

Psych 3102

Introduction to Behavior Genetics

Small World

by Tom Briscoe

Lecture 14



Identifying
genes in
humans

Review of vocabulary:

markers

genetic linkage

haplotype the combination of alleles inherited together

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

allelic association

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

For behavioral traits: Identifying aggregate effect of SNPs currently available

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 Taylor⁷, James C Lee⁸, Philippe Goyette^{2,3}, Marcin Imielinski⁹, Anna Latiano¹⁰, Caroline Lagacé^{2,3}, Regan Scott¹¹, Leila Amininejad¹², Suzannah Bumpstead¹, Leonard Baidoo¹¹, Robert N Baldassano¹³, Murray Barclay¹⁴, Theodore M Bayless¹⁵, Stephan Brand¹⁶, Carsten Büning¹⁷, Jean-Frédéric Colombel¹⁸, Lee A Denson¹⁹, Martine De Vos²⁰, Marla Dubinsky²¹, Cathryn Edwards²², David Ellinghaus⁵, Rudolf S N Fehrmann²³, James A B Floyd¹, Timothy Florin²⁴, Denis Franchimont²⁵, Lude Franke²³, Michel Georges²⁶, Jürgen Glas¹⁶, Nicole L Glazer²⁷, Stephen L Guthery²⁸, Talin Haritunians²⁹, Nicholas K Hayward³⁰, Jean-Pierre Hugot³¹, Gilles Jobin^{2,32}, Debby Laukens³⁰, Ian Lawrance³³, Marc Lémann³⁴, Arie Levine³⁵, Cecile Libioulle³⁶, Edouard Louis³⁶, Dermot P McGovern^{7,29}, Monica Milla³⁷, Grant W Montgomery³⁸, Katherine I Morley¹, Craig Mowat³⁹, Aylwin Ng^{40,41}, William Newman⁴², Roel A Ophoff⁴³, Laura Papi⁴⁴, Orazio Palmieri¹⁰, Laurent Peyrin-Biroulet⁴⁵, Julián Panés⁴⁶, Anne Phillips³⁹, Natalie J Prescott⁴⁷, Deborah D Proctor⁴⁸, Rebecca Roberts¹⁴, Richard Russell⁴⁹, Paul Rutgeerts⁵⁰, Jeremy Sanderson⁵¹, Miquel Sans⁵², Philip Schumm⁵³, Frank Seibold⁵⁴, Yashoda Sharma⁴⁸, Lisa A Simms⁵⁵, Mark Seielstad^{56,57}, A Hillary Steinhart⁵⁸, Stephan R Targan⁷, Leonard H van den Berg⁵⁹, Morten Vatn⁶⁰, Hein Verspaget⁶¹, Thomas Walters⁶², Cisca Wijmenga²³, David C Wilson^{49,63}, Harm-Jan Westra²³, Ramnik J Xavier^{40,41}, Zhen Z Zhao³⁸, Cyriel Y Ponsioen⁶⁴, Vibeke Andersen⁶⁵, Leif Torkvist⁶⁶, Maria Gazouli⁶⁷, Nicholas P Anagnou⁶⁷, Tom H Karlsen⁶⁰, Limas Kupcinskas⁶⁸, Jurgita Sventoraityte⁶⁸, John C Mansfield⁶⁹, Subra Kugathasan⁷⁰, Mark S Silverberg⁵⁸, Jonas Halfvarson⁷¹, Jerome I Rotter²⁹, Christopher G Mathew⁴⁷, Anne M Griffiths⁶², Richard Gearry¹⁴, Tariq Ahmad⁷², Steven R Brant¹⁵, Mathias Chamaillard⁷³, Jack Satsangi⁴, Rudy H Cho^{48,74}, Stefan Schreiber^{5,75}, Mark J Daly⁷⁶, Jeffrey C Barrett¹, Miles Parkes⁸, Vito Annesse^{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 ($P < 5 \times 10^{-9}$), 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 individuals¹. 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 pathogenesis².

*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



ORIGINAL ARTICLE

Genomewide pharmacogenomic study of metabolic side effects to antipsychotic drugs

DE Adkins¹, K Åberg¹, JL McClay¹, J Bukszár¹, Z Zhao², P Jia², TS Stroup³, D Perkins³, JP McEvoy⁴, JA Lieberman⁵, PF Sullivan^{6,7} and EJCG van den Oord¹

¹Center for Biomarker Research and Personalized Medicine, School of Pharmacy, Medical College of Virginia of Virginia Commonwealth University, Richmond, VA, USA; ²Department of Biomedical Informatics, Psychiatry, and Cancer Biology, Vanderbilt University Medical Center, Nashville, TN, USA; ³Department of Psychiatry, University of North Carolina at Chapel Hill, NC, USA; ⁴Department of Psychiatry, Duke University, Durham, NC, USA; ⁵Department of Psychiatry, Columbia University, New York, NY, USA; ⁶Department of Genetics, Psychiatry, & Epidemiology, University of North Carolina at Chapel Hill, NC, USA and ⁷Department 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 *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 findings were also found for SNPs in *PRKAR2B*, *GPR98*, *RHOD3*, *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 rates³ and potential irreversibility.⁴

Clozapine was reintroduced in the year 1989, marking the advent of a second generation of 'atypical' antipsychotics.^{5A} 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

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

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

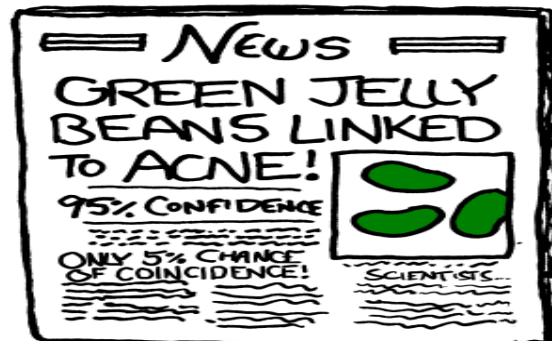
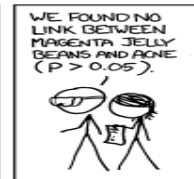
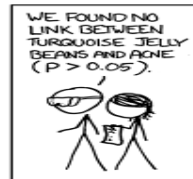
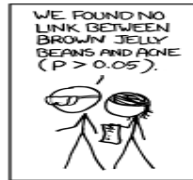
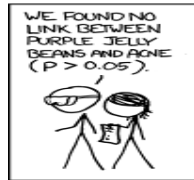
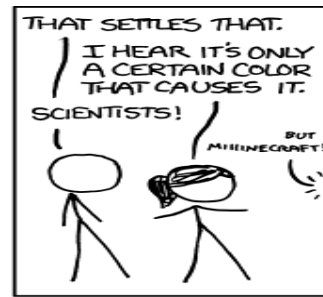
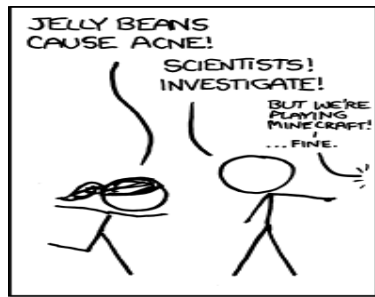
risk allele present in 25% of cases, 15% of controls

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
 - few studies produce replicable results



Why have results been so inconsistent for behavior, few solid genes located, few studies able to be replicated?

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 :

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

Type 2 errors

2. Very small gene effects

3. Too small a sample size

} both reduce power to detect

Need <500,000 SNPs, <10,000 sample size to reliably detect associations

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

-

melanin-producing locus / sickle cell disease

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)

Other recently developed methods

- **expression pattern studies** - look at actual gene product (or mRNA) differences between those with & without disorder
- **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

Knowing when to stop

Polymorphisms in *CHRNA5* — which encodes the $\alpha 5$ 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 $\alpha 5$ -containing nAChRs are crucially involved in an inhibitory motivational pathway that limits nicotine consumption.

The authors began by studying mice lacking the $\alpha 5$ 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 $\alpha 5$ -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 $\alpha 5$ -containing nAChRs. Lentiviral delivery of the $\alpha 5$ subunit to the MHb of *Chrna5*^{-/-} mice rescued the inhibitory effect of high-dose nicotine on consumption. Moreover, in rats, delivery of $\alpha 5$ -specific short hairpin RNA ($\alpha 5$ 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 $\alpha 5$ shRNA-treated rats, the threshold for reward remained low even if they received high doses of nicotine. Thus, in the absence of $\alpha 5$ -containing nAChR signalling, high doses of nicotine do not have an inhibitory effect on reward circuitries.

These studies highlight the importance of the $\alpha 5$ nAChR subunit in limiting nicotine consumption, suggesting that it could be a therapeutic target for smoking cessation.

Katie Klingwell



ORIGINAL RESEARCH PAPER Fowles C. D. et al. Habenular $\alpha 5$ nicotinic receptor subunit signalling controls nicotine intake. *Nature* 30 Jan 2011 | doi:10.1038/nature09797

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 polygenic architecture (1000's of genes)
- genetic effect sizes of common SNP 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)