1-1-2012

Multigene interactions and the prediction of depression in the Wisconsin Longitudinal Study

N S Roetker

J A Yonker

C. Lee

V Chang

J J Basson

See next page for additional authors

Roetker, N., Yonker, J., Lee, C., Chang, V., Basson, J., Roan, C., Hauser, T., Hauser, R., & Atwood, C. (2012). Multigene interactions and the prediction of depression in the Wisconsin Longitudinal Study. BMJ Open, 2(4). Available here

Authors

N S Roetker, J A Yonker, C. Lee, V Chang, J J Basson, C L Roan, T S Hauser, R M Hauser, and Craig Atwood

BMJ Open

Multigene interactions and the prediction of depression in the Wisconsin Longitudinal Study

Nicholas S Roetker, James A Yonker, Chee Lee, et al.

Updated information and services can be found at:

These include:

Data Supplement

"Supplementary Data"

References

This article cites 34 articles, 3 of which can be accessed free at:

Open Access

Email alerting service

Receive free email alerts when new articles cite this article. Sign up in the box at the top right corner of the online article.

Topic Collections

Articles on similar topics can be found in the following collections

Epidemiology (519 articles)
Genetics and genomics (37 articles)
Mental health (154 articles)
Neurology (117 articles)
Public health (472 articles)

To request permissions go to:

To order reprints go to:

To subscribe to BMJ go to:

Notes

Multigene interactions and the prediction of depression in the Wisconsin Longitudinal Study

Nicholas S Roetker,[1] James A Yonker,[1] Chee Lee,[1] Vicky Chang,[1] Jacob J Basson,[3] Carol L Roan,[1] Taissa S Hauser,[1] Robert M Hauser,[1] Craig S Atwood[2,3,4]

[1]Department of Sociology, University of Wisconsin, Madison, Wisconsin, USA
[2]Geriatric Research, Education and Clinical Center, William S. Middleton Memorial Veterans Hospital, Madison, WI
[3]Department of Medicine, University of Wisconsin-Madison School of Medicine and Public Health, Madison, Wisconsin, USA
[4]School of Exercise, Biomedical and Health Sciences, Edith Cowan University, Joondalup, Australia

Correspondence to
Dr Craig S Atwood;
csa@medicine.wisc.edu

ABSTRACT

Objectives: Single genetic loci offer little predictive power for the identification of depression. This study examined whether an analysis of gene—gene (G × G) interactions of 78 single nucleotide polymorphisms (SNPs) in genes associated with depression and age-related diseases would identify significant interactions with increased predictive power for depression.

Design: A retrospective cohort study.

Setting: A survey of participants in the Wisconsin Longitudinal Study.

Participants: A total of 4811 persons (2464 women and 2347 men) who provided saliva for genotyping; the group comes from a randomly selected sample of Wisconsin high school graduates from the class of 1957 as well as a randomly selected sibling, almost all of whom are non-Hispanic white.

Primary outcome measure: Depression as determine by the Composite International Diagnostic Interview—Short-Form.

Results: Using a classification tree approach (recursive partitioning (RP)), the authors identified a number of candidate G × G interactions associated with depression. The primary SNP splits revealed by RP (ANKK1 rs1800497 (also known as DRD2 Taq1A) in men and DRD2 rs224592 in women) were found to be significant as single factors by logistic regression (LR) after controlling for multiple testing ($p=0.001$ for both). Without considering interaction effects, only one of the five subsequent RP splits reached nominal significance in LR (FTO rs1421085 in women, $p=0.008$). However, after controlling for G × G interactions by running LR on RP-specific subsets, every split became significant and grew larger in magnitude (OR (before) → (after): men: GNRH1 novel SNP: (1.43 → 1.57); women: APOC3 rs2854116: (1.28 → 1.55), ACVR2B rs3749386: (1.11 → 2.17), FTO rs1421085: (1.32 → 1.65), IL6 rs1800795: (1.12 → 1.85)).

Conclusions: The results suggest that examining G × G interactions improves the identification of genetic associations predictive of depression. 4 of the SNPs identified in these interactions were located in two pathways well known to impact depression: neurotransmitter (ANKK1 and DRD2) and neuroendocrine (GNRH1 and ACVR2B) signalling. This study demonstrates the utility of RP analysis as an

ARTICLE SUMMARY

Article focus
- Single genetic loci offer little predictive power for the identification of depression. This study examined whether an analysis of G × G interactions of SNPs in genes associated with depression and age-related diseases would identify significant interactions with increased predictive power for depression.

Key messages
- Using a classification tree approach (RP), we identified a number of candidate G × G interactions associated with depression. After controlling for G × G interactions by running LR on RP-specific subsets, every split became significant and grew larger in magnitude. Four of the SNPs identified in these interactions were located in two pathways well known to impact depression: neurotransmitter (ANKK1 and DRD2) and neuroendocrine (GNRH1 and ACVR2B) signalling.

Strengths and limitations of this study
- Our results suggest that examining G × G interactions improves the identification of genetic associations predictive of depression. This study demonstrates the utility of RP analysis as an efficient and powerful exploratory analysis technique for uncovering genetic and molecular pathway interactions associated with disease aetiology.

efficient and powerful exploratory analysis technique for uncovering genetic and molecular pathway interactions associated with disease aetiology.

INTRODUCTION

Depression is a widespread mental disorder associated with a host of undesirable health, social and economic outcomes. One in six Americans is diagnosed with depression in his or her lifetime.[1] While many environmental

factors—such as socioeconomic status, childhood abuse and major life events—have important ties with depression, so too does gender and many genetic and epigenetic factors, making the disorder heterogeneous in nature.[2] Another major risk factor for depression is age, with depression reaching its highest levels in adults aged 80 years and older.[3]

It has been demonstrated from twin studies that genetic factors typically account for 40%−70% of the risk for developing major depressive disorder, and adoption studies have confirmed the role of genetic risk factors in the development of major depressive disorder (see Zubenko *et al*[4] and references therein). Genetic studies, including recent genome-wide association studies (GWAS), have identified genetic alterations in over 50 genes known to be associated with depression.[5] However, individually, the genetic alterations found within these genes (primarily single nucleotide polymorphisms (SNPs)) have little predictive value. There is a similar lack of predictive value from GWAS of other major age-related diseases.[6]

Given this lack of predictive power among individual genetic alterations for depression together with the complex nature of ageing-related diseases, it would seem prudent to examine epistatic effects on this age-related condition. In this respect, we have previously demonstrated that gene-gene (G × G) interactions greatly modulate risk for complex age-related diseases.[7 8] Recent studies of depression also have identified epistatic effects. In particular, associations have been identified between *BDNF* Val66Met (brain-derived neurotrophic factor; rs6265) and *5-HTTLPR* (serotonin transporter-linked promoter region[9]); *GSK3B* rs6782799 (glycogen synthase kinase 3β), *BDNF* rs7124442 and *BDNF* Val66Met[10]; *BDNF* Val66Met and SNPs in *NTRK2* (neurotrophic tyrosine kinase receptor 2[11]); and *5-HTTLPR* short allele and a chromosome 4 gene.[12] The machine learning tool recursive partitioning (RP) has recently been used by Wong *et al*[13] to assess complex G × G interactions in depression. Wong *et al* note that RP is useful in that it quickly explores high-dimensional data for non-linear effects that are non-biased and easily interpretable.

The goals of this study were therefore to (1) explore G × G interactions that might better predict the genetic factors involved in the aetiology of depression and (2) to further demonstrate the utility of machine learning algorithms (RP) to identify genetic interactions. Using genotypic data from the Wisconsin Longitudinal Study (WLS), we identified associations between dopaminergic genes and depression in men and women, as well as G × G interactions involving neuroendocrine signalling pathways, with increased significance compared with single genetic associations.

METHODS
Study participants and surveys
Data were collected from the WLS, a random sample originally comprised 10,317 men and women who graduated from Wisconsin high schools in 1957. Later in 1977, the WLS began interviewing one randomly selected sibling of each graduate, when possible. The cohort reflects the ancestral makeup of the late-1950s Wisconsin population in that participants are almost entirely non-Hispanic white men and women. In general, the sample is broadly representative of older white Americans with at least a high school education.[14] Further characteristics of the WLS cohort may be found in detail elsewhere.[15] Health and psychological well-being phenotypic data were taken from mail and phone surveys given in 2004−2005. Inclusion criteria for depression included any member of the WLS cohort who was depressed according to the Composite International Diagnostic Interview−Short-Form. Individuals who answered YES to the question 'Have you ever had a time in life lasting two weeks or more when nearly every day you felt sad, blue, depressed, or when you lost interest in most things like work, hobbies, or things you usually liked to do for fun?' and whose depression was not caused by alcohol, drugs, medications or physical illness were asked further depression symptom questions. Symptom questions asked whether the 2-week period was accompanied with (1) any weight loss, (2) trouble sleeping, (3) feeling tired, (4) feeling bad upon waking, (5) losing interest, (6) trouble concentrating or (7) thoughts about death. Those answering YES to three or more of these symptom questions were classified as having depression.[16] Those answering YES to two or fewer symptom questions and all those answering NO to the initial stem question were classified as controls.

Genotyping
Seven thousand one hundred and one participants (4569 graduates and 2532 siblings) provided saliva samples in Oragene DNA sample collection kits (DNA Genotek, Kanata, Canada) from which DNA was extracted and genotyped for 78 SNPs that were selected based on their association with depression and age-related conditions and diseases (see supplementary information 1). Genotyping was performed by KBioscience (Hoddesdon, UK) with use of a homogeneous Fluorescent Resonance Energy Transfer technology coupled to competitive allele-specific PCR. All SNP genotypes described in our results were in Hardy−Weinberg equilibrium and their frequencies matched those reported in the literature for European samples.

Statistical analysis
Analyses were limited to the 4811 pooled graduates and siblings for whom we had depression and genotype information (note: individuals with more than 10% missing genotype data were not included). The average age among this sample was just younger than 65 years in 2004. Eighty per cent were married, and the average amount of post-high school educational attainment was 2 years. Median household income in 1993 was $56 700.

Recursive partitioning

RP is a data mining tool for revealing trends that relate a dependent variable (depressed vs non-depressed) to various predictor variables (SNPs). Zhang and Bonney[17] have shown how RP can be used in genetic association studies to identify disease genes. RP helps control for heterogeneity in the population and confounding factors by allowing for the segregation of the sample population according to any condition. Thus, RP is a useful way to handle complex data sets that might confound regression analysis due to the complexity of the relationship between the independent and dependent variables and due to missing information.

RP classification trees (using R package rpart) were used to identify potential interactions among the 78 SNPs in relation to depression. The trees split the data along branches according to the criteria determined by the rpart package algorithm, which is originally based off the work of Breiman's classification and regression trees algorithm.[18] Basically, the classification and regression trees algorithm first considers all depressed and non-depressed subjects pooled together in a heterogeneous root node. Based on considering every possible 'yes–no' binary partition that can be made by each independent variable, the single split, which maximises homogeneity between the two resulting subnodes as compared with the root node, is made. Each subnode can then be treated independently as a new root node for all subsequent splits, and the pattern continues until every subject constitutes a terminal node, resulting in a very large and complex tree. A 10-part cross-validation procedure seeking to minimise misclassification and complexity determines optimal pruning. See Therneau and Atkinson[19] for specific details of the rpart package. Priors were set to 0.5, 0.5. The use surrogate parameter was set to 0 so that subjects missing the primary split variable do not progress further down the tree, and maxsurrogate was set to 0 to cut computation time in half. The threshold complexity parameter was set to 0.01. Tree nodes were re-created in Microsoft Visio to display percentage depressed and the default number of controls/cases as presented by the rpart.

Logistic regression

Variables found in association with depression based on RP analysis were considered in single-factor logistic regression (LR) models, separate by gender, using the specific dichotomous splitting of genotypes as designated by RP trees. Regression models for all seven SNP splits were first run on the full data set to represent single main factor effects. Then, each split was run on the respective subset of data as represented by the preceding RP split criteria. Thus, we attempt to mirror RP splits within a more formal LR framework in order to measure the significance of interactions presented by the trees. Multiple testing of 78 SNPs in RP for both men and women followed by 14 LR models resulted in a modified false discovery rate (FDR) significance level of 0.009.

RESULTS

Of the 4811 participants (2464 women and 2347 men) under examination in this study, we identified 713 participants (481 women and 232 men) with depression (14.8%). Given that the independent variable gender (when included as a factor in the full data set) was the primary split on RP trees; that women are over two times as likely to be diagnosed with depression than men and since the female aetiology of depression has been reported to be associated with unique social, psychological, and biological factors,[20] all subsequent analyses were performed by gender.

Recursive partitioning analysis

To examine multigene interactions for association with depression, we screened our data set using RP. The two-factor RP tree (ANKK1/GNRH1) was the optimised pruning for men (figure 1), while the five-factor tree (DRD2/APOC3/ACVR2B/FTO/IL6) was the optimised pruning for women (figure 2). For more detailed information on the seven SNPs found by RP, see supplementary information 2.

The best overall split for men was ANKK1 rs1800497 (historically known as the DRD2 Taq1A allele), where the incidence of depression increased 2.2-fold in those with
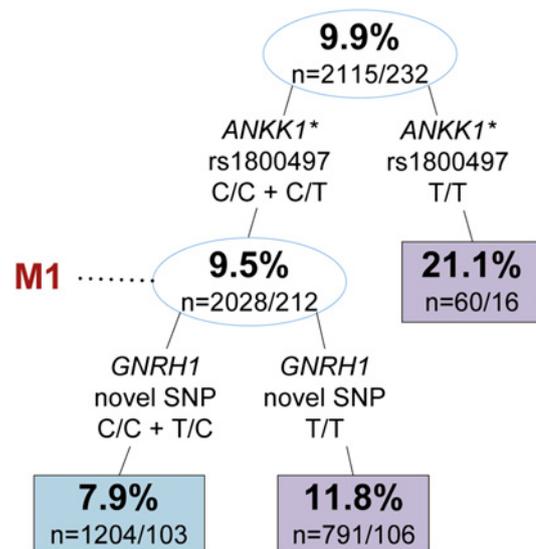

Figure 1 Recursive partitioning tree of Composite International Diagnostic Interview–Short-Form depression in men of the Wisconsin Longitudinal Study. Upper and lower numbers in nodes represent the percentage of participants with depression and the number of controls/cases in that node, respectively. Blue and purple boxes/circles indicate lower and higher rates of depression relative to the primary node, respectively. Split information indicates gene, single nucleotide polymorphism (SNP), and genotype criteria, respectively. M1 is subset of data referenced in table 1. Sensitivity: 0.526, specificity: 0.598, accuracy: 0.591. Due to missing genotype information, we lose approximately 1.5% of participants per split. *rs1800497 is historically known as the DRD2 Taq1A allele.
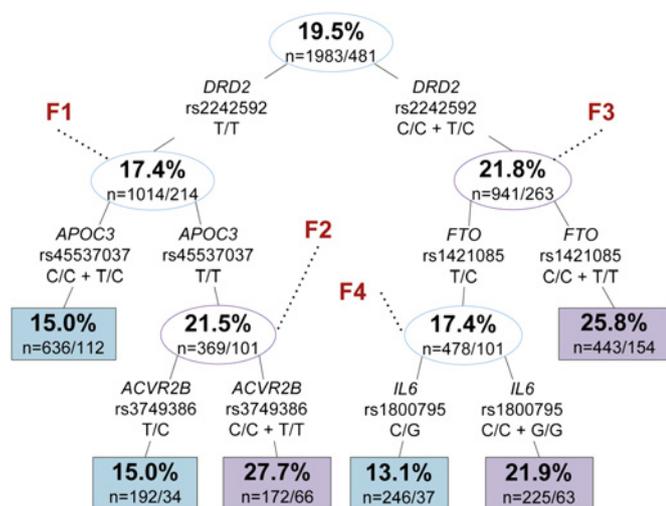
Figure 2 Recursive partitioning tree of Composite International Diagnostic Interview–Short-Form depression in women of the Wisconsin Longitudinal Study. Upper and lower numbers in nodes represent the percentage of participants with depression and the number of controls/cases in that node, respectively. Blue and purple boxes/circles indicate lower and higher rates of depression relative to the primary node, respectively. Split information indicates gene, single nucleotide polymorphism (SNP), and genotype criteria, respectively. F1–F4 are subsets referenced in table 1. Sensitivity: 0.607, specificity: 0.563, accuracy: 0.572. Due to missing genotype information, we lose approximately 1.4% of participants per split.

no C-alleles compared with those with one or two C-alleles. Considering interaction between *ANKK1* and *GNRH1* widened the disparity in incidence, where those with at least one C-allele in both *ANKK1* rs1800497 and the novel SNP in *GNRH1* had a 2.7-fold lower incidence than those without a C-allele in *ANKK1* rs1800497.

For women, the best overall split was *DRD2* rs2242592, where those with one or two C-alleles had 1.3-fold higher incidence of depression compared with those without any C-alleles. G × G interactions associated with the highest incidence of depression included *DRD2* rs2242592 T/T + *APOC3* rs45537037 T/T + *ACVR2B* rs3749386 C/C or T/T, accounting for a 1.4-fold increase in depression compared with baseline incidence.

Single main factor effects
Specific SNP interactions identified by RP were next analysed by LR (see table 1 for full data). The primary SNP splits in men and women were significant at the modified FDR level. Men with no C-alleles for *ANKK1* rs1800497 had 2.55 times higher odds (p=0.001 (95% CI 1.44 to 4.51)) of depression compared with men with at least one C-allele. Women with at least one C-allele for *DRD2* rs2242592 had 1.32 times higher odds (p=0.006 (95% CI 1.08 to 1.62)) of depression compared with women with no C-alleles. One other split reached nominal significance; women homozygous (C/C or T/T) for *FTO* rs1421085 had 1.32 times higher odds

(p=0.008 (95% CI 1.08 to 1.62)) for depression than women with a heterozygous genotype. SNP splits of *GNRH1, APOC3, ACVR2B,* and *IL6* did not significantly associate with depression.

Gene–gene interactions enhance predictability for depression
Specific SNP interactions identified by RP were next analysed by LR as RP-specific subsets (see table 1, RP-subsetted data). All five of the secondary and tertiary RP splits were found to be significant at the modified FDR level when considered as subsets. Among only men with at least one C-allele in *ANKK1* rs1800497, those with no C-allele in the novel SNP of *GNRH1* had 1.57 times higher odds (p=0.002 (95% CI 1.18 to 2.08)) for depression than men with one or two C-alleles. For the subset of women in the first right-hand split of figure 2, those homozygous for *FTO* rs1421085 had 1.65 times higher odds (p=0.0005 (95% CI 1.24 to 2.18)) for depression than women with a heterozygous genotype. For the remaining subset of women in the second right-hand split of figure 2, those homozygous for *IL6* rs1800795 had 1.85 times higher odds (p=0.006 (95% CI 1.19 to 2.89)) for depression than women with a heterozygous genotype. For the subset of women in the first left-hand split of figure 2, those with no C-alleles for *APOC3* rs45537037 had 1.55 times higher odds (p=0.004 (95% CI 1.15 to 2.09)) for depression than women with one or two C-alleles. For the subset of women in the second left-hand split of figure 2, those homozygous for *ACVR2B* rs3749386 had 2.17 times higher odds (p=0.001 (95% CI 1.37 to 3.44)) for depression than women with a heterozygous genotype.

DISCUSSION
Using RP as a screening tool to find potential multigene interactions, followed by verification by LR, our data demonstrate that multigene interactions predict depression with a greater certainty than single main factor associations. RP provided us with primary dichotomous genotype splits in men and women (*ANKK1* rs1800497 and *DRD2* rs2242592, respectively) that were both significant in LR models at the modified FDR level (table 1). When considering the five subsequent RP splits over the entire data set with LR, only one reached a nominal level of significance (barely), which was *FTO* rs1421085 in women. However, after running LR on specific subsets of data according to the pattern of RP branches, every split was found to be significant and every OR grew larger (table 1; OR (before) → (after): male left: 1.43 → 1.57, female left 1: 1.28 → 1.55, female left 2: 1.11 → 2.17, female right 1: 1.32 → 1.65, female right 2: 1.12 → 1.85). Thus, RP provides two unique and important criteria: dichotomous genotype splitting instructions and G × G interaction patterns. These criteria go beyond the traditional single-factor SNP approach to genetic association studies and allow identification of important multigene pathways that more suitably characterise the aetiology of complex diseases.

Table 1 Single-factor LR models based directly off male and female RP tree split criteria (see figures 1 and 2)

Gender	RP split	Gene	SNP	Genotypes	Full data		RP-subsetted data		
					OR (95% CI)	p Value	Subset	OR (95% CI)	p Value
Male	Primary	ANKK1#	rs1800497	T/T versus C/C + C/T	2.55 (1.44 to 4.51)	0.001*	–	–	–
	Left	GNRH1	novel SNP	T/T versus C/C + T/C	1.43 (1.09 to 1.88)	0.011	M1	1.57 (1.18 to 2.08)	0.002*
Female	Primary	DRD2	rs2242592	C/C + T/C versus T/T	1.32 (1.08 to 1.62)	0.006*	–	–	–
	Left 1	APOC3	rs2854116	T/T versus C/C + T/C	1.28 (1.04 to 1.57)	0.018	F1	1.55 (1.15 to 2.09)	0.004*
	Left 2	ACVR2B	rs3749386	C/C + T/T versus T/C	1.11 (0.91 to 1.36)	0.302	F2	2.17 (1.37 to 3.44)	0.001*
	Right 1	FTO	rs1421085	C/C + T/T versus T/C	1.32 (1.08 to 1.62)	0.007*	F3	1.65 (1.24 to 2.18)	0.0005*
	Right 2	IL6	rs1800795	C/C + G/G versus C/G	1.12 (0.92 to 1.37)	0.269	F4	1.85 (1.19 to 2.89)	0.006*

Each SNP split was first run on the full data set to represent single main factor effects ('full data') for both men and women. Then, the same SNP splits were run on specific subsets of data per RP tree splits (M1, F1–F4; 'RP-subsetted data').
M1: LR analysis was run for only those with genotype DRD2 rs1800497 C/C or C/T.
F1: LR analysis was run for only those with genotype DRD2 rs2242592 T/T.
F2: LR analysis was run for only those with genotypes DRD2 rs2242592 T/T and APOC3 rs2854116 T/T.
F3: LR analysis was run for only those with genotype DRD2 rs2242592 C/C or T/C.
F4: LR analysis was run for only those with genotypes DRD2 rs2242592 C/C or T/C and FTO rs1421085 T/C.
#rs1800497 is historically known as the DRD2 Taq1A allele.
*p<0.009
LR, logistic regression; RP, recursive partitioning; SNP, single nucleotide polymorphism.

The utility of RP and LR for identification of gene–gene interactions

With recent advances in genotyping allowing for high-dimensional SNP identification, it is now possible to examine genetic data sets for single main factor effects and also for G × G interactions. The requirement for G × G analyses as a better predictor of age-related diseases is obvious from the standpoint that humans are complex biological systems composed of numerous molecular interactions and from recent studies indicating disease risk is modulated by G × G interactions.[7] Notwithstanding this, the development of analytical tools for the identification of G × G interactions has not kept pace with the technological advances in identifying genetic alterations among individuals. In this respect, we have previously used multifactor dimensionality reduction (MDR), LR and linkage disequilibrium (LD) to identify G × G interactions among a small set of SNPs.[7] However, large data sets require a screening tool to identify potential multigene interactions. In this study, we have used RP to screen for multigene interactions, a data mining technique that is currently underused in genetic studies. RP serves as an efficient and powerful exploratory analysis technique, especially when looking for interactions in data sets with a large number of independent variables. This screening allows for the identification of G × G interactions (with greater explanatory power) that might otherwise not have been identified and that can then be confirmed using more traditional statistical techniques. As illustrated in this paper, this data mining methodology has the advantage of identification of genetic interactions *between* pathways involved in the aetiology of depression, in keeping with the etiological heterogeneity of this disorder (see later).

Our study provides proof of principle for the use of RP in higher dimensional analyses such as GWAS, where a comprehensive list of SNPs may fully explore genetic predisposition to depression and other age-related disease. The WLS is an ideal candidate for future GWAS studies, given its large sample size, rich covariate composition and longitudinal nature.

In this genetic study, we aimed to identify underlying genetic predispositions to depression and thus have not yet tested environmental, health, socio-behavioral or other non-genetic factors. Future analyses using RP to examine the impact of these factors on the development of depression would be anticipated to identify gene–non-genetic factor interactions. Indeed, the predictive gains of G × G analyses were stronger for men than for women, despite the fact that depression occurs disproportionately in women ($\sim$2:1 female-to-male ratio[21–25]). This suggests that environmental factors may be needed in addition to genetic factors in understanding the aetiological pathways for women. Indeed, biological factors such as hormonal changes related to reproductive status[26 27] may impact environmental factors such as psychosocial experiences (trauma, stress, interpersonal relationships, etc) and general health issues in the development of depression.

Genetic and biological correlates of depression

Numerous studies have identified SNPs that associate with depression. Many of the SNPs associated with depression from other studies were not significantly associated in our study. This is perhaps not surprising since a single factor is unlikely to provide consistent association especially in a complex condition such as depression, where multiple pathways intersect in regulating the risk of the disease. For example, if a SNP within the serotonin pathway also requires a SNP in the glutamatergic pathway in order for the patient to present with depression, the presence of either SNP in the absence of the other will not be predictive of depression. Moreover, as indicated by Shi and Weinberg,[28] since the human genome contains genetic redundancy, disruption of a single gene may be selectively neutral, but the malfunction of several genes in a pathway might result in expression of a particular phenotype.

Both the primary splits in men and women were SNPs linked with *DRD2* (dopamine receptor D2), a gene that has previously been linked with depression and social phobia.[29-31] The primary male genotype split rs1800497, technically found in gene *ANKK1*, is historically known as the *DRD2* Taq1A allele because of its known association with decreased dopamine receptor D2 density (in those with T-alleles).[32-35] The Taq1A allele has also been previously associated with depressive symptoms in children, where those with the A1 allele (T) were more likely to have depressive symptoms.[36] We saw a similar association between A1 and depression in WLS men, where those with two A1 alleles had 2.6 times higher odds for depression compared with those with one or no A1 alleles. The primary split in women (*DRD2* rs2242592) has previously been found to be associated with schizophrenia, where the C-allele was associated with higher susceptibility for the disease.[37] Interestingly, this same study also found the Taq1A allele to also associate with schizophrenia.

The secondary and tertiary right-hand splits in the female RP tree—*FTO* (fat mass and obesity associated) rs1421085 and *IL6* (interleukin 6) rs1800795—have also been found to relate with mental illness and depression in previous studies.[38][39] There is evidence that activin receptor signalling also is involved in affective disorders, especially when considering interaction with GABAergic pathways.[40] Although we did not see an interaction between SNPs in GABA/activin receptor genes and depression, *ACVR2B* was associated with depression in women. No previous associations between depression and *APOC3*, *ACVR2B*, or *GNRH1* have been reported.

That these genetic variants are associated with *neuroendocrine* pathways (*GNRH1*, *ACVR2B*) that are known to regulate *neurotransmitter* release and cognitive behaviour[39][40] supports these associations as relevant to the aetiology of depression and underlines the benefits of using RP to identify meaningful G × G interactions associated with disease.

Limitations

Given the numerous genetic and non-genetic influences that are linked to depression and the small number of SNPs analysed, it is not surprising that predictability from our models was low (although our predictability was superior to previous studies examining only single main factors). Also, the predictive value of our statistical models was further limited due to user bias in selection of SNPs (from nearly 2 million SNPs in the human genome) used in this study. As a result of this, interactions we have found could potentially be moderated by another gene that we have not considered in this study. Nonetheless, we identified significant G × G interactions between known, and newly identified, loci associated with depression. Importantly, four of the seven SNPs identified in these interactions were primarily located in two pathways well known to impact depression: neurotransmitter and neuroendocrine signalling.

The results from the RP analyses conducted in this study were confirmed by LR, demonstrating the utility of RP as a screening tool for identifying meaningful G × G interactions. Future development of algorithms for RP analysis should maximise the distance between branches of the next best split (ie, rpart) and consider subsequent future split combinations that could potentially result in trees with 'better' overall predictability.

Summary

Our data indicate that G × G interaction analyses allow for enhanced predictability of conditions and diseases of ageing. RP is an efficient and powerful exploratory analysis technique for elucidating G × G interactions in large data sets and combined with LR provides an important statistical analysis for the identification of well-supported G × G interactions. We predict that such analytical methods will play an increasingly important role in the identification of epistatic effects in future GWAS. Finally, our studies illustrate how RP analyses can be used to find interacting pathways involved in the aetiology of a disease or condition such as depression.

REFERENCES

1. CDC. *Anxiety and Depression*. Atlanta, GA: Centers for Disease Control and Prevention, 2009. http://www.cdc.gov/Features/dsBRFSSDepressionAnxiety/ (accessed 18 Oct 2010).
2. Eley TC, Sugden K, Corsico A, *et al*. Gene-environment interaction analysis of serotonin system markers with adolescent depression. *Mol Psychiatry* 2004;9:908–15.
3. Mirowsky J, Ross CE. Age and depression. *J Health Soc Behav* 1992;33:187–205.
4. Zubenko GS, Zubenko WN, Spiker DG, *et al*. Malignancy of recurrent, early-onset major depression: a family study. *Am J Med Genet* 2001;105:690–9.
5. Raymer KA, Waters RF, Price CR. Proposed multigenic Composite Inheritance in major depression. *Med Hypotheses* 2005;65:158–72.
6. Harold D, Abraham R, Hollingworth P, *et al*. Genome-wide association study identifies variants at CLU and PICALM associated with Alzheimer's disease. *Nat Genet* 2009;41:1088–93.
7. Haasl R, Ahmadi MR, Meethal S, *et al*. A luteinizing hormone receptor intronic variant is significantly associated with decreased risk of Alzheimer's disease in males carrying an apolipoprotein E epsilon4 allele. *BMC Med Genet* 2008;9:37.
8. Basson J, Wilson A, Haasl R, *et al*. Multi locus interactions in steroidogenic pathway genes predict Alzheimer's disease. *Alzheimer's Dementia* 2008;4(Suppl 2):T579, P3–199.
9. Pezawas L, Meyer-Lindenberg A, Goldman AL, *et al*. Evidence of biologic epistasis between BDNF and SLC6A4 and implications for depression. *Mol Psychiatry* 2008;13:709–16.
10. Zhang K, Yang C, Xu Y, *et al*. Genetic association of the interaction between the BDNF and GSK3B genes and major depressive disorder in a Chinese population. *J Neural Transm* 2010;117:393–401.
11. Lin E, Hong CJ, Hwang JP, *et al*. Gene-gene interactions of the brain-derived neurotrophic-factor and neurotrophic tyrosine kinase receptor 2 genes in geriatric depression. *Rejuvenation Res* 2009;12:387–93.
12. Neff CD, Abkevich V, Potter J, *et al*. Evidence for epistasis between SLC6A4 and a chromosome 4 gene as risk factors in major depression. *Am J Med Genet B Neuropsychiatr Genet* 2010;153B:321–2.
13. Wong ML, Dong C, Andreev V, *et al*. Prediction of susceptibility to major depression by a model of interactions of multiple functional genetic variants and environmental factors. *Mol Psychiatry* 2012;17:624–33.
14. Carr D, Khodyakov D. End-of-life health care planning among young-old adults: an assessment of psychosocial influences. *J Gerontol B Psychol Sci Soc Sci* 2007;62:S135–S41.
15. Sewell WH. *As we age: the Wisconsin longitudinal study, 1957-2001*. Madison, WI: Center for Demography and Ecology University of Wisconsin–Madison, 2001.
16. Nelson CB, Kessler RC, Mroczek D. *Scoring the World health Organization's Composite International Diagnostic Interview short form (CIDI-SF; v1.0 NOV98)*. 2001. http://www.hcp.med.harvard.edu/ncs/ftpdir/cidisf_readme.pdf (accessed 27 Jan 2012).
17. Zhang H, Bonney G. Use of classification trees for association studies. *Genet Epidemiol* 2000;19:323–32.
18. Breiman L, Friedman J, Olshen R, *et al*. *Classification and Regression Trees*. Boca Raton: Chapman and Hall/CRC, 1984.
19. Therneau T, Atkinson E, eds. *Technical Report Series No. 61, An Introduction to Recursive Partitioning Using the RPART Routines*. Rochester, MN: Department of Health Science Research, Mayo Clinic, 1997.
20. Accortt EE, Freeman MP, Allen JJB. Women and major depressive disorder: clinical perspectives on causal pathways. *J Womens Health (Larchmt)* 2008;17:1583–90.
21. Kessler RC, McGonagle KA, Swartz M, *et al*. Sex and depression in the National Comorbidity Survey. I: lifetime prevalence, chronicity and recurrence. *J Affect Disord* 1993;29:85–96.
22. Kessler RC, McGonagle KA, Nelson CB, *et al*. Sex and depression in the National Comorbidity Survey. II: cohort effects. *J Affect Disord* 1994;30:15–26.
23. Kessler RC, McGonagle KA, Zhao S, *et al*. Lifetime and 12-month prevalence of DSM-III-R psychiatric disorders in the United States. Results from the National Comorbidity Survey. *Arch Gen Psychiatry* 1994;51:8–19.
24. Weissman MM, Bland R, Joyce PR, *et al*. Sex differences in rates of depression: cross-national perspectives. *J Affect Disord* 1993;29:77–84.
25. Weissman MM, Olfson M. Depression in women: implications for health care research. *Science* 1995;269:799–801.
26. Pajer K. New strategies in the treatment of depression in women. *J Clin Psychiatry* 1995;56(Suppl 2):30–7.
27. Paykel ES. Depression in women. *Br J Psychiatry Suppl* 1991:22–9. Review.
28. Shi M, Weinberg CR. How much are we missing in SNP-by-SNP analyses of genome-wide association studies? *Epidemiology* 2011;22:845–7.
29. Lawford BR, Young R, Noble EP, *et al*. The D2 dopamine receptor (DRD2) gene is associated with co-morbid depression, anxiety and social dysfunction in untreated veterans with post-traumatic stress disorder. *Eur Psychiatry* 2006;21:180–5.
30. Klimek V, Schenck JE, Han H, *et al*. Dopaminergic abnormalities in amygdaloid nuclei in major depression: a postmortem study. *Biol Psychiatry* 2002;52:740–8.
31. Schneier FR, Liebowitz MR, Abi-Dargham A, *et al*. Low dopamine D (2) receptor binding potential in social phobia. *Am J Psychiatry* 2000;157:457–9.
32. Noble EP, Blum K, Ritchie T, *et al*. Allelic association of the D2 dopamine receptor gene with receptor-binding characteristics in alcoholism or gene ism. *Arch Gen Psychiatry* 1991;48:648–54.
33. Thompson J, Thomas N, Singleton A, *et al*. D2 dopamine receptor gene (DRD2) Taq1 A polymorphism: reduced dopamine D2 receptor binding in the human striatum associated with the A1 allele. *Pharmacogenetics* 1997;7:479–84.
34. Pohjalainen T, Rinne JO, Nagren K, *et al*. The A1 allele of the human D2 dopamine receptor gene predicts low D2 receptor availability in healthy volunteers. *Mol Psychiatry* 1998;3:256–60.
35. Jonsson EG, Nothen MM, Grunhage F, *et al*. Polymorphisms in the dopamine D2 receptor gene and their relationships to striatal dopamine receptor density of healthy volunteers. *Mol Psychiatry* 1999;4:290–6.
36. Hayden EP, Klein DN, Dougherty LR, *et al*. The dopamine D2 receptor gene and depressive and anxious symptoms in childhood: associations and evidence for gene-environment correlation and gene-environment interaction. *Psychiatr Genet* 2010;20:304–10.
37. Dubertret C, Bardel C, Ramoz N, *et al*. A genetic schizophrenia-susceptibility region located between the ANKK1 and DRD2 genes. *Prog Neuropsychopharmacol Biol Psychiatry* 2010;34:492–9.
38. Kivimaki M, Jokela M, Hamer M, *et al*. Examining overweight and obesity as risk factors for common mental disorders using fat mass and obesity-associated (FTO) Genotype-Instrumented Analysis. *Am J Epidemiol* 2011;173:421–9.
39. Bull SJ, Huezo-Diaz P, Binder EB, *et al*. Functional polymorphisms in the interleukin-6 and serotonin transporter genes, and depression and fatigue induced by interferon-alpha and ribavirin treatment. *Mol Psychiatry* 2009;14:1095–104.
40. Zheng F, Adelsberger H, Muller MR, *et al*. Alzheimer C. Activin tunes GABAergic neurotransmission and modulates anxiety-like behavior. *Mol Psychiatry* 2008;14:332–46.