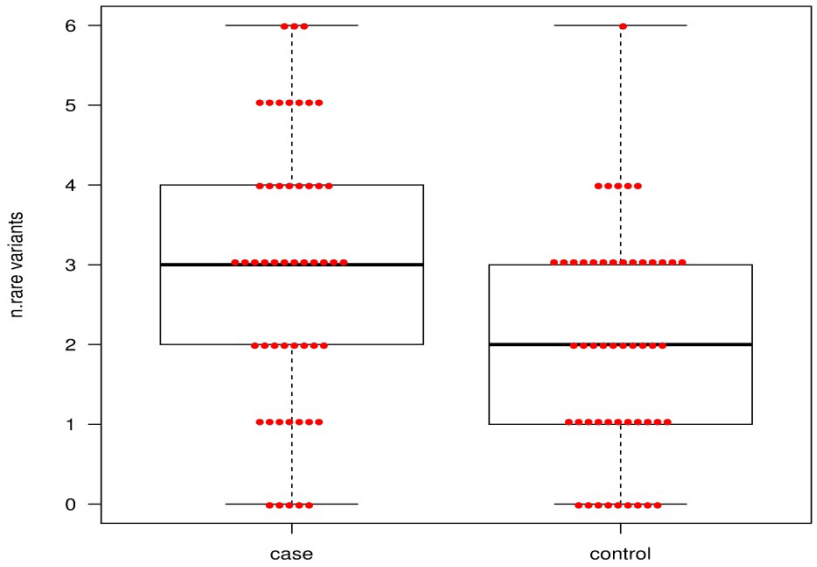
Collapsing rare variants

- Collapsed genotype :
  - Gene-by-gene recoding genotype
    - Any rare variant present or not (yes/no)
  - Associate collapsed GT with phenotype

	No rare variant	At least 1 rare variant	Total
Case	850	150	1000
Control	889	111	1000

Burden test

- Test the burden of rare variants:
  - Count #rare variants within each gene
  - Associate #variants with phenotype

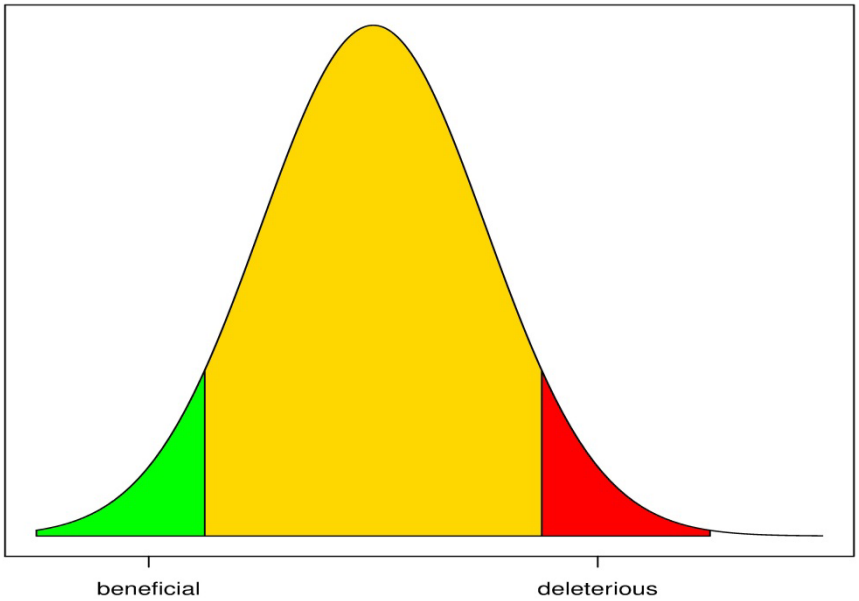


Problem with burden tests

- Collapsing & burden test assume all rare alleles act in one direction
  - Assume deleterious effect
  - Ignore neutral/beneficial alleles
  - Controls may be enriched in beneficial variants

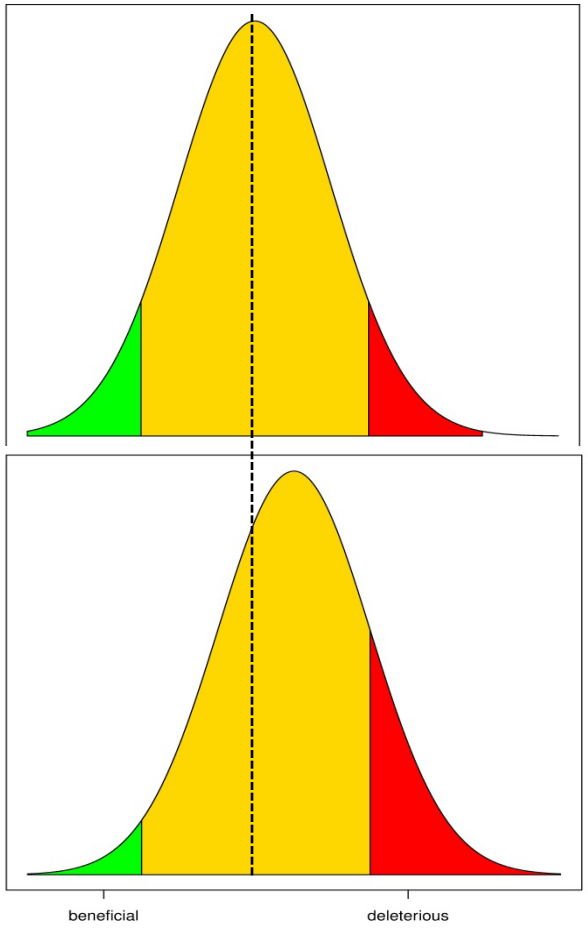
Non-burden test

- Effects of rare alleles represent a distribution
  - Beneficial and deleterious alleles



Non-burden test

- Sequence Kernel Association Test (SKAT)
- Estimate allelic effect sizes for all SNPs
  - Compare distribution of effects between cases and controls
  - Collective effect of all SNPs per gene



Recent results on rare variants

- Require combination of :
  - GWAS data
  - Whole exome/targeted resequencing/whole genome sequencing
- Often
  - Exome sequencing form GWAS cohorts
  - Targeted resequencing of previous GWAS hits
- N : 5,000 – 10,000

Study1 : Amino acid levels

- 9 phenotypes = levels of 9 AA
  - Risk factors for (ao.) T2D, Alzheimers)
  - Exome sequencing on 8,800 ID + GWAS data
- Single-marker test:
  - 17 associations G-W significant at 12 loci
  - 3 novel loci

Study1 : Amino acid levels

- 9 phenotypes = levels of 9 AA
  - Risk factors for (ao.) T2D, Alzheimers)
  - Exome sequencing on 8,800 ID + GWAS data
- Single-marker test
- Gene-based (SKAT) test
  - 1 additional gene
  - p=9^{E}-8 on aggregated effect of all variants
  - most sig. single SNP:10^{E}-5

Study1 : Amino acid levels

- 9 phenotypes = levels of 9 AA
  - Risk factors for (ao.) T2D, Alzheimers)
  - Exome sequencing on 8,800 ID + GWAS data
- Single-marker test
- Gene-based (SKAT) test
- Missing heritability?
  - Common variants (GWAS) : 6%
  - Plus rare variants (exome) : 15-20%

Study 2 : Type2 diabetes

- Meta-analysis of 23k cases and 40k controls
- Combine:
  - (old) GWAS data
  - Whole exome resequencing (50x depth)
  - Whole genome resequencing (4x depth)

Study 2 : Type2 diabetes

- Meta-analysis of 23k cases and 40k controls
- Combine:
- Results:
  - 175K LF variants identified
  - No LF variants were genome-wide significant
  - Only one highly sig. LF variants in previously unidentified gene
  - SKAT (of LF variants): no genome-wide sig. genes

  → no LF variants with high effect in T2D

Study 2 : Type2 diabetes

- Meta-analysis of 23k cases and 40k controls
- Combine:
- Results:

  →no LF variants with high effect in T2D

  - LF variants in previously known loci
    - LF variants underly old GWAS signal
    - Old GWAS hit often due to >1 LF variant

  → better pinpointing causal variant

Conclusion LF variants

- Most, but not all, LF analyses point to loci/genes previously identified by GWAS
- LF variants with large effect size : not found (yet?)
- LF variants may explain part of missing heritability, but not all

Missing heritability

- Rare variants
- Common variants of smaller effect
- Structural variants
- Gene-gene interactions
- Inflated heritability

Heritability in common variants

- GWAS results on human adult height
  - N = 34,000 → 20 loci associated
  - N = 180,000 → 180 loci associated
- All associated SNPs account for 10% of phenotypic variation

Heritability in common variants

- GWAS results on human adult height
  - N = 34,000 → 20 loci associated
  - N = 180,000 → 180 loci associated
  - N= 400,000 → 460 loci
- All associated SNPs account for 20% of phenotypic variation
  - SNPs with $p<5.0^{E}\text{-}8$

Common SNPs for height

Published in final edited form as:
Nat Genet. 2010 July ; 42(7): 565–569. doi:10.1038/ng.608.

Common SNPs explain a large proportion of heritability for human height

Jian Yang[1], **Beben Benyamin**[1], **Brian P McEvoy**[1], **Scott Gordon**[1], **Anjali K Henders**[1], **Dale R Nyholt**[1], **Pamela A Madden**[2], **Andrew C Heath**[2], **Nicholas G Martin**[1], **Grant W Montgomery**[1], **Michael E Goddard**[3], and **Peter M Visscher**[1,*]

[1]Queensland Institute of Medical Research, 300 Herston Road, Brisbane, Queensland 4006, Australia

[2]Department of Psychiatry, Washington University St. Louis, MO, USA

[3]Department of Food and Agricultural Systems, University of Melbourne, Parkville 3011, Australia

Common SNPs for height

- Missing heritability in common SNPs ?
  - SNPs not reaching genome-wide significance
  - Causal variants not in complete LD with typed SNPs

Common SNPs for height

- Missing heritability in common SNPs ?
  - SNPs not reaching genome-wide significance
  - Causal variants not in complete LD with typed SNPs
- Effect of all SNPs together:
  - No individual SNPs pinpointed
  - Aggregated effect of all SNPs
  - 45% < H^2

Common SNPs for height

- Missing heritability in common SNPs ?
  - SNPs not reaching genome-wide significance
  - Causal variants not in complete LD with typed SNPs
- Correct for incomplete LD :
  - Correction depends on MAF of causal SNPs
  - MAF ~ typed SNPs : 54%
  - MAF lower : 80%

Conclusion

- Typed SNPs can explain over 50% of H^2
  - Remainder : incomplete LD and (compatible with) lower MAF
- Larger GWAS can identify more SNPs
  - P lowers as N increases
  - Deep resequencing will identify more causal variants

Psychiatric disorders

nature

Vol 460|6 August 2009|doi:10.1038/nature08185

LETTERS

Common polygenic variation contributes to risk of schizophrenia and bipolar disorder

The International Schizophrenia Consortium*

Schizophrenia is a severe mental disorder with a lifetime risk of about 1%, characterized by hallucinations, delusions and cognitive deficits, with heritability estimated at up to 80%[1,2]. We performed a genome-wide association study of 3,322 European individuals with schizophrenia and 3,587 controls. Here we show, using two analytic approaches, the extent to which common genetic variation underlies the risk of schizophrenia. First, we implicate the major histocompatibility complex. Second, we provide molecular genetic evidence for a substantial polygenic component to the risk of schizophrenia involving thousands of common alleles of very small effect. We show that this component also contributes to the risk of bipolar disorder, but not to several non-psychiatric diseases.

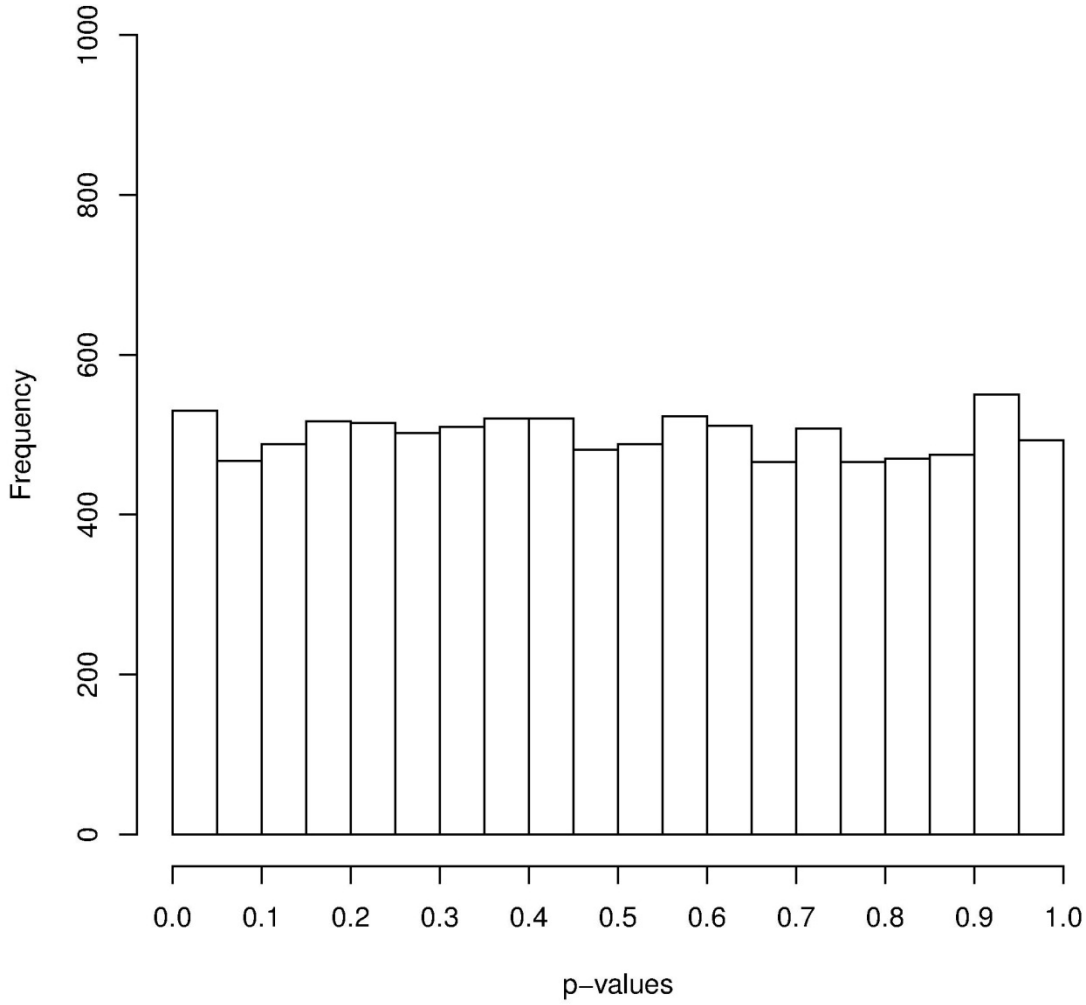
Table 2, Supplementary Fig. 2 and section 5 and 6 in Supplementary Information).

The best imputed SNP, which reached genome-wide significance (rs3130297, $P = 4.79 \times 10^{-8}$, T allele odds ratio = 0.747, minor allele frequency (MAF) = 0.114, 32.3 megabases (Mb)), was also in the MHC, 7 kilobases (kb) from *NOTCH4*, a gene with previously reported associations with schizophrenia[4]. We imputed classical human leukocyte antigen (HLA) alleles; six were significant at $P < 10^{-3}$, found on the ancestral European haplotype[5] (Table 1, Supplementary Table 3 and section 3 in Supplementary Information). However, it was not possible to ascribe the association to a specific HLA allele, haplotype or region (Supplementary Table 3 and

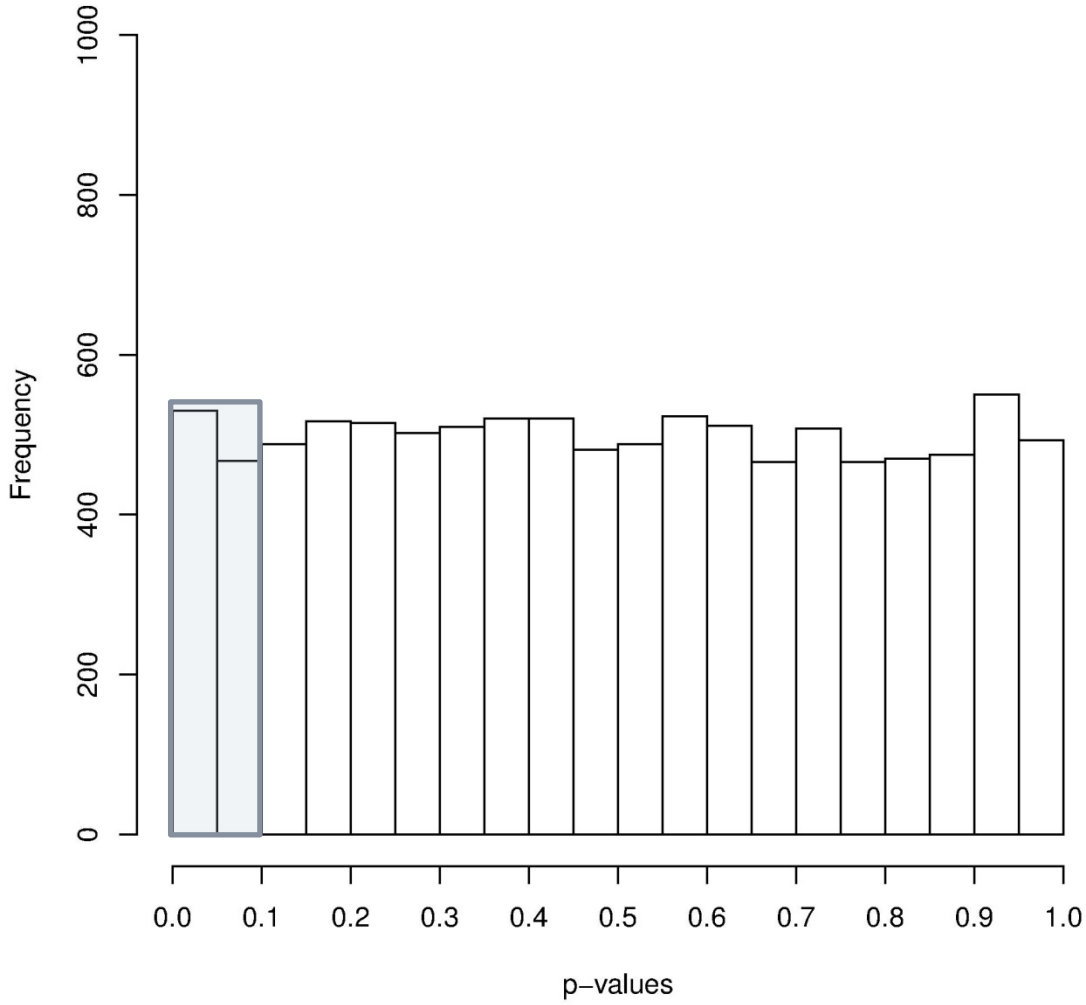
Psychiatric disorders

- (Almost) no major genes identified
- How much variance is explained by non-significant SNPs ?
  - GWAS in schizophrenia
  - Select SNPs using varying cutoff
    - $p<0.1$; $p<0.2$; $p<0.3$; $p<0.4$; $p<0.5$
  - Build regression model using included SNPs
  - Use regression model to predict disease status in independent populations

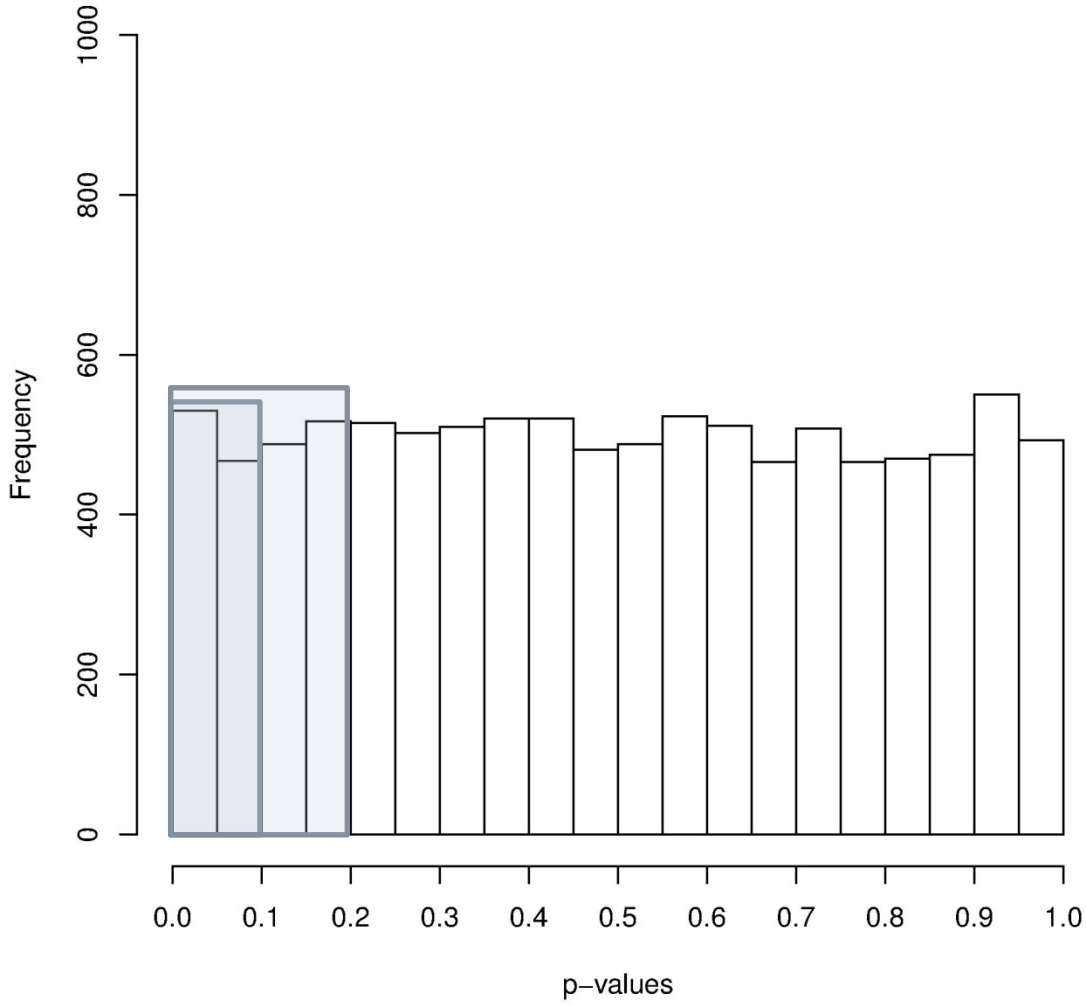
P-value distribution



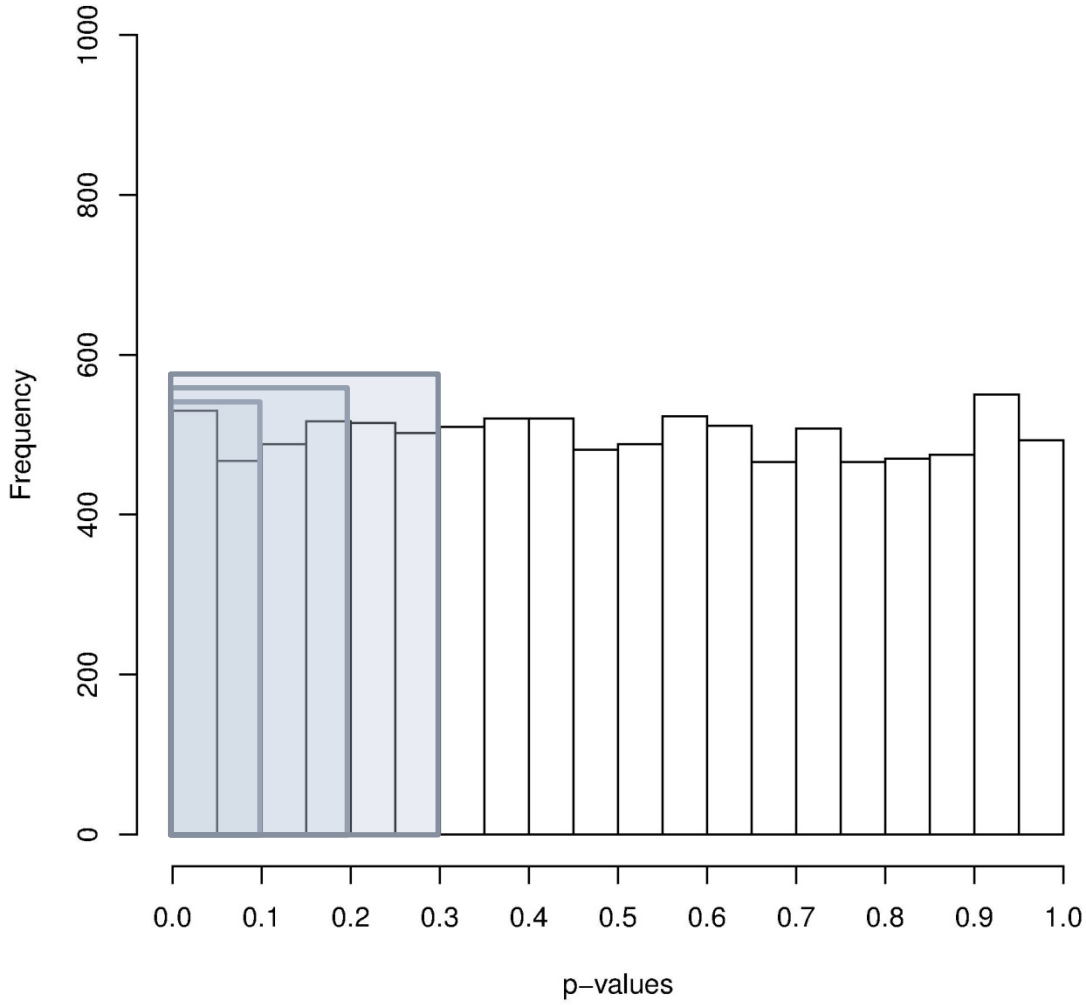
P-value distribution



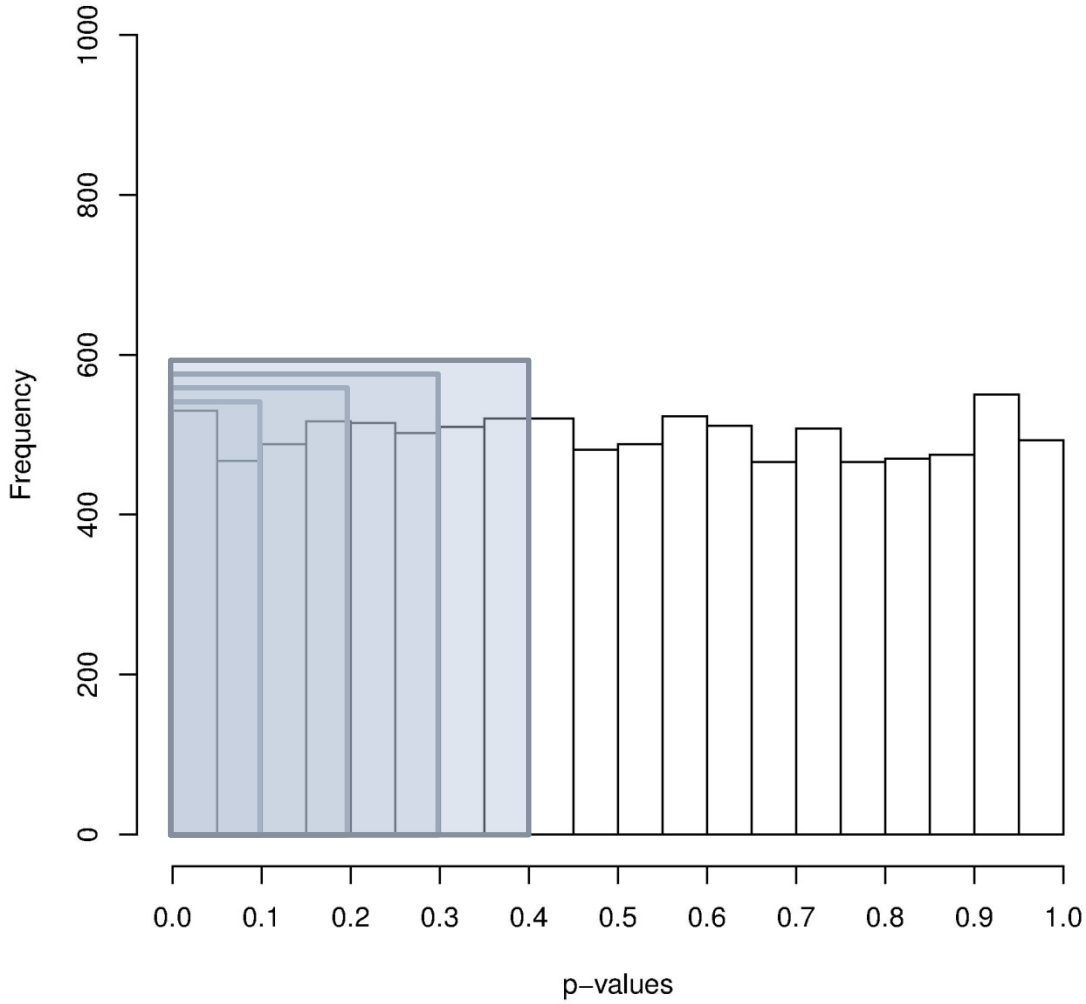
P-value distribution



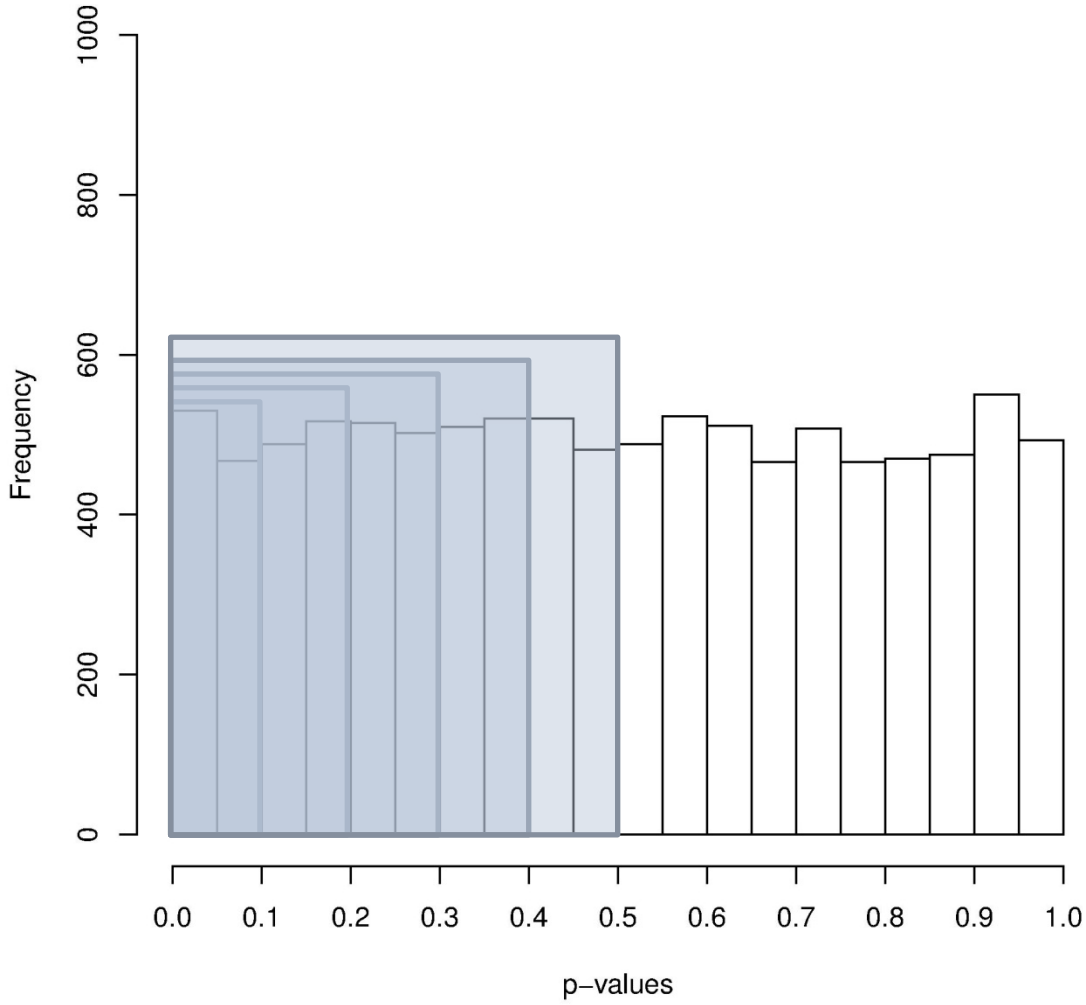
P-value distribution



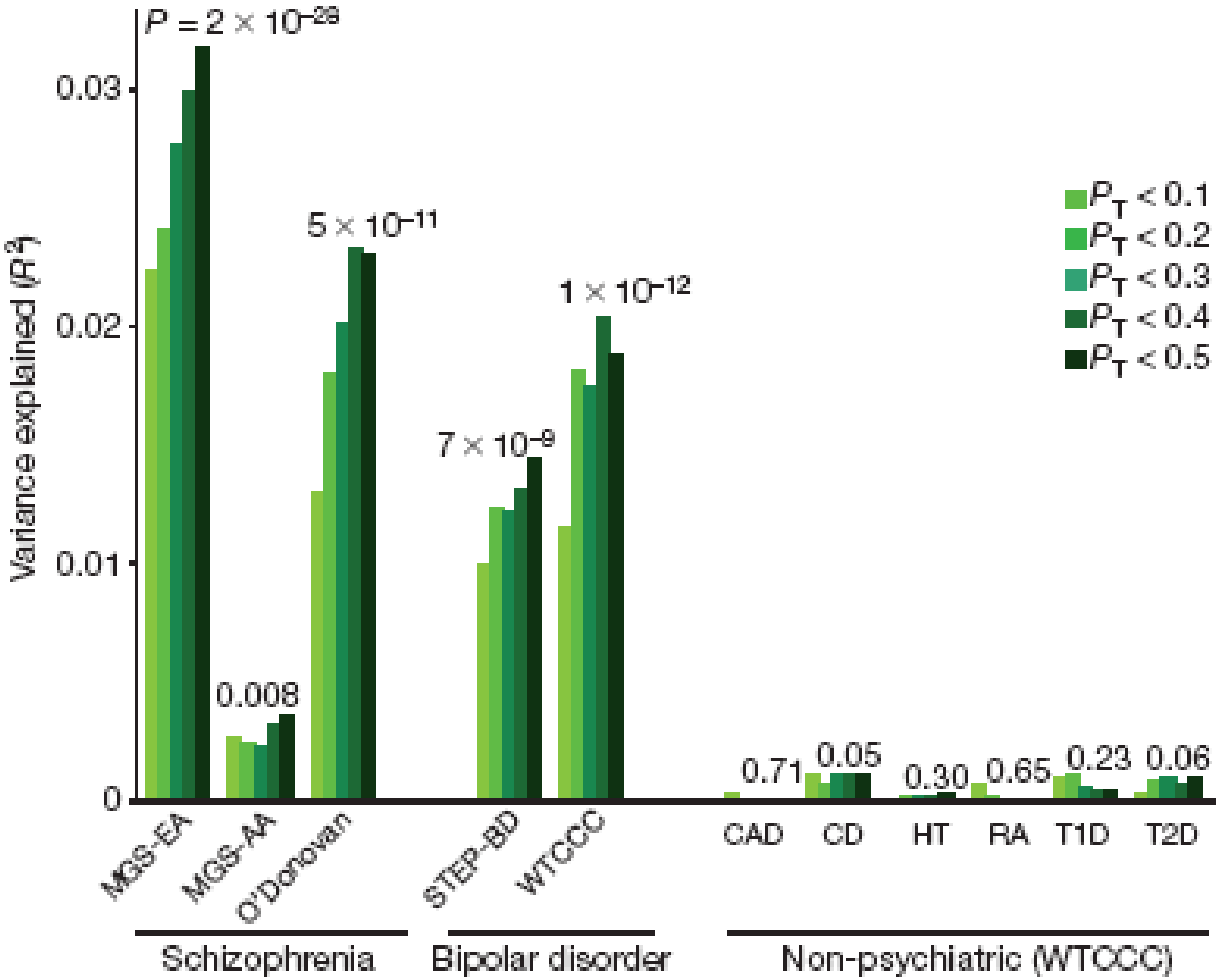
P-value distribution



P-value distribution



Prediction of phenotype



Conclusion

- Low-significant SNPs enriched for causal alleles
  - SNPs (for schizophrenia) with low significance can predict disease status (for schizophrenia) in independent population
  - Schizophrenia SNPs also predicting bipolar disorder

Future prospect

- Prediction model built using SNPs from GWAS
- Higher sample size :
  - → SNPs with p<< more enriched of truly causative SNPs
  - → Prediction model becomes more accurate

→ Polygenic risk score may become diagnostic tool to predict phenotype (even in absence of individually pinpointed SNPs)