Open

ORIGINAL ARTICLE

Prediction of susceptibility to major depression by a model of interactions of multiple functional genetic variants and environmental factors

M-L Wong[1], C Dong[2], V Andreev[3], M Arcos-Burgos[1] and J Licinio[1]

[1]Department of Translational Medicine, John Curtin School of Medical Research, Australian National University, Canberra, ACT, Australia; [2]Department of Neurology, University of Miami Miller School of Medicine, Miami, FL, USA and [3]Department of Psychiatry, University of Miami Miller School of Medicine, Miami, FL, USA

Major depressive disorder (MDD) is the most common psychiatric disorder and the second overall cause of disability. Even though a significant amount of the variance in the MDD phenotype is explained by inheritance, specific genetic variants conferring susceptibility to MDD explain only a minimal proportion of MDD causality. Moreover, genome-wide association studies have only identified two small-sized effect loci that reach genome-wide significance. In this study, a group of Mexican-American patients with MDD and controls recruited for a pharmacogenetic study were genotyped for nonsynonymous single-nucleotide polymorphisms (nsSNPs) and used to explore the interactions of multiple functional genetic variants with risk-classification tree analysis. The risk-classification tree analysis model and linkage disequilibrium blocks were used to replicate exploratory findings in the database of genotypes and phenotypes (dbGaP) for major depression, and pathway analysis was performed to explore potential biological mechanisms using the branching events. In exploratory analyses, we found that risk-classification tree analysis, using 15 nsSNPs that had a nominal association with MDD diagnosis, identified multiple increased-MDD genotype clusters and significant additive interactions in combinations of genotype variants that were significantly associated with MDD. The results in the dbGaP for major depression disclosed a multidimensional dependent phenotype constituted of MDD plus significant modifiers (smoking, marriage status, age, alcohol abuse/dependence and gender), which then was used for the association tree analysis. The reconstructed tree analysis for the dbGaP data showed robust reliability and replicated most of the genes involved in the branching process found in our exploratory analyses. Pathway analysis using all six major events of branching (*PSMD9*, *HSD3B1*, *BDNF*, *GHRHR*, *PDE6C* and *PDLIM5*) was significant for positive regulation of cellular and biological processes that are relevant to growth and organ development. Our findings not only provide important insights into the biological pathways underlying innate susceptibility to MDD but also offer a predictive framework based on interactions of multiple functional genetic variants and environmental factors. These findings identify novel targets for therapeutics and for translation into preventive, clinical and personalized health care.
Molecular Psychiatry (2012) **17**, 624–633; doi:10.1038/mp.2012.13; published online 27 March 2012

Introduction

Most genetic diseases, Mendelian or not, are now known to be complex;[1] the broad strategy used to identify Mendelian disease genes, which requires that each susceptibility factor have a large independent effect on disease risk, has not been successful for common and complex disorders.[2] In spite of the recent application of genome-wide association studies,[3] which have identified hundreds of genetic variants associated with chronic illnesses, the search for and interpretation of these genetic associations continue to be a daunting task. In many cases, associated genetic variants account for a very small component of the phenotypic variance ($\sim 3\%$), despite high estimates of heritability (~ 0.80).[4] Because of this disappointingly small power for phenotypic prediction, several authors have suggested that genome-wide association study is more informative about the novel biological pathways rather than for the clinical diagnosis and use in predictive medicine.[5,6]

It has been proposed that multifactorial disorders fit the 'common disease–common alleles model',[2]

in which the cumulative linear impact of multiple common small-effect genetic variants (gene–gene interactions) in the context of environmental exposures to increase the individual risk, exceed a biological threshold and produce an altered phenotype (genetic–environment interactions).[3] It is also reasonable to assume that larger risks will emerge from the nonlinear interaction of rare genetic variant with major effect. Nonsynonymous single-nucleotide polymorphisms (nsSNPs) are a main type of 'functional' SNPs, as they result in true amino-acid variation, and consequently may have a major impact on phenotype.

Major depressive disorder (MDD) is a complex illness characterized by a triad of symptoms: low or depressed mood, loss of interest or pleasure in almost all activities (anhedonia) and low energy/fatigability;[7] MDD represents the second cause of disability.[7,8] In the United States, suicide, mostly a consequence of depression, is the eleventh overall cause of death; the third cause of death in the age group of 15–24 years and the fourth cause of death in the age group of 25–44 years.[9] The economic burden of depression to the US economy is $100 billion annually.[10] MDD heritability is estimated at 0.36–0.7 based on twin studies,[11] that is, a significant genetic component underlies MDD susceptibility. Several candidate-gene approaches disclosed intriguing association of MDD to genetic variants, whereas genome-wide association studies only reported two loci that reached genome-wide significance.[12,13]

This study was designed to investigate relationships/gene–gene interactions defined by nsSNPs in the pathways relevant to central nervous system function in case-control samples to find recurrent functional process themes in MDD.

Materials and methods

Subjects

Mexican-American sample. This sample of cases and controls consists of 321 healthy controls and 278 MDD aged 19 years or older (Supplementary Table 1). All participants were Mexican-Americans and had at least three grandparents born in Mexico. MDD was defined using the DSM-IV (Diagnostic and Statistical Manual of Mental Disorders, 4th edn) diagnosis of current, unipolar major depressive episode as assessed by the structured clinical interview for DSM-IV and a HAM-D21 (21-Item Hamilton Depression Rating Scale) score of ⩾18 with item number 1 (depressed mood) rated ⩾2. MDD patients were recruited in a pharmacogenetic study approved by the Institutional Review Boards of the University of California Los Angeles, University of Miami and Australian National University, and it has been registered in ClinicalTrials.gov (NCT00265292) as previously described.[14–16] Briefly, in their primary language, all MDD patients had a comprehensive psychiatric and medical assessment based on the diagnostic and ratings instruments that had been fully validated in English and Spanish.

Exclusion criteria included active medical illnesses that could be etiologically related to the ongoing depressive episode, current or active suicidal ideation with a plan and strong intent, pregnancy, lactation, current use of medications with significant central nervous system activity, which interfere with electroencephalographic activity (for example, benzodiazepines) or any other antidepressant treatment within the 2 weeks prior to enrollment, illicit drug use and/ or alcohol abuse in the last 3 months or current enrollment in psychotherapy.[14–16] Control individuals for our genomic studies were in general good health but were not screened for medical or psychiatric illness, they were age- and gender-matched and recruited from the same Mexican-American community in Los Angeles by the same bilingual clinical research team.[14–16]

Database of genotypes and phenotypes (dbGaP) sample data. We used genome-wide association study genotype information from the major depression dbGaP[17] for replication purposes. Briefly, dbGaP provides open access to large samples of genetic and phenotypic datasets (this particular dataset included 1862 participants affected with major depression and 1857 unaffected controls).[17]

Genomic DNA collection and genotyping

At the initial visit, blood samples were collected under informed consent from the participating individuals into EDTA (K2EDTA) BD Vacutainer EDTA tubes (Becton Dickinson, Franklin Lakes, NJ, USA), and genomic DNA was isolated from those samples using Gentra Puregene DNA purification kits (Gentra Systems, Indianapolis, IN, USA). A total of 372 autosomal nsSNPs were selected in 188 genes from dbSNP (build 121). SNP assays were designed and typed with the Golden Gate assay as part of a 1536 multiplex reaction (59). DNAs with poor results (50% GC score < 0.65) as well as loci with a low clustering score (< 0.3) were removed. The threshold for retaining individual genotype calls was set to a score of 0.25. Data quality was assessed using duplicate DNAs across all plates. Genotypes from no matching or missing duplicates were dropped.

Data analysis

Standard genetic association analysis. The Mexican-American sample was used as a discovery-exploratory sample. Quality control and the standard genetic filters were applied following the STREGA statement checklist.[18] Testing for Hardy–Weinberg equilibrium, a more appropriate approach when one allele is very rare, was performed as implemented in PLINK (http://pngu.mgh.harvard.edu/~purcell/plink/). Deviation from Hardy–Weinberg equilibrium was tested separately for control and depressed groups. Nonsynonymous SNPs were excluded from the data analysis if they met the following criteria: SNP assays were successfully genotyped on $< 90\%$ of samples, failed an exact test of Hardy–Weinberg equilibrium in

controls ($p < 0.01$) or were monoallelic in the whole sample. To control and identify the presence of hidden case-control genetic stratification, a major cause of discovering spurious associations, we applied rigorous strategies to discriminate genetic subdivision as described elsewhere.[14,19]

Student's t-test was conducted to compare age and baseline HAM-D21 score means, and Pearson's χ² test was performed to compare gender ratios between the subgroups using SAS version 9.1 (SAS Institute, Cary, NC, USA). Genotypic association with disease status was investigated based on two genetic models: dominant (homozygous for major allele versus the remaining genotypes) and recessive (homozygous for minor allele versus the remaining genotypes). Either the Pearson's χ² test or the Fisher's exact test, whenever any cell count was < 5, was applied using PLINK. Odds ratio (OR) and its 95% confidence interval were calculated using either the Woolf's method or the exact logistic regression model as implemented in SAS to appropriately handle frequency tables that contain cells with zero counts.[20]

Gene–gene interaction analyses. Additive interaction between two genetic variants was measured based on three measures: relative excess risk due to interaction ($RERI = OR_{11} - OR_{10} - OR_{01} + 1$), attributable proportion due to interaction ($AP = RERI/OR_{11}$) and Rothman's synergy index ($S = (OR_{11} - 1)/(OR_{10} + OR_{01} - 2)$), where subscript 0 denotes the absence of the risk genotype at the SNP and OR denotes the odds ratio. No interaction corresponds to $RERI = 0$, $AP = 0$ and $S = 1$, whereas $RERI > 0$, $AP > 0$ and $S > 1$ indicate an increase in the effect due to the additive interaction.[21,22]

Linkage disequilibrium analysis and SNP tagging for dbGaP. HapMap databases (genotypes and phased haplotypes, phase III, release 2) were used to identify linkage disequilibrium blocks harboring SNPs significantly associated to depression risk in the Mexican-American sample. We then selected the dbGaP SNPs in the identified linkage disequilibrium blocks and performed the same comprehensive and extensive array of genetic association tests applied to the Mexican-American sample.

Advance recursive partition (tree-based) approach (ARPA). We applied tree-based approaches because they are currently the most used in predictive analyses as they account for non-linear effects, fast solution for hidden complex substructure (gene–gene and gene–demography interactions) and release of truly non-biased, statistically significant analysis of highly dimensional, seemingly unrelated data. Furthermore, results supplied by tree-based analytics are easier to be visually and logically interpreted. Thus, we employed ARPA to test high-order gene–gene and gene–covariates (demographic data) interactions as it is implemented in HelixTree. Briefly, the basic process in recursive partitioning analysis is to divide a data set into parts where the individuals in each segment share

common features. The dependent variable was the response MDD affection status, and genetic as well as demographic variables were used as predictors. Trees were interactively explored with the following restrictive criteria: (i) at least five elements per child to avoid small, possibly outlier groups, a common artifact consequence of recursive partitioning; (ii) an exact $O(n^2)$ algorithm for segmenting; (iii) a maximum cardinality of five for the multi-way split; (iv) a resampling approach with 5000 iterations to define the smallest bound for the P-value; and (v) split-stopping criteria defined as P-value $\leqslant 0.05$ (corrected by Bonferroni).

Prediction of functional effect of human nsSNPs. We used the freely available Polymorphism phenotyping (PolyPhen-2) tool for predicting whether a specific nsSNP may be damaging[23] (http://genetics.bwh.harvard.edu/pph2/).

Pathway and network analyses. In order to evaluate potential common ontogenetic as well as cellular process functionality of those genes disclosed by the ARPA analysis, we used Metacore 6.8 software build 29806 (GeneGo, St Joseph, MI, USA) tools to build networks using the list of six genes that had significant events of branching process.

Results

Discovery and exploratory phase with the Mexican-American data

After quality control filters were applied and Hardy–Weinberg equilibrium tested, a total of 252 nsSNPs in 155 genes were available for analytical procedures (Supplementary Table 2). It is relevant to mention that SNP frequency may vary widely among different ethnic populations. We studied Mexican-American individuals (Supplementary Table 3 shows the influence of ethnicity in significant nsSNPs); consequently, those frequencies may influence the risk-assessment tree structure. For this reason, we replicated this work in the dbGap sample.

In all, 15 nsSNPs showed significant genotypic association with the diagnosis of MDD (Table 1); 11 and 6 of those nsSNPs fit better a dominant and a recessive model, respectively.

ARPA analyses performed to assess MDD risk in clusters of variants showed that eight nsSNPs had high-order interactions in the control of tree growth (Figure 1). Our tree model provided 40% sensitivity to the diagnosis of MDD, 83% specificity for prediction of controls and 63% accuracy for the corrected prediction of MDD or control subjects. Significant additive interactions between two genetic variants were found for nine genotype combinations (Table 2).

Replication phase with the dbGaP sample

A preliminary univariate phenotype analysis reported significant associations to markers harbored at markers tagging *HSD3B1*, *PDLIM5*, *INMT*, *ADRA1A* and *PSMD9* (Table 3). By controlling effects of

Table 1 nsSNPs showing significant genotypic association with depression

Model	Gene (location)	SNP	Genotype		Case	Control	OR (95% CI)	P[a]	PolyPhen-2 score, prediction
Dominant	HSD3B1 (1p13.1)	rs6205	F286L	CC/CT	49	30	2.10 (1.29, 3.41)	0.002	Score 0.000
				TT	226	290	Reference		Benign
	PSMB4 (1q21)	rs4603	I234T	CC/CT	125	173	0.70 (0.51, 0.97)	0.03	Score 0.971
				TT	151	146	Reference		Probably damaging
	ALDH9A1 (1q23.1)	rs1065756	S221T	CC/CG	34	24	1.73 (1.00, 3.00)	0.05	Score 0.000
				GG	240	293	Reference		Benign
	PDLIM5 (4q22)	rs7690296	T272A	GG/GA	206	202	1.65 (1.16, 2.35)	0.005	Score 0.000
				AA	71	115	Reference		Benign
	NFKB1 (4q24)	rs4648072	M507V	GG/GA	1	8	0.14 (0.02, 1.17)	0.04	Score 0.0069
				AA	276	312	Reference		Benign
	INMT (7p15)	rs4723010	D28N	AA/AG	126	181	0.64 (0.46, 0.89)	0.007	Score 0.002
				GG	149	137	Reference		Benign
	GHRHR (7p14)	rs4988496	A57T	AA/AG	14	33	0.47 (0.24, 0.89)	0.02	Score 0.000
				GG	262	288	Reference		Benign
	ADRA1A (8p21-p11)	rs2229125	I200S	GG/GT	38	27	1.73 (1.02, 2.91)	0.04	Score 1.000
				TT	239	293	Reference		Probably damaging
	BDNF (11p13)	rs6265	V66M	AA/AG	57	98	0.58 (0.40, 0.85)	0.005	Score 0.604
				GG	220	221	Reference		Possibly damaging
	ACE (17q23.3)	rs4318	S49G	GG/GA	1	10	0.11 (0.01, 0.89)	0.01	Score 0.013
				AA	274	310	Reference		Benign
	USP36 (17q25.3)	rs3744793	V271I	AA/AG	195	197	1.55 (1.10, 2.19)	0.01	Score 0.009
				GG	78	122	Reference		Benign
	COMT (22q11.21)	rs6267	A72S	TT/TG	13	32	0.48 (0.25, 0.94)	0.03	Score 0.215
				GG	235	279	Reference		Benign
Recessive	MYOC (1q23-q24)	rs2234926	R76K	AA	7	1	8.36 (1.02, 68.41)	0.03	Score 0.988
				AG/GG	267	319	Reference		Probably damaging
	INMT (7p15)	rs2302339	M205V	GG	6	22	0.32 (0.13, 0.81)	0.01	Score 1.000
				GA/AA	251	296	Reference		Probably damaging
	GHRHR (7p14)	rs2228078	M422T	CC	22	11	2.40 (1.14, 5.05)	0.02	Score 0.002
				CT/TT	254	305	Reference		Benign
	NAT2 (8p22)	rs1208	R268K	GG	29	58	0.53 (0.33, 0.86)	0.01	Score 0.062
				GA/AA	243	260	Reference		Benign
	PDE6C (10q24)	rs701865	S270T	AA	58	41	1.84 (1.19, 2.85)	0.006	Score 0.000
				AT/TT	212	276	Reference		Benign
	PSMD9 (12q24)	rs1043307	E197G	GG	65	47	1.76 (1.16, 2.67)	0.007	Score 0.174
				GA/AA	211	269	Reference		Benign

Abbreviations: CI, confidence interval; OR, odds ratio; nsSNPs, nonsynonymous single-nucleotide polymorphisms; SNP, single-nucleotide polymorphism.
[a]Based on Fisher's exact test if cell count < 5.

significantly associated demographic factors, such as age, sex and comorbidities (substance abuse/dependence (nicotine and alcohol), among other variables), all of them available at dbGaP (Supplementary Table 4), we created a multidimensional MDD plus modifiers phenotype that explained a significant phenotype variance associated to the MDD phenotype. Using the multidimensional MDD plus modifiers phenotype, ARPA reconstructed a tree extraordinarily resembling the tree generated during the exploratory phase with the Mexican-American MDD sample. This dbGaP-based tree replicated most of the genes involved in the branching process; that is, the first five very significant events of branching are provided by: PDE6C, PSMD9, HSD3B1, BDNF, GHRHR and PDLIM5 (Figure 2).

Pathway analyses
We used all six genes generated by significant events of branching provided by our dbGaP analyses, that is, PDE6C, PSMD9, HSD3B1, BDNF, GHRHR and PDLIM5, to build networks and perform the pathway analyses (Figure 3). These genes form a network with 156 interactions. Significant biological processes
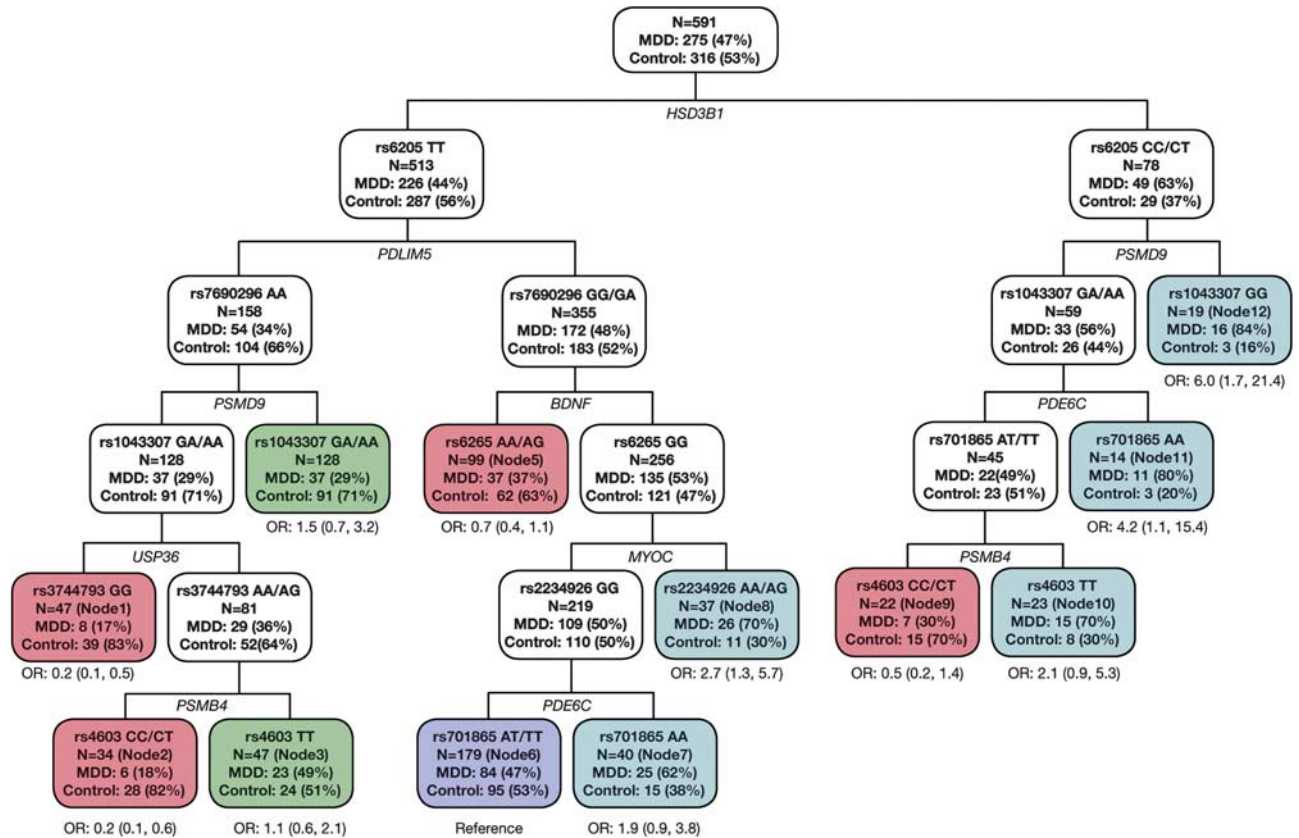
Figure 1 Tree analyses performed to assess MDD risk in cluster of variants. Eight nsSNPs had high-order interactions in the control of tree growth; they are located in the following genes: *HSD3B1*, *PDLIM5*, *PSMD9*, *BDNF*, *USP36*, *PDE6C*, *PSMB4* and *MYOC* genes. Light-pink nodes = low MDD risk; light-green nodes = moderate MDD risk; light-blue nodes = high MDD risk; and purple node = reference MDD risk. Sensitivity = 40%, specificity = 83% and accuracy = 63% (sensitivity refers to the proportion of cases who are correctly predicted as case; specificity refers to the proportion of controls who are correctly predicted as control; and accuracy refers to the proportion of cases and controls who are correctly predicted as case or control).

involved in this network include positive regulation of cellular and biological processes, nerve growth-factor receptor signaling pathway, and regulation of cell death (Supplementary Table 5).

Prediction of functional effect of human nsSNPs
Most variations are predicted to be benign (Table 1). Variations rs4603 (*PSMB4*), rs2229125 (*ADRA1A*), rs2234926 (*MYOC*) and rs2302339 (*INMT*) are predicted to be probably damaging and variation rs6265 (*BDNF*) is predicted to be possibly damaging.

Discussion

We showed that genetic analyses as well as classificatory multidimensional tree techniques applied to nsSNPs associated to MDD diagnosis provided a reliable branching-tree framework that predicted, with reasonable sensibility and specificity, the clustering of MDD patients.

The allelic variation rs6265 in the *BDNF* gene has been replicated in at least three independent studies in MDD;[24] this variation is predicted to be possibly damaging and it has been also implicated in a number of neuropsychiatric conditions (OMIM 113505). Recently, allelic variations in the *PDLIM5* gene have been implicated in recurrent MDD,[25] bipolar[26] and schizophrenia.[27] We have reported data supporting the role of cGMP phosphodiesterases (*PDE*) genes[28] and the role of immune-related genes relevant to T-cell function (*TBX21* and *PSMB4*) in MDD.[29]

In our discovery/exploratory analyses, the ubiquitin-proteasome pathway provided a recurrent theme in the tree structure (Figure 1). Three nsSNPs with high-order interactions in the control of the tree growth in the *PSMB4*, *PSMD9* and *USP36* genes belong to this pathway, however, only *PSMD9* remained in the replication tree (Figure 2). *PSMD9* is a subunit of the 26S proteasome, which consists of a 20S proteolytic core capped at both ends by the 19S regulatory complex that recognizes the polyubiquitin-tagged substrates. Studies have linked the 26S proteasome with transcriptional activities through proteolysis of steroid-hormone receptors to limit their transcriptional output and recycling of transcriptional complexes on chromatin to facilitate multiple rounds of transcriptional initiation.[30] The ubiquitin-proteasome pathway is the major system of selective degradation

Table 2 nsSNPs showing synergic effects on the risk of depression

Genotype		Case	Control	OR (95% CI)	P	Measures of additive interaction		
SNP1 (gene 1)	SNP2 (gene 2)					RERI	AP	S
rs6205 *(HSD3B1)*	rs1043307 *(PSMD9)*							
TT	GA/AA	177	243	Reference				
	GG	49	44	1.53 (0.97, 2.40)	0.06			
CC/CT	GA/AA	33	26	1.74 (1.01, 3.02)	0.05			
	GG	16	3	7.32 (2.10, 15.04)	0.0003	5.05	0.69	4.97
rs7690296 *(PDLIM5)*	rs6265 *(BDNF)*							
AA	AA/AG	11	32	Reference				
	GG	60	83	2.10 (0.98, 4.50)	0.05			
GG/GA	AA/AG	46	66	2.03 (0.93, 4.43)	0.07			
	GG	158	135	3.40 (1.65, 6.99)	0.0005	0.27	0.08	1.13
rs7690296 *(PDLIM5)*[a]	rs6265 *(BDNF)*[a]							
AA	AA/AG	10	27	Reference				
	GG	44	77	1.54 (0.68, 3.48)	0.29			
GG/GA	AA/AG	37	62	1.61 (0.70, 3.70)	0.26			
	GG	135	121	3.01 (1.40, 6.47)	0.003	0.86	0.28	1.74
rs7690296 *(PDLIM5)*	rs2234926 *(MYOC)*							
AA	GG	55	90	Reference				
	AA/AG	16	25	1.05 (0.51, 2.13)	0.90			
GG/GA	GG	164	169	1.59 (1.07, 2.37)	0.02			
	AA/AG	40	32	2.05 (1.15, 3.55)	0.01	0.41	0.20	1.65
rs7690296 *(PDLIM5)*[a]	rs2234926 *(MYOC)*[a]							
AA	GG	41	82	Reference				
	AA/AG	13	22	1.18 (0.54, 2.58)	0.68			
GG/GA	GG	137	155	1.77 (1.14, 2.74)	0.01			
	AA/AG	35	28	2.50 (1.34, 4.56)	0.004	0.55	0.22	1.58
rs6265 *(BDNF)*	rs2234926 *(MYOC)*							
AA/AG	GG	45	71	Reference				
	AA/AG	12	27	0.70 (0.32, 1.52)	0.37			
GG	GG	174	188	1.46 (0.95, 2.24)	0.08			
	AA/AG	44	30	2.31 (1.28, 4.05)	0.005	1.15	0.50	8.14
rs6265 *(BDNF)*[b]	rs2234926 *(MYOC)*[b]							
AA/AG	GG	28	45	Reference				
	AA/AG	9	17	0.85 (0.33, 2.17)	0.73			
GG	GG	109	110	1.59 (0.93, 2.74)	0.09			
	AA/AG	26	11	3.80 (1.63, 7.80)	0.002	2.36	0.62	6.31
rs3744793 *(USP36)*	rs4603 *(PSMB4)*							
GG	CC/CT	38	67	Reference				
	TT	41	53	1.36 (0.77, 2.41)	0.29			
AA/AG	CC/CT	87	105	1.46 (0.90, 2.38)	0.13			
	TT	109	91	2.11 (1.30, 3.41)	0.002	0.29	0.14	1.35
rs3744793 *(USP36)*[c]	rs4603 *(PSMB4)*[c]							
GG	CC/CT	5	19	Reference				
	TT	3	20	0.57 (0.12, 2.72)	0.48			
AA/AG	CC/CT	6	28	0.81 (0.22, 3.05)	0.76			
	TT	23	24	3.64 (1.17, 11.41)	0.02	3.26	0.89	−4.29

Abbreviations: AP, attributable proportion due to interaction; CI, confidence interval; nsSNPs, nonsynonymous single-nucleotide polymorphisms; OR, odds ratio; RERI, relative excess risk due to interaction; S: synergy index; SNP, single-nucleotide polymorphism.
[a]rs6205 = TT.
[b]rs6205 = TT and rs7690296 = GG/GA.
[c]rs6205 = TT, rs7690296 = AA and rs1043307 = GA/AA.

Table 3 Replication results with the dbGap sample

Chromosome	SNP and tagged gene	Model	Case	Control	Chi-square (χ^2)	df	P
1	rs6428830 (*HSD3B1*)	Allelic	1003/2579	1109/2503	6.332	1	0.012
		Dominant	866/925	941/865	5.062	1	0.024
		Genotypic	137/729/925	168/773/865	6.388	2	0.041
	rs7553527 (*HSD3B1*)	Allelic	1577/1947	1519/2061	3.888	1	0.049
4	rs12649976 (*PDLIM5*)	Recessive	310/1476	265/1539	4.748	1	0.029
	rs13135257 (*PDLIM5*)	Recessive	307/1469	263/1530	4.556	1	0.033
	rs17021917 (*PDLIM5*)	Dominant	44/1745	26/1781	4.906	1	0.027
		Allelic	44/3534	26/3588	4.858	1	0.028
7	rs4988498 (*INMT*)	Recessive	3/1750	12/1756	5.346	1	0.021
8	rs505138 (*ADRA1A*)	Genotypic	15/285/1483	7/358/1432	12.03	2	0.002
		Dominant	300/1483	365/1432	7.191	1	0.007
		Allelic	315/3251	372/3222	4.75	1	0.029
12	rs1168658 (*PSMD9*)	Trend	997/2565	934/2666	3.894	1	0.048
	rs7313252 (*PSMD9*)	Allelic	1720/1774	1638/1892	5.616	1	0.018
		Dominant	1300/447	1256/509	4.686	1	0.030

Abbreviation: SNP, single-nucleotide polymorphism.
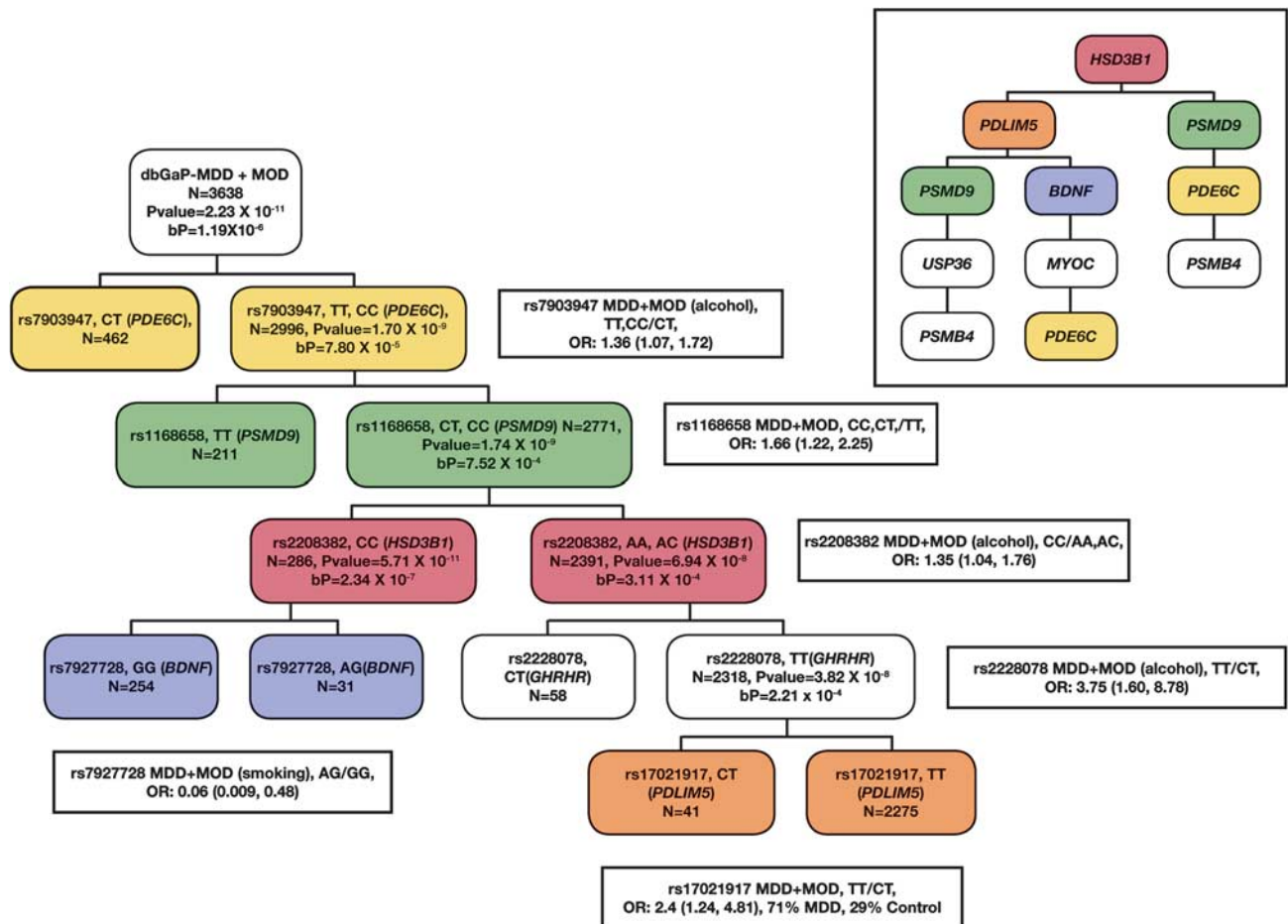
Figure 2 ARPA using dbGaP data reconstructed a tree resembling the tree generated during the exploratory phase with the Mexican-American MDD sample in Figure 1. Most of the genes involved in the branching process, that is, *PDE6C*, *PSMD9*, *HSD3B1*, *BDNF* and *PDLIM5*, were replicated. This tree has one gene (*GHRHR*) not present in the exploratory tree. Reconstruction of the tree highlighted that some covariates used by the multidimensional phenotype, that is, nicotine and alcohol use, provided a significant weight to the process of discriminating MDD-affected individuals from unaffected during the branching. For easier understanding, Figure 1 tree is reproduced in the right upper corner inset and genes in the two trees are color-coordinated.
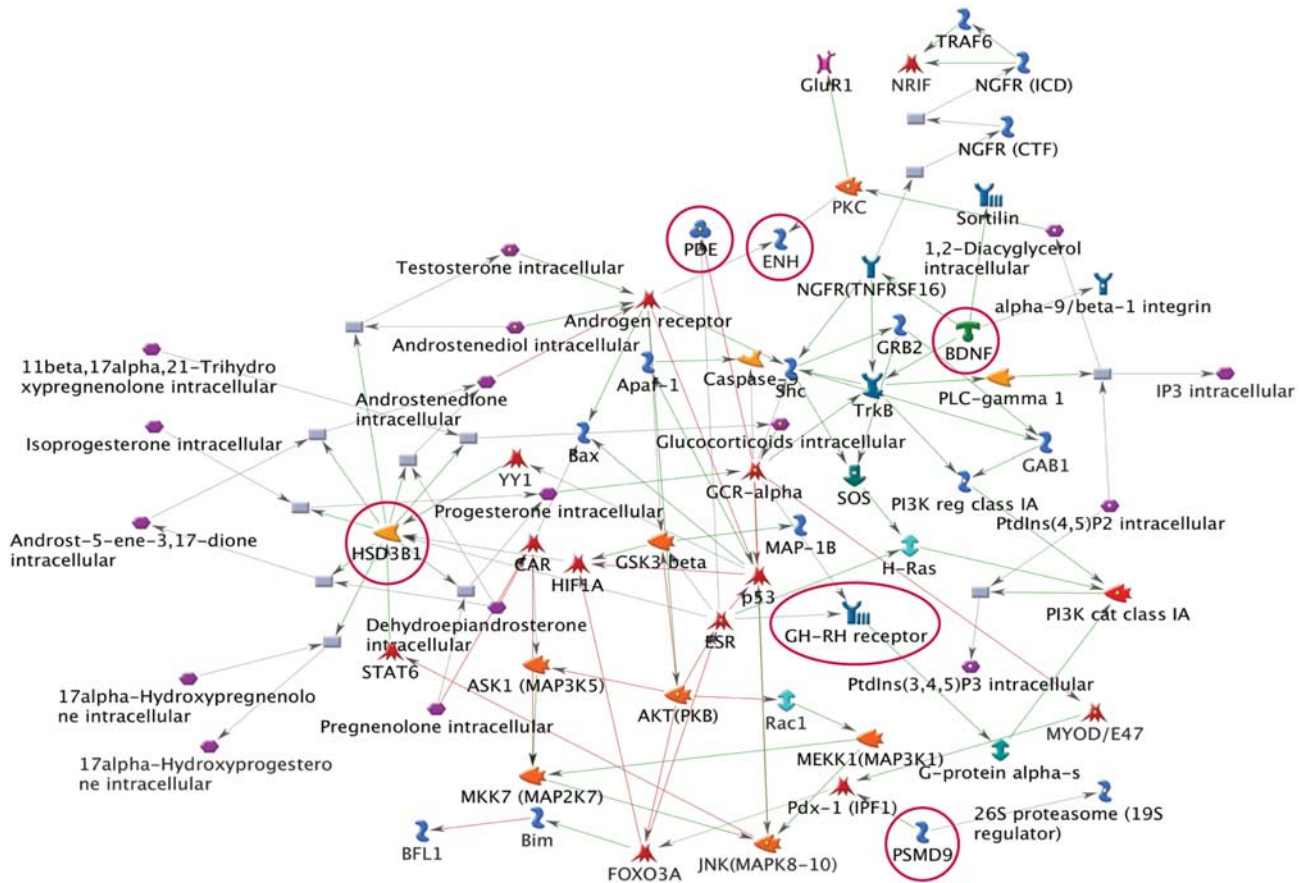
Figure 3 Network analyses showed that all the genes involved in the major branching events in Figure 2 are related to positive regulation of cellular, biological processes and relevant to growth and organ development (the nerve growth-factor signaling pathway and retinal cone-cell development, also see Supplementary Table 5). All the symbols are explained in the MetaCore legend (Supplementary Material, Supplementary Figure 1).

of short-lived regulatory proteins in eukaryotic cells, but it also has an important roles in cell-cycle regulation, signal transduction, differentiation, antigen processing, immune function and degradation of tumor suppressors.[30]

In Figure 1, the variation rs6205 in the *HSD3B1* gene, which defined the root (topmost) node in our risk-assessment tree, is located in a putative membrane-binding domain.[31] The genotype TT in rs6205 (*HSD3B1)* is prevalent in the population, whereas the CC/CT variants of rs6205 (*HSD3B1)* conferred a small increment in MDD risk, which in combination with the *PSMD9* genotype GG (in rs1043306) defined the highest MDD percentage-increased risk (84%, node 12). This cluster of genotypes represents an increase of six-fold in MDD risk compared with our reference node 6 (47%). Clusters of genotypes can confer low risk (nodes 1, 2, 5 and 9) or increased risk for MDD (3 and 4 are moderate MDD risk and nodes 7, 8, 10, 11 and 12 are high MDD risk, OR ⩾ 1.9). Three subtrees emerged in our main tree structure: two outer (right and left) and one middle subtrees. Outer subtrees contained several representations of the ubiquitin-proteasome pathway components (*PSMD9*, *USP36*

and *PSMB4* in the left subtree, and *PSMD9* and *PSMB4* in the right subtree) and the middle subtree contained the *BDNF* rs6265 variation. Of note, is the presence of genes relevant to visual function (*MYOC* and *PDE6C*) in the subtrees that contain high-risk MDD nodes (Figure 1 middle and right subtrees) and their absence in the left subtree, which carries low- and moderate-risk MDD nodes. Variation rs2234926 (*MYOC*) is predicted to be possibly damaging; variations in the *MYOC* gene have been associated with glaucoma. *PDE6C* is a phosphodiesterase enzyme abundantly found in the eye. In our replication tree, only *PDE6C* remained (Figure 2).

Our replication study used an MDD-independent sample (dbGaP data) and tagged markers in linkage disequilibrium to the nsSNPs (Figure 2). It showed an extraordinary reliability for most genes involved in the branching-tree process depicted by the exploratory analysis. It also revealed that comorbidities, such as, nicotine, alcohol abuse, among others, included as covariate information might provide a better prediction of the status outcome. Additionally, network analyses showed that genes involved in those major branching events are related to positive regulation of

cellular, biological processes and relevant to growth and organ development.

Our study provides an initial functional pathway map which helps elucidate the pathophysiology of mild to moderate forms of MDD. This approach has the potential to enhance our understanding of the contributions of gene–gene variations to MDD risk. We have not conducted an extensive nsSNP survey; therefore, other pathways may also contribute to MDD risk. Future work should further assess the interactions of nsSNPs and environmental factors in major depression in diverse population groups.

Conflict of interest

The authors declare no conflict of interest.

References

1 Badano JL, Katsanis N. Beyond Mendel: an evolving view of human genetic disease transmission. *Nat Rev Genet* 2002; **3**: 779–789.

2 Altmuller J, Palmer LJ, Fischer G, Scherb H, Wjst M. Genomewide scans of complex human diseases: true linkage is hard to find. *Am J Hum Genet* 2001; **69**: 936–950.

3 Couzin J, Kaiser J. Genome-wide association. Closing the net on common disease genes. *Science* 2007; **316**: 820–822.

4 Kuntsi J, Neale BM, Chen W, Faraone SV, Asherson P. The IMAGE project: methodological issues for the molecular genetic analysis of ADHD. *Behav Brain Funct* 2006; **2**: 27.

5 Maher BS, Riley BP, Kendler KS. Psychiatric genetics gets a boost. *Nat Genet* 2008; **40**: 1042–1044.

6 Hirschhorn JN, Gajdos ZK. Genome-wide association studies: results from the first few years and potential implications for clinical medicine. *Annu Rev Med* 2011; **62**: 11–24.

7 Licinio J, Wong ML. Advances in depression research: 2011. *Mol Psychiatry* 2011; **16**: 686–687.

8 Licinio J, Wong ML. Why does the United States need a national center for new cures? *Mol Psychiatry* 2011; **16**: 882–884.

9 Milane MS, Suchard MA, Wong ML, Licinio J. Modeling of the temporal patterns of fluoxetine prescriptions and suicide rates in the United States. *PLoS Med* 2006; **3**: e190.

10 Greenberg PE, Kessler RC, Birnbaum HG, Leong SA, Lowe SW, Berglund PA *et al.* The economic burden of depression in the United States: how did it change between 1990 and 2000? *J Clin Psychiatry* 2003; **64**: 1465–1475.

11 Kendler KS, Gardner CO, Neale MC, Prescott CA. Genetic risk factors for major depression in men and women: similar or different heritabilities and same or partly distinct genes? *Psychol Med* 2001; **31**: 605–616.

12 McMahon FJ, Akula N, Schulze TG, Muglia P, Tozzi F, Detera-Wadleigh SD *et al.* Meta-analysis of genome-wide association data identifies a risk locus for major mood disorders on 3p21.1. *Nat Genet* 2010; **42**: 128–131.

13 Hek K, Mulder CL, Luijendijk HJ, van Duijn CM, Hofman A, Uitterlinden AG *et al.* The PCLO gene and depressive disorders: replication in a population-based study. *Hum Mol Genet* 2010; **19**: 731–734.

14 Dong C, Wong ML, Licinio J. Sequence variations of ABCB1, SLC6A2, SLC6A3, SLC6A4, CREB1, CRHR1 and NTRK2: association with major depression and antidepressant response in Mexican-Americans. *Mol Psychiatry* 2009; **14**: 1105–1118.

15 Luo HR, Wu GS, Dong C, Arcos-Burgos M, Ribeiro L, Licinio J *et al.* Association of PDE11A global haplotype with major depression and antidepressant drug response. *Neuropsychiatr Dis Treat* 2009; **5**: 163–170.

16 Licinio J, Dong C, Wong ML. Novel sequence variations in the brain-derived neurotrophic factor gene and association with major depression and antidepressant treatment response. *Arch Gen Psychiatry* 2009; **66**: 488–497.

17 Mailman MD, Feolo M, Jin Y, Kimura M, Tryka K, Bagoutdinov R *et al.* The NCBI dbGaP database of genotypes and phenotypes. *Nat Genet* 2007; **39**: 1181–1186.

18 Little J, Higgins JPT, Ioannidis JPA, Moher D, Gagnon F, von Elm E *et al.* STrengthening the REporting of Genetic Association Studies (STREGA)–an extension of the STROBE statement. *Genet epidemiol* 2009; **33**: 581–598.

19 Wu GS, Luo HR, Dong C, Mastronardi C, Licinio J, Wong ML. Sequence polymorphisms of MC1R gene and their association with depression and antidepressant response. *Psychiatr Genet* 2011; **21**: 14–18.

20 Mehta CR, Patel NR. Exact Logistic Regression: Theory and Examples. *Stat Med* 1995; **14**: 2143–2160.

21 Rothman KJ. The estimation of synergy or antagonism. *Am J Epidemiol* 1976; **103**: 506–511.

22 Rothman KJ. Modern Epidemiology. 2nd edn. Publishers L-R, editor: Philadelphia, PA, 1998.

23 Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P *et al.* A method and server for predicting damaging missense mutations. *Nat Methods* 2010; **7**: 248–249.

24 Lopez-Leon S, Janssens AC, Gonzalez-Zuloeta Ladd AM, Del-Favero J, Claes SJ, Oostra BA et al. Meta-analyses of genetic studies on major depressive disorder. Mol Psychiatry 2008; 13: 772–785.

25 Liu Z, Zhu F, Wang G, Xiao Z, Tang J, Liu W et al. Association study of corticotropin-releasing hormone receptor1 gene polymorphisms and antidepressant response in major depressive disorders. Neurosci Lett 2007; 414: 155–158.

26 Kato T, Iwayama Y, Kakiuchi C, Iwamoto K, Yamada K, Minabe Y et al. Gene expression and association analyses of LIM (PDLIM5) in bipolar disorder and schizophrenia. Mol Psychiatry 2005; 10: 1045–1055.

27 Li C, Tao R, Qin W, Zheng Y, He G, Shi Y et al. Positive association between PDLIM5 and schizophrenia in the Chinese Han population. Int J Neuropsychopharmacol 2008; 11: 27–34.

28 Wong ML, Whelan F, Deloukas P, Whittaker P, Delgado M, Cantor RM et al. Phosphodiesterase genes are associated with susceptibility to major depression and antidepressant treatment response. Proc Natl Acad Sci USA 2006; 103: 15124–15129.

29 Wong ML, Dong C, Maestre-Mesa J, Licinio J. Polymorphisms in inflammation-related genes are associated with susceptibility to major depression and antidepressant response. Mol Psychiatry 2008; 13: 800–812.

30 Collins GA, Tansey WP. The proteasome: a utility tool for transcription? Curr Opin Genet Dev 2006; 16: 197–202.

31 Wang L, Salavaggione E, Pelleymounter L, Eckloff B, Wieben E, Weinshilboum R. Human 3beta-hydroxysteroid dehydrogenase types 1 and 2: Gene sequence variation and functional genomics. J Steroid Biochem Mol Biol 2007; 107: 88–99.

Supplementary Information accompanies the paper on the Molecular Psychiatry website (http://www.nature.com/mp)