Psych 3102
Introduction to Behavior Genetics

Lecture 14

Identifying genes in humans

Review of vocabulary:

markers variable loci (polymorphisms) of known location people can be genotyped to see which marker alleles they have at these loci

SNPs single nucleotide polymorphisms 8 million now available

- these make good markers for association studies but only 2 alleles possible
- for linkage studies, the ideal marker has many alleles (eg 10) equally frequent in population

genetic linkage tendency of a short chromosomal segment to be inherited intact from parent to offspring

- used in linkage methods

haplotype the combination of alleles inherited together

-stay together over many generations, only broken up by recombination

allelic association excessive co-occurrence of a particular combination of alleles due to tight linkage (or other reasons)

- used in association methods

Humans

- not possible to manipulate genes
- not possible to design matings
- not possible to eliminate environmental effects that may influence gene effects
- forced to deal with naturally occurring genetic and environmental variation
- results from research WILL generalize to world outside lab, more likely to be clinically relevant for diagnoses, treatment, unlike some animal research

Success so far:

Identifying genes for single gene disorders
Identifying QTLs for some medical conditions

macular degeneration IBS type 2 diabetes
blood group O allele associated with duodenal ulcers - very small effect, only 1% of variance

Meta-analysis identifies 29 additional ulcerative colitis risk loci, increasing the number of confirmed associations to 47

Carl A Anderson[1*], Gabrielle Boucher[2,3,80], Charlie W Lees[4,80], Andre Franke[5,80], Mauro D'Amato[6,80], Kent D Taylor[7], James C Lee[8], Philippe Goyette[2,3], Marcin Imielinski[9], Anna Latiano[10], Caroline Lagacé[2,3], Regan Scott[11], Leila Amininejad[12], Suzannah Bumpstead[1], Leonard Baidoo[11], Robert N Baldassano[13], Murray Barclay[14], Theodore M Bayless[15], Stephan Brand[16], Carsten Büning[17], Jean-Frédéric Colombel[18], Lee A Denson[19], Martine De Vos[20], Marla Dubinsky[21], Cathryn Edwards[22], David Ellinghaus[5], Rudolf S N Fehrmann[23], James A B Floyd[1], Timothy Florin[24], Denis Franchimont[25], Lude Franke[23], Michel Georges[26], Jürgen Glas[16], Nicole L Glazer[27], Stephen L Guthery[28], Talin Haritunians[29], Nicholas K Hayward[30], Jean-Pierre Hugot[31], Gilles Jobin[2,32], Debby Laukens[20], Ian Lawrance[33], Marc Lémann[34], Arie Levine[35], Cecile Libioulle[36], Edouard Louis[36], Dermot P McGovern[7,29], Monica Milla[37], Grant W Montgomery[38], Katherine I Morley[1], Craig Mowat[39], Aylwin Ng[40,41], William Newman[42], Roel A Ophoff[43], Laura Papi[44], Orazio Palmieri[10], Laurent Peyrin-Biroulet[45], Julián Panés[46], Anne Phillips[39], Natalie J Prescott[47], Deborah D Proctor[48], Rebecca Roberts[14], Richard Russell[49], Paul Rutgeerts[50], Jeremy Sanderson[51], Miquel Sans[52], Philip Schumm[53], Frank Seibold[54], Yashoda Sharma[48], Lisa A Simms[55], Mark Seielstad[56,57], A Hillary Steinhart[58], Stephan R Targan[7], Leonard H van den Berg[59], Morten Vatn[60], Hein Verspaget[61], Thomas Walters[62], Cisca Wijmenga[23], David C Wilson[49,63], Harm-Jan Westra[23], Ramnik J Xavier[40,41], Zhen Z Zhao[38], Cyriel Y Ponsioen[64], Vibeke Andersen[65], Leif Torkvist[66], Maria Gazouli[67], Nicholas P Anagnou[67], Tom H Karlsen[60], Limas Kupcinskas[68], Jurgita Sventoraityte[68], John C Mansfield[69], Subra Kugathasan[70], Mark S Silverberg[58], Jonas Halfvarson[71], Jerome I Rotter[29], Christopher G Mathew[47], Anne M Griffiths[62], Richard Gearry[14], Tariq Ahmad[72], Steven R Brant[15], Mathias Chamaillard[73], Jack Satsangi[4], Judy H Cho[48,74], Stefan Schreiber[5,75], Mark J Daly[76], Jeffrey C Barrett[1], Miles Parkes[8], Vito Annese[10,37], Hakon Hakonarson[13,77,81], Graham Radford-Smith[55,81], Richard H Duerr[11,78,81], Séverine Vermeire[50,81], Rinse K Weersma[79,81] & John D Rioux[2,3]

Genome-wide association studies and candidate gene studies in ulcerative colitis have identified 18 susceptibility loci. We conducted a meta-analysis of six ulcerative colitis genome-wide association study datasets, comprising 6,687 cases and 19,718 controls, and followed up the top association signals in 9,628 cases and 12,917 controls. We identified 29 additional risk loci ($P < 5 \times 10^{-8}$), increasing the number of ulcerative colitis–associated loci to 47. After annotating associated regions using GRAIL, expression quantitative trait loci data and correlations with non-synonymous SNPs, we identified many candidate genes that provide potentially important insights into disease pathogenesis, including *IL1R2*, *IL8RA-IL8RB*, *IL7R*, *IL12B*, *DAP*, *PRDM1*, *JAK2*, *IRF5*, *GNA12* and *LSP1*. The total number of confirmed inflammatory bowel disease risk loci is now 99, including a minimum of 28 shared association signals between Crohn's disease and ulcerative colitis.

Ulcerative colitis and Crohn's disease represent the two major forms of inflammatory bowel disease (IBD, MIM#266600), which together affect approximately 1 in 250 people in Europe, North America and Australasia. Clinical features, epidemiological data and genetic evidence suggest that ulcerative colitis and Crohn's disease are related polygenic diseases. In contrast to Crohn's disease, bowel inflammation in ulcerative colitis is limited to the colonic mucosa. Although disease-related mortality is low, morbidity remains high, and 10%–20% of affected individuals will undergo colectomy. Though the precise etiology is unknown, the current hypothesis is a dysregulated mucosal immune response to commensal gut flora in genetically susceptible individuals[1]. Recent genome-wide and candidate gene association studies have identified 18 susceptibility loci for ulcerative colitis, including seven that overlap with Crohn's disease (for example, *IL23* pathway genes, *NKX2-3* and *IL10*). Established risk loci specific for ulcerative colitis (*HNF4A*, *CDH1* and *LAMB1*) have highlighted the role of defective barrier function in disease pathogenesis[2].

*A full list of author affiliations appears at the end of the paper.

Received 15 September 2010; accepted 14 January 2011; published online 6 February 2011; doi:10.1038/ng.764

99 loci accounts for ~16% of genetic variance

Molecular Psychiatry (2011) 16, 321–332

www.nature.com/mp

ORIGINAL ARTICLE

Genomewide pharmacogenomic study of metabolic side effects to antipsychotic drugs

DE Adkins[1], K Åberg[1], JL McClay[1], J Bukszár[1], Z Zhao[2], P Jia[2], TS Stroup[3], D Perkins[3], JP McEvoy[4], JA Lieberman[5], PF Sullivan[6,7] and EJCG van den Oord[1]

[1]Center for Biomarker Research and Personalized Medicine, School of Pharmacy, Medical College of Virginia of Virginia Commonwealth University, Richmond, VA, USA; [2]Department of Biomedical Informatics, Psychiatry, and Cancer Biology, Vanderbilt University Medical Center, Nashville, TN, USA; [3]Department of Psychiatry, University of North Carolina at Chapel Hill, NC, USA; [4]Department of Psychiatry, Duke University, Durham, NC, USA; [5]Department of Psychiatry, Columbia University, New York, NY, USA; [6]Department of Genetics, Psychiatry, & Epidemiology, University of North Carolina at Chapel Hill, NC, USA and [7]Department of Medical Epidemiology & Biostatistics, Karolinska Institutet, Stockholm, Sweden

Understanding individual differences in the susceptibility to metabolic side effects as a response to antipsychotic therapy is essential to optimize the treatment of schizophrenia. Here, we perform genomewide association studies (GWAS) to search for genetic variation affecting the susceptibility to metabolic side effects. The analysis sample consisted of 738 schizophrenia patients, successfully genotyped for 492K single nucleotide polymorphisms (SNPs), from the genomic subsample of the Clinical Antipsychotic Trial of Intervention Effectiveness study. Outcomes included 12 indicators of metabolic side effects, quantifying antipsychotic-induced change in weight, blood lipids, glucose and hemoglobin A1c, blood pressure and heart rate. Our criterion for genomewide significance was a pre-specified threshold that ensures, on average, only 10% of the significant findings are false discoveries. A total of 21 SNPs satisfied this criterion. The top finding indicated that a SNP in Meis homeobox 2 (*MEIS2*) mediated the effects of risperidone on hip circumference ($q=0.004$). The same SNP was also found to mediate risperidone's effect on waist circumference ($q=0.055$). Genomewide significant finding were also found for SNPs in *PRKAR2B*, *GPR98*, *FHOD3*, *RNF144A*, *ASTN2*, *SOX5* and *ATF7IP2*, as well as in several intergenic markers. *PRKAR2B* and *MEIS2* both have previous research indicating metabolic involvement, and *PRKAR2B* has previously been shown to mediate antipsychotic response. Although our findings require replication and functional validation, this study shows the potential of GWAS to discover genes and pathways that potentially mediate adverse effects of antipsychotic medication.
Molecular Psychiatry (2011) **16,** 321–332; doi:10.1038/mp.2010.14; published online 2 March 2010

Keywords: genomewide association; antipsychotics; pharmacogenomics; personalized medicine; metabolic side effects

Introduction

Antipsychotics are the cornerstone for acute and long-term treatment for schizophrenia.[1,2] The first generation, sometimes referred to as the 'typical' antipsychotics (for example, haloperidol) was introduced in the 1950s. Despite treatment with these first-generation antipsychotics, a substantial proportion of schizophrenia patients do not improve or relapse frequently. Furthermore, these drugs are often associated with significant side effects, including extrapyramidal symptoms—involuntary movements that may occur in schizophrenia patients after long-term treatment with antipsychotic medication. Tardive dyskinesia is a particularly worrisome extrapyramidal symptom because of its high annual incidence rates[3] and potential irreversibility.[4]

Clozapine was reintroduced in the year 1989, marking the advent of a second generation of 'atypical' antipsychotics.[5,6] It has enhanced therapeutic effects in patients who respond poorly to treatment and has a much lower risk of side effects such as tardive dyskinesia. Clozapine has, however, been associated with severe agranulocytosis, necessitating hematological monitoring and making it unsuitable as a first-line drug. Clozapine's success stimulated efforts to develop new antipsychotics, resulting in other second-generation drugs such as risperidone and olanzapine.[7] These newer second-generation drugs differ pharmacologically from first-generation antipsychotics principally in their lower affinity for

Correspondence: Dr DE Adkins, Department of Psychiatry, Virginia Commonwealth University, McGuire Hall, PO Box 980533, Room 216A, 1112 East Clay Street, Richmond, VA 23223, USA.
E-mail: deadkins@vcu.edu
Received 30 July 2009; revised 9 December 2009; accepted 20 January 2010; published online 2 March 2010

Size of individual gene effects

For polygenic traits

- some of the largest gene effects seen for macular degeneration

5 variants in 3 genes account for ~50% of genetic variation

- smallest effect sizes for height

180 loci explain about 13% of genetic variation

Few polygenic traits have all heritability accounted for ('missing heritability' dilemma)

- current GWAS SNPs and sample sizes not adequate to detect very small effect sizes the larger the sample, the more heritability accounted for
- epistasis?
- G X E?

Recent study in yeast – polygenic traits, able to account for 100% of genetic variance, epistasis ranged from 0 – 50% of genetic variance depending on trait. Huge sample size, environment held constant (Bloom et al, Nature, Feb 2013)

Linkage methods : single gene disorders

- linkage will result in alleles of loci that are close together on a chromosome being passed on together down the generations within a population of related individuals

- test for cosegregation (cotransmission) of a DNA marker allele along with an inferred disease locus in individuals in a large pedigree
- locus is inferred by looking for affected phenotype

- provides only an approximate location of a gene for the trait

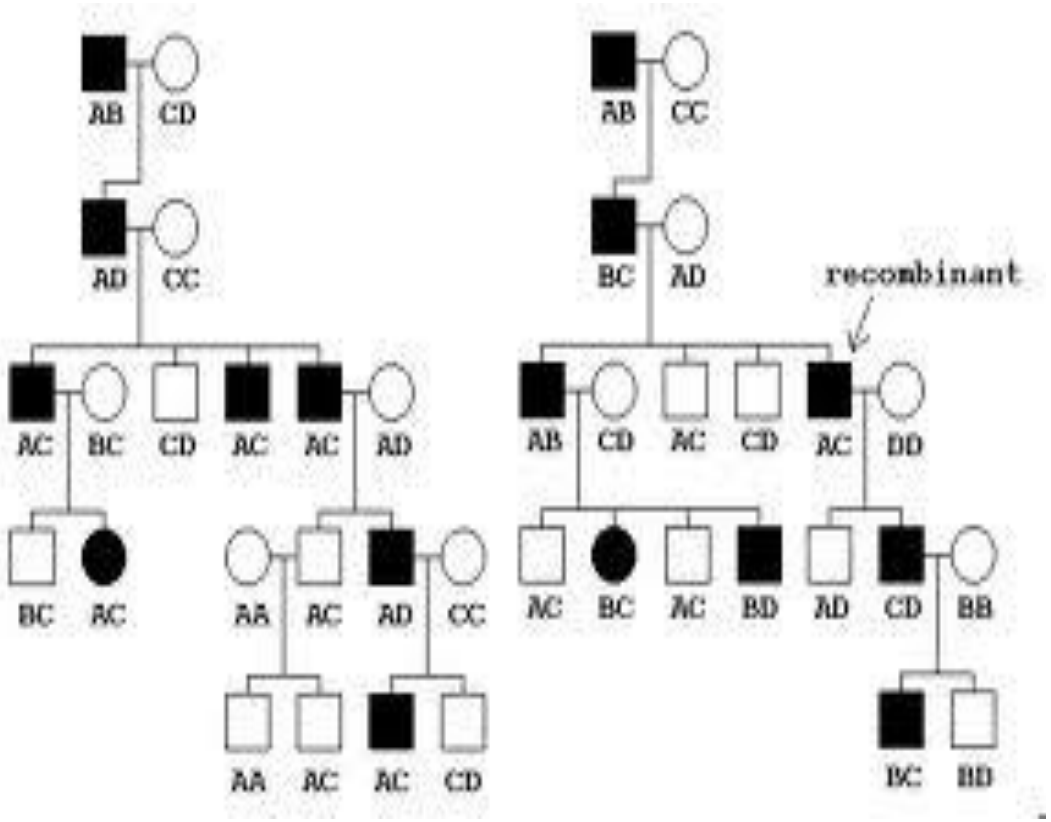
5cM region several different genes

Examples: location of genes for Huntingtons, fragile X, PKU

Genome wide linkage analysis

- large number of markers (SNPs) now available makes it possible to systematically search the genome for markers linked to phenotypes
- small effect of each gene in a complex trait makes this difficult in practice
- linkage methods are not powerful enough to detect genes of very small effect

2 Huntington pedigrees showing marker alleles



Family 1 Family 2

which marker allele is linked to the disease locus in family 1?

which marker allele is initially linked to the disease locus in family 2?

Linkage methods: complex traits

Much larger sample sizes needed than for single gene traits

Linkage analysis using pedigrees not powerful enough

Allele-sharing (affected sib-pair) QTL linkage design

- can be used for dichotomous traits or quantitative traits
- uses pairs of sibs from many different families, allows larger sample sizes
- look for over-representation of markers in sib-pairs that both have the trait of interest or are more similar for a quantitative trait
- expect 50% of sib-pairs to share a marker, even if not linked with the trait
- based on identity-by-descent (ibd)

Example: identification and replication of linkage for reading disability on chr 6 (6p21, Cardon et al, 1994), since replicated many times

Association methods: candidate genes

- look for association between particular allelic variants – often SNPs alleles - within a gene and variation in phenotype for the trait the gene is suspected of affecting
- not systematic
- needs candidate loci (genes suspected of being influential in the trait)
- related individuals not needed
- more powerful than linkage

Examples:

Replicated association of a risk allele for late-onset Alzheimers disease
apolipoprotein E gene, chr 19
risk allele present in 40% of cases, only 15% of controls

Replicated association of DRD4 7-repeat allele with risk for ADHD
risk allele present in 25% of cases, 15% of controls

for dichotomous traits

- use chi-square test with null hypothesis of NO association (ie. no difference in marker occurrence between cases and controls)
- significant result indicates allele IS associated with the trait

Genome-wide association studies (GWAS)

- systematic search of the genome
- very large number of SNPs densely distributed across entire genome are used as markers, essentially using every SNP location as a ‘candidate gene’
- use of microarrays capable of genotyping millions of SNPs at once makes this possible
- SNPs located close together are inherited together in haplotype blocks, allows imputation and a reduction in number of markers genotyped (only tag SNPs genotyped)
- identify which SNPs are associated with phenotype
- Weaknesses:

  marker itself either has to be risk allele (direct association) or very close to it (indirect association or linkage disequilibrium)

  SNP coverage discovered to be inadequate, even when millions of SNPs used (genome sequencing may solve this problem)

Molecular Psychiatry (2013) 18, 2–14

www.nature.com/mp

LETTERS TO THE EDITOR

Power in GWAS: lifting the curse of the clinical cut-off

Molecular Psychiatry (2013) **18,** 2–3; doi:10.1038/mp.2012.65; published online 22 May 2012

Although genome-wide association studies (GWAS), in general, facilitated important discovery of new biological knowledge about diseases,[1–3] identified variants for psychiatric disorders explain little variation, and insight into the role of genes in highly heritable psychiatric traits remains poor.[4,5] Low statistical power is seen as the main reason for the failure to locate more variants, and has resulted in a call for larger samples. Our present study, however, shows that better use of (available) phenotypic information can also increase power considerably.

Many GWAS of psychiatric disorders involve comparisons of allele frequencies between cases and controls. This binary phenotype is usually based on (diagnostic) questionnaire data (e.g., Center for Epidemiologic Studies Depression Scale (CES-D),[6] Conners' Rating Scales[7]) or the number of endorsed symptoms as listed in the Diagnostic and Statistical Manual of Mental Disorders.[8] Such clinical measures are developed to facilitate diagnosis and include items referring to extreme behavior (e.g., 'I had crying spells', 'Temper outbursts and unpredictable behavior'). Although these items distinguish well between cases and controls, they provide little information about phenotypic differences among controls who rarely report such extreme behavior. Indeed, if subjects answer initial screening questions negatively, entire sections of diagnostic questionnaires are regularly skipped under the assumption that none of the items in these sections will be endorsed. Consequently, although controls clearly do show phenotypic variation, they receive very similar test scores (calculated across all items), i.e., the existing phenotypic variation is not well represented in the test scores. Scores on diagnostic instruments are therefore highly skewed in the general population in which controls are in the majority.

Rather than using these skewed test scores as dependent variable in GWAS, however, researchers generally apply clinical cut-off rules to obtain a case–control status measure, for example, subjects scoring $\geqslant 16$ on the CES-D are considered depression cases (obtaining phenotypic score = 1), whereas subjects scoring <16 are considered controls (obtaining phenotypic score = 0). Such cut-off criteria, yielding, for instance, ~88% controls and ~12% cases, are clinically relevant, but detrimental for the power of GWAS to identify causal variants. After all, the common-trait common-variant hypothesis underlying GWAS predicts that causal variants may also produce appreciable phenotypic variation among controls, whose phenotypic scores are now all coded as 0. Application of a clinical cut-off thus reduces phenotypic variation, and consequently diminishes the power to detect genetic effects.

To illustrate this detrimental effect, we simulated test scores for 5000 subjects from a general population on an instrument consisting of 30 extreme items (see Supplementary Material for details). This skewed sum score was then categorized into two groups (using cut-off criteria at 50 or 88%, the latter corresponding to a clinical cut-off), or three groups (each representing ~33% of the sample). In addition, we simulated test scores based on 30 items that covered the entire phenotypic range (i.e., −3.5 to +3.5, assuming standard normal distribution), resulting in an instrument that also measures phenotypic differences among controls (see Supplementary Material for suggestions on how to create such an instrument). This sum score is normally distributed in the general population. Ten causal single-nucleotide polymorphisms (SNPs) with minor allele frequency (MAF) of 0.2 were simulated, explaining 0.2 to 2% of the variance in the underlying trait. This simulation was repeated 2000 times, and we counted the number of times the SNPs were detected ($\alpha = 1e\text{-}07$) using the 5 different phenotypic measures (see Supplementary Material for results of various additional phenotypic operationalizations). Figure 1 shows that the power to detect the genetic variants varies widely across the different phenotypic measures: a skewed sum score (1) performs worse than a normally distributed sum score (5), and the power decreases dramatically when the skewed sum score is categorized (2–4), especially when a clinical cut-off criterion is used (2). Clearly, when the trait is polygenic (rather than Mendelian), and cases and controls differ quantitatively, as stipulated in the common-trait common-variant hypothesis underlying GWAS, the test statistic associated with the correlation between the genotype and the case–control phenotype, is generally smaller compared with the test statistics associated with the other phenotypic measures. This is mainly due to the larger s.e. of the estimate. Consequently, the power to detect the causal locus drops dramatically. Similar results were obtained for MAF = 0.05 and 0.5, and for varying levels of inter-item correlations (Supplementary Material).

In conclusion, the power to detect a trait-associated SNP diminishes considerably if clinical cut-off criteria are used to dichotomize skewed trait measures before analysis. If diagnostic

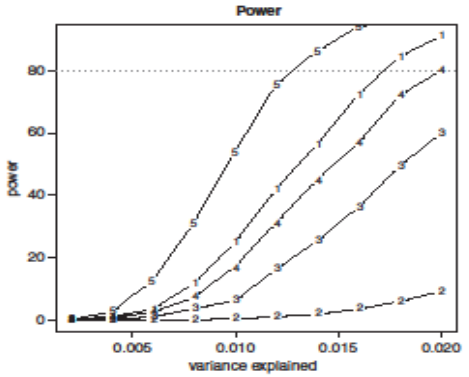
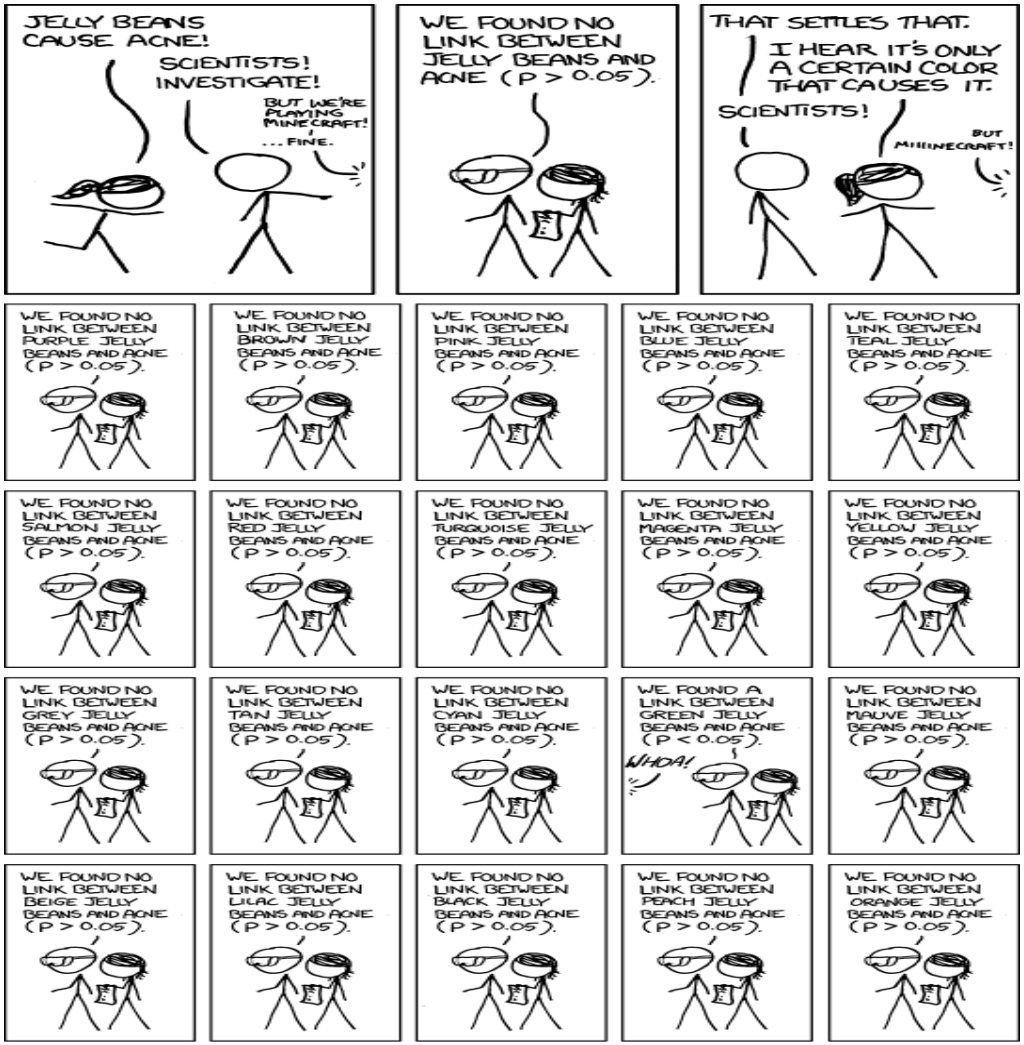


Figure 1. Power plots: on the *x* axis the effect sizes of the 10 simulated causal genetic variants (MAF = 0.2), which explain 0.2 to 2% of the trait variance, and on the *y* axis the power to detect these causal variants for the following phenotypic operationalizations: (1) skewed sum score based on 30 extreme items (as in diagnostic instruments), (2) 88–12% categorized skewed sum score (corresponding to clinical cut-off), (3) 50–50% categorized skewed sum score, (4) 33–33–33% categorized skewed sum score, and (5) normally distributed sum score based on 30 items covering the entire phenotypic range.


JELLY BEANS CAUSE ACNE!
SCIENTISTS! INVESTIGATE!
BUT WE'RE PLAYING MINECRAFT!
... FINE.
WE FOUND NO LINK BETWEEN JELLY BEANS AND ACNE (P > 0.05).
THAT SETTLES THAT.
I HEAR IT'S ONLY A CERTAIN COLOR THAT CAUSES IT.
SCIENTISTS!
BUT MIIINECRAFT!
WE FOUND NO LINK BETWEEN PURPLE JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN BROWN JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN PINK JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN BLUE JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN TEAL JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN SALMON JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN RED JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN TURQUOISE JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN MAGENTA JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN YELLOW JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN GREY JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN TAN JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN CYAN JELLY BEANS AND ACNE (P > 0.05).
WE FOUND A LINK BETWEEN GREEN JELLY BEANS AND ACNE (P < 0.05).
WHOA!
WE FOUND NO LINK BETWEEN MAUVE JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN BEIGE JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN LILAC JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN BLACK JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN PEACH JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN ORANGE JELLY BEANS AND ACNE (P > 0.05).
NEWS
GREEN JELLY BEANS LINKED TO ACNE!
95% CONFIDENCE
ONLY 5% CHANCE OF COINCIDENCE!
SCIENTISTS...

In the past, candidate gene association studies were designed poorly, many false positive results published. What were the problems?

1. Linkage and association methods are *non-hypothesis* driven

- prone to false positives and false negatives
  - especially for association studies
  - large number of studies report results that are result of

Type 1 errors : false positives – finding an association when there really is not one, leads to failure to replicate positive findings when study is repeated

$p<0.05$ ½ million SNPs gives expected 25000 false positives

need $p<10^{-7}$ to give expectation of <1 false positive

but taking up **too** stringent significance criteria leads to

Type 2 errors: false negatives - not finding an association when there is one again, leads to failure to replicate negative findings

2. Very small gene effects

3. Too small a sample size

} both reduce power to detect

Need <1m SNPs, <30,000 sample size to reliably detect associations

4. Population stratification

- another cause of errors

- allele frequencies vary across ethnic groups (‘genetic’ populations)

- between group differences will confound search for biologically relevant within group differences

ie. allele frequency differences between cases and controls will be confounded with between ethnic group differences

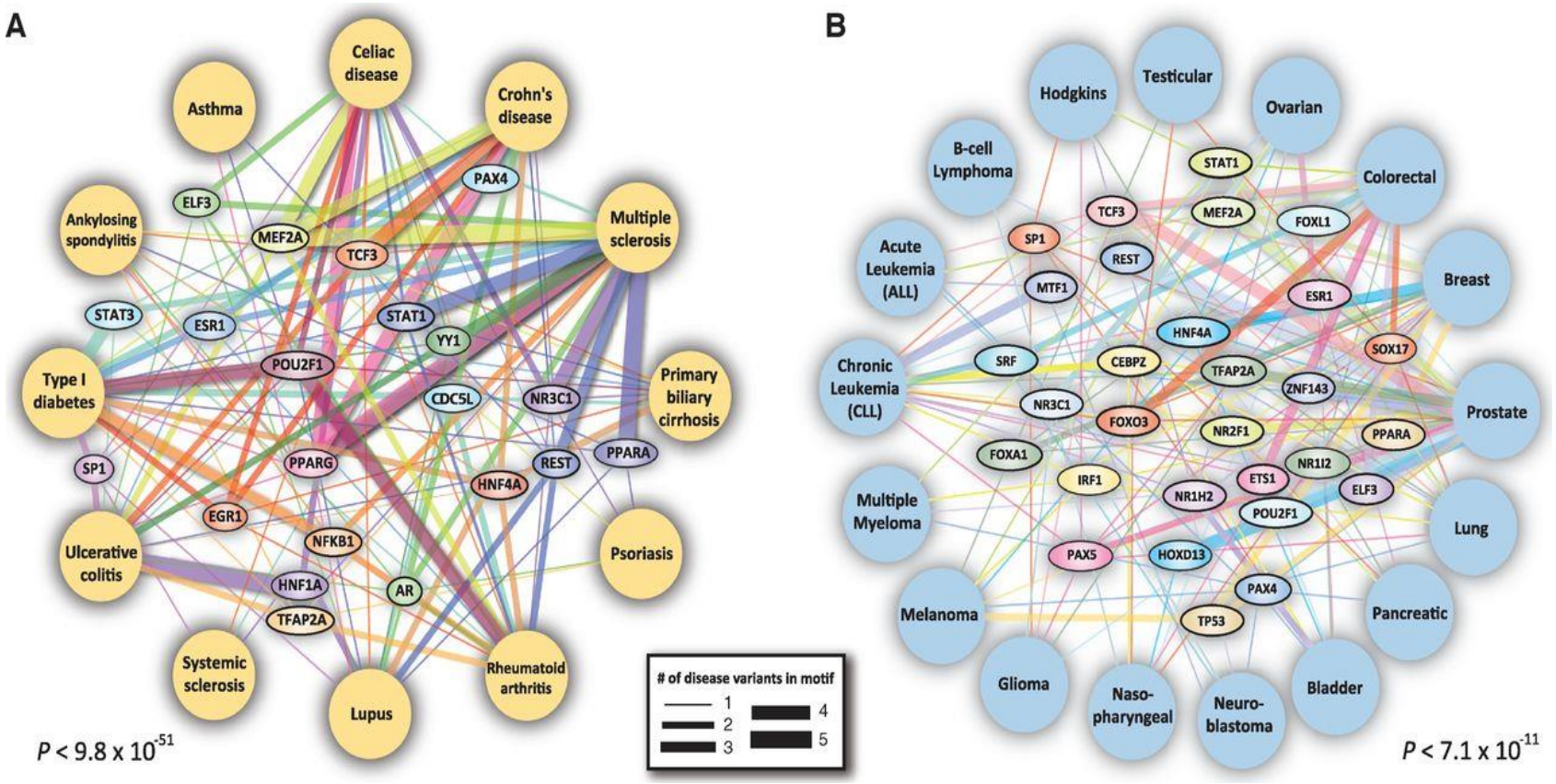
- will produce spurious associations unless control group is from same population as affected group

melanin-producing locus / sickle cell disease

Gene pathway analysis

- based on assumption that risk alleles for a disorder will be found in genes with functions more closely related to each other than random sets of genes
- using results from all other gene-locating methods, use an analytical method to look for nonrandom functional relationships between genes containing risk alleles
- algorithms test whether a given set of loci in the genome is enriched for genetic variants that show some relationship with a disorder compared to a null expectation

Gene pathway to be tested needs to be developed independently from results of gene-finding studies , not biased by including genes found by genetic analysis (post hoc bias)



Common disease networks.

GWAS SNPs from related diseases repeatedly perturb recognition sequences of common transcription factors. Shown are factors whose recognition sequences harbor ≥8 or ≥6 GWAS SNPs in inflammatory or autoimmune diseases (A) and cancer (B), respectively. Edge thickness represents number of associations between transcription factor and disease in DHSs in relevant tissues. Both networks are significantly enriched for overlap with disease-relevant GWAS SNPs and include many well-studied regulators.

Maurano et al, Science 2012, 337, 1190

Maurano et al, Science 2012, 337

Systematic Localization of Common Disease-Associated Variation in Regulatory DNA
Matthew T. Maurano *et al.*
Science 337, 1190 (2012);
DOI: 10.1126/science.1222794

Other recently developed methods

- expression pattern studies - look at actual gene product (mRNA) differences between those with & without disorder, use RNA microarrays
- methylation arrays - look at which genes are methylated to help determine activity, how it relates to disorders
- locating copy number variants (CNVs) - whole genes present as extra copies or missing on one chromosome, thought to be cause of some genetic variation, not recognized by linkage or association methods
- genome sequencing (resequencing) - allows all variation to be studied, not just SNPs

ADDICTION

Knowing when to stop

Polymorphisms in *CHRNA5* — which encodes the α5 subunit of the nicotinic acetylcholine receptor (nAChR) — are associated with an increased risk of tobacco addiction, but the reason for this has remained unclear. Now, Kenny and colleagues show that α5-containing nAChRs are crucially involved in an inhibitory motivational pathway that limits nicotine consumption.

The authors began by studying mice lacking the α5 subunit (*Chrna5*$^{-/-}$ mice) which, like wild-type mice, showed a vigorous response of self administration to low doses of nicotine. However, wild-type mice limited their nicotine intake when higher unit doses were available, whereas knockout mice continued to consume more nicotine. Thus, the inhibitory effects of high doses of nicotine that normally limit nicotine intake are absent in mice lacking α5-containing nAChRs.

The authors next focused on the medial habenula (MHb)–interpeduncular nucleus (IPN) pathway, as it is activated by high doses of nicotine and is enriched for α5-containing nAChRs. Lentiviral delivery of the α5 subunit to the MHb of *Chrna5*$^{-/-}$ mice rescued the inhibitory effect of high-dose nicotine on consumption. Moreover, in rats, delivery of α5-specific short hairpin RNA (α5 shRNA) to the MHb produced an increase in nicotine consumption that was most pronounced at high doses. The increased neuronal activity in the IPN of wild-type mice in response to high-dose nicotine, as measured by FOS immunoreactivity, was almost completely abolished in *Chrna5*$^{-/-}$ mice. Glutamatergic transmission between the MHb and the IPN seems to play a key part in these processes, as pharmacological inhibition of NMDA receptors specifically in these sites increased self administration of nicotine in rats.

It was previously shown that low doses of nicotine reduce the threshold for experiencing reward in a brain-stimulation reward paradigm, whereas high doses of nicotine increase it. However, in α5 shRNA-treated rats, the threshold for reward remained low even if they received high doses of nicotine. Thus, in the absence of α5-containing nAChR signalling, high doses of nicotine do not have an inhibitory effect on reward circuitries.

These studies highlight the importance of the α5 nAChR subunit in limiting nicotine consumption, suggesting that it could be a therapeutic target for smoking cessation.

Katie Kingwell

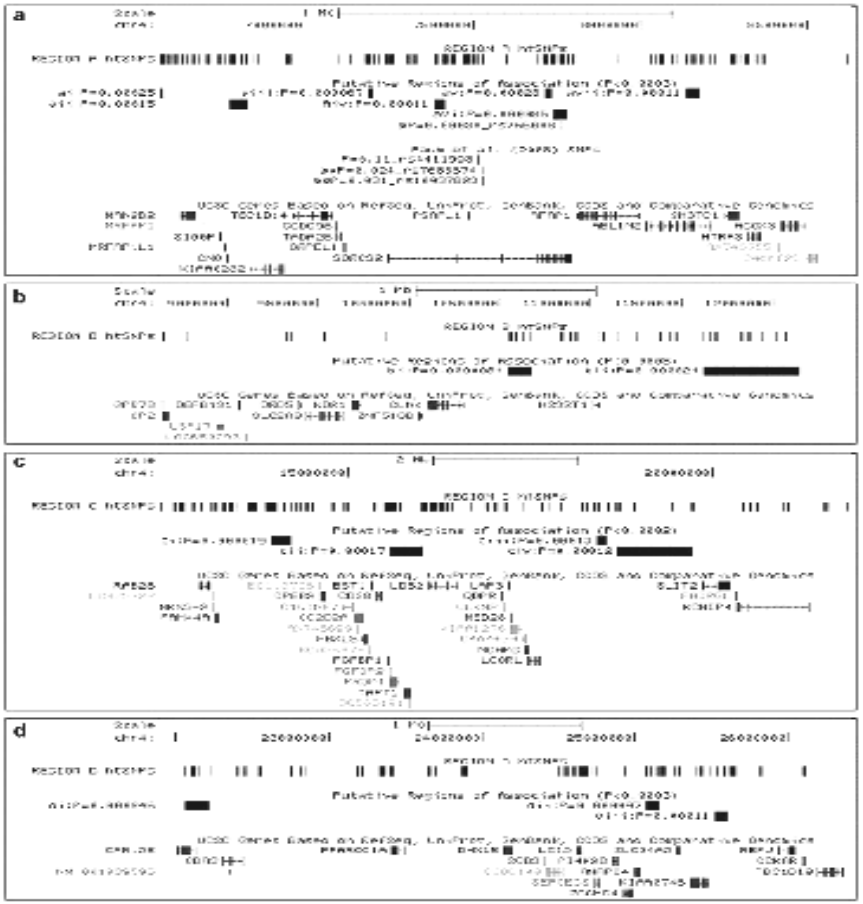
ORIGINAL RESEARCH PAPER Fowler, C. D. *et al.* Habenular α5 nicotinic receptor subunit signalling controls nicotine intake. *Nature* 30 Jan 2011 (doi:10.1038/nature09797)

Convergence of linkage, association and GWAS findings for a candidate region for bipolar disorder and schizophrenia on chromosome 4p

Molecular Psychiatry (2011) **16**, 240–242; doi:10.1038/mp.2010.25; published online 30 March 2010

Several strong candidate genes and regions have been implicated in bipolar disorder (BP) and schizophrenia (SCZ) through linkage and association studies. These disorders have also recently been studied in genome-wide association studies (GWAS), identifying further putative candidate loci, albeit with lower levels of significance and reproducibility than in GWAS of other complex disorders.[1] Our study focuses on a well-established candidate region for psychiatric illness and independently implicates one of the top candidate genes to emerge from GWAS of BP.

The chromosome 4p15–p16 linkage region was first identified in a large Scottish family multiply affected with major affective disorder.[2,3] Subsequently, it has been repeatedly implicated in BP, SCZ and related



Major outcomes of GWAS in human genetics of complex traits

GWAS = largest biological investigations humans have ever conducted

total number of people genotyped to date > 1 million

- most common diseases have highly polygenic architecture (1000's of genes)
- genetic effect sizes of common SNVs (variants) are very small (<0.1%)
- genes and biological processes not previously suspected as being involved have been identified
- some loci are involved in several different diseases once thought to be completely independent in terms of etiology

HUGE sample sizes needed are only made possible by collaborations, often on worldwide scale (eg International Schizophrenia Consortium)

most collaborations were self-organized, emerged rapidly from grassroots origins (actual researchers, not government or business corporations)

Genome-wide association study identifies five new schizophrenia loci

The Schizophrenia Psychiatric Genome-Wide Association Study (GWAS) Consortium[1]

We examined the role of common genetic variation in schizophrenia in a genome-wide association study of substantial size: a stage 1 discovery sample of 21,856 individuals of European ancestry and a stage 2 replication sample of 29,839 independent subjects. The combined stage 1 and 2 analysis yielded genome-wide significant associations with schizophrenia for seven loci, five of which are new (1p21.3, 2q32.3, 8p23.2, 8q21.3 and 10q24.32-q24.33) and two of which have been previously implicated (6p21.32-p22.1 and 18q21.2). The strongest new finding ($P = 1.6 \times 10^{-11}$) was with rs1625579 within an intron of a putative primary transcript for *MIR137* (microRNA 137), a known regulator of neuronal development. Four other schizophrenia loci achieving genome-wide significance contain predicted targets of *MIR137*, suggesting *MIR137*-mediated dysregulation as a previously unknown etiologic mechanism in schizophrenia. In a joint analysis with a bipolar disorder sample (16,374 affected individuals and 14,044 controls), three loci reached genome-wide significance: *CACNA1C* (rs4765905, $P = 7.0 \times 10^{-9}$), *ANK3* (rs10994359, $P = 2.5 \times 10^{-8}$) and the *ITIH3-ITIH4* region (rs2239547, $P = 7.8 \times 10^{-9}$).

In stage 1, we conducted a mega-analysis combining genome-wide association study (GWAS) data from 17 separate studies (with a total of 9,394 cases and 12,462 controls; Table 1 and Supplementary Tables 1,2). We imputed allelic dosages for 1,252,901 autosomal SNPs (Table 1, Supplementary Table 3 and Supplementary Note) using HapMap3 as the reference panel[1]. We tested for association using logistic regression of imputed dosages with sample identifiers and three principal components as covariates to minimize inflation in significance testing caused by population stratification. The quantile-quantile plot (Supplementary Fig. 1) deviated from the null distribution with a population stratification inflation factor of $\lambda = 1.23$. However, λ_{1000}, a metric that standardizes the degree of inflation by sample size, was only 1.02, similar to that observed in other GWAS meta-analyses[2,3]. This deviation persisted despite comprehensive quality control and inclusion of up to 20 principal components (Supplementary Fig. 1). Thus, we interpret this deviation as indicative of a large number of weakly associated SNPs consistent with polygenic inheritance[4]. We also examined 298 ancestry-informative markers (AIMs) that reflect European-ancestry population substructure[5]. Unadjusted analyses showed greater inflation in the test statistics than we saw for all markers (AIMs $\lambda = 2.26$ compared to all markers $\lambda = 1.56$). After inclusion of principal components, the distributions of the test statistics did not differ between AIMs ($\lambda = 1.18$) and all markers ($\lambda = 1.23$), a result inconsistent with population stratification explaining the residual deviation seen in Supplementary Figure 1. Moreover, the results of a meta-analysis using summary results generated using study specific principal components (Supplementary Note) were highly correlated with those from the mega-analysis (Pearson correlation = 0.94, with a similar $\lambda = 1.20$; Supplementary Fig. 2). Of the ten SNPs in Table 2, four increased and six decreased in significance, suggesting that the most extreme values did not result from systematic inflation artifacts. Therefore, our primary analysis used unadjusted *P* values (nevertheless, see Table 2 for stage 1 *P* values adjusted for λ (ref. 6).

In stage 1 (Table 2, Supplementary Table 4 and Supplementary Figs. 3 and 4), 136 associations reached genome-wide significance ($P < 5 \times 10^{-8}$)[7]. The majority of these associations ($N = 129$) mapped to 5.5 Mb in the extended major histocompatibility complex (MHC, 6p21.32-p22.1), a region of high linkage disequilibrium (LD) previously implicated in schizophrenia in a subset of the samples used here[4,8,9]. The other stage 1 regions included new regions (10q24.33 and 8q21.3) and previously reported regions (18q21.2 at *TCF4* (encoding transcription factor 4) and 11q24.2 (ref. 8)). The signal at 11q24.2 is ~0.85 Mb from *NRGN* (encoding neurogranin) and is uncorrelated with the previously associated variant near this gene[8].

In Table 2 and Supplementary Table 4, we denote regions of association by the most significant marker. Associated SNPs with $r^2 \geq 0.2$ in HapMap3 (CEU+TSI populations) were not considered independent. However, we noticed instances where multiple SNPs within 250 kb of each other yielded evidence for association ($P < 10^{-5}$) despite weak LD ($r^2 < 0.2$) between them. For regions with $P < 10^{-6}$, we performed a conditional analysis using as covariates the dosages of the strongest associated SNP, principal components 1–4 and 6 and study indicator. We observed multiple statistically independent signals at the MHC. Although a number of SNPs within the MHC were potentially independent per HapMap r^2 values, only rs9272105 withstood formal conditional analysis, showing $P = 1.8 \times 10^{-6}$ conditional on association to the best SNP, rs2021722 (stage 1 $P = 4.3 \times 10^{-11}$, inter-SNP distance = 2.4 Mb, $r^2 = 0.01$ in HapMap). Excluding the MHC region, we identified six regions with at least one SNP associated at $P < 10^{-5}$ and a second SNP with a conditionally independent $P < 10^{-3}$

[1]A full list of authors and affiliations appears at the end of the paper.

Received 16 February; accepted 19 August; published online 18 September 2011; doi:10.1038/ng.940

Psychology's Bold Initiative

In an unusual attempt at scientific self-examination, psychology researchers are scrutinizing their field's reproducibility

PICK UP THE JANUARY 2008 ISSUE OF *Psychological Science*, turn to page 49, and you'll find a study showing that people are more likely to cheat on a simple laboratory task if they have just read an essay arguing that free will is an illusion. It was a striking study that drew widespread attention both from psychologists and from many media outlets. But should you believe the result?

There's no reason to think that the study, conducted by psychologists Kathleen Vohs of the University of Minnesota Carlson School of Management in Minneapolis, and Jonathan Schooler, who is now at the University of California, Santa Barbara (UCSB), is incorrect. Yet according to many psychologists, their field has a credibility problem at the moment, and it affects thousands of studies like this one.

Part of the angst stems from recent high-profile cases of scientific misconduct, most dramatically the extensive fraud perpetrated by Dutch social psychologist Diederik Stapel (*Science*, 4 November 2011, p. 579), that have cast a harsh light on psychological science. Yet there is no evidence that psychology is more prone to fraud than any other field of science. The greater concern arises from several recent studies that have broadly critiqued psychological research practices, highlighting lax data collection, analysis, and reporting, and decrying a scientific culture that too heavily favors new and counterintuitive ideas over the confirmation of existing results. Some psychology researchers argue that this has led to too many findings that are striking for their novelty and published in respected journals—but are nonetheless false.

As a step toward testing that disturbing idea, one project begun this year offers an online site (PsychFileDrawer.org) where psychologists can quickly and easily post, in brief form, the results of replications of experiments—whether they succeed or fail. University of California, San Diego, psychologist Hal Pashler, one of the project's developers, says the goal is to counteract the "file drawer problem" that plagues all of science, including psychology; researchers usually just file away straightforward replication studies because most journals decline to publish such work.

Double trouble? Brian Nosek leads a large-scale effort to replicate recent psychology studies

In an even more daring effort, a group of more than 50 academic psychologists, which calls itself the Open Science Collaboration (OSC), has begun an unprecedented, large-scale project to systematically replicate psychological experiments recently published in leading journals. "We're wringing our hands worrying about whether reproducibility is a problem or not," says psychologist Brian Nosek of the University of Virginia in Charlottesville, who is coordinating the effort. "If there is a problem, we're going to find out, and then we'll figure out how to fix it."

Robert Kail, a Purdue University developmental psychologist and editor of *Psychological Science*—one of the three journals whose papers the OSC is attempting to replicate—is optimistic that a high percentage of published findings will be replicated. Nonetheless, he views the field's recent attention to the issue of false positives as healthy. "There has been a lot of speculation about the extent to which it's a problem," he says. "But nobody has actually set it up as an empirical project. It's a great thing for somebody to actually do that."

Schooler, who is not directly involved with the project but whose free will study