Contents lists available at ScienceDirect

Brain, Behavior, & Immunity - Health

journal homepage: www.editorialmanager.com/bbih/default.aspx

Differential expression of MDGA1 in major depressive disorder

Yijie (Jamie) Li, Elizabeth Kresock, Rayus Kuplicki, Jonathan Savitz, Brett A. McKinney *

University of Tulsa, Laureate Institute for Brain Research, USA

ABSTRACT

The identification of gene expression-based biomarkers for major depressive disorder (MDD) continues to be an important challenge. In order to identify candidate biomarkers and mechanisms, we apply statistical and machine learning feature selection to an RNA-Seq gene expression dataset of 78 unmedicated individuals with MDD and 79 healthy controls. We identify 49 genes by LASSO penalized logistic regression and 45 genes at the false discovery rate threshold 0.188. The MDGA1 gene has the lowest P-value (4.9e-5) and is expressed in the developing brain, involved in axon guidance, and associated with related mood disorders in previous studies of bipolar disorder (BD) and schizophrenia (SCZ). The expression of MDGA1 is associated with age and sex, but its association with MDD remains significant when adjusted for covariates. MDGA1 is in a co-expression cluster with another top gene, ATXN7L2 (ataxin 7 like 2), which was associated with MDD in a recent GWAS. The LASSO classification model of MDD includes MDGA1, and the model has a cross-validation accuracy of 79%. Another noteworthy top gene, IRF2BPL, is in a close co-expression cluster with MDGA1 and may be related to microglial inflammatory states in MDD. Future exploration of MDGA1 and its gene interactions may provide insights into mechanisms and heterogeneity of MDD.

1. Introduction

Major depressive disorder (MDD) is a leading cause of disability globally (Rehm and Shield, 2019). Genome-wide association studies (GWAS) of MDD have found numerous variants with small effect sizes (Hyde et al., 2016; Shi et al., 2011). However, there remain gaps in our understanding of the biological mechanisms of MDD and our ability to translate genetic effects clinically. Gene expression is an intermediate level of analysis that can act as a bridge between genetic associations and biological pathways to symptom dimensions. In addition to being an intermediate phenotype, gene expression is dynamic, allowing it to vary with symptom state and environment.

It has been difficult to identify significant single-gene effects at the expression level for MDD. In a large RNA-Seq study of 922 subjects, 29 genes were found to have associations with MDD status at the relaxed false discovery rate (FDR) threshold of 0.25 (Mostafavi et al., 2014). The set of top genes was significantly enriched for the IFNα/β signaling pathway. No genes were differentially expressed at the .05 FDR threshold; however, their results support the role of immune system signaling in the pathogenesis of MDD. A more recent study found one gene with differential expression between HC and MDD with adjusted P = 0.008 (Cole et al., 2021). While they did not find strong single-gene signals for MDD, Cole et al. did find a statistically significant association with a biological aging signature derived from the residuals of a multi-gene model of chronological age (Cole et al., 2021).

In our previous analysis of whole-blood RNA-Seq of MDD, we used a read alignment protocol that enriched for the expression of antisense RNA (Wanowska et al., 2018), and we developed an approach that identified an antisense gene module for MDD (Le et al., 2018, 2020). Antisense expression analysis can complement regular gene expression by revealing regulatory effects. For example, antisense transcripts can act in cis or trans to inhibit the expression of a gene. In the current study, we use standard gene expression, which uses a sense alignment protocol (strandedness set to reverse). We compare statistical and machine learning feature selection approaches to identify multivariate classification models of MDD diagnosis. We then use hierarchical cluster analysis to explore relationships between the top MDD genes.

2. Materials and methods

2.1. RNA-sequencing alignment and annotation

We use RNA-Seq gene expression data from a study of major depressive disorder (MDD) with 78 unmedicated MDD and 79 healthy controls (HC) described in Ref (Le et al., 2018). The data contain 66 males and 91 females and an age range of 18–55. Participants between the ages of 18 and 55 years were recruited from the clinical services of the Laureate Psychiatric Clinic and Hospital (LPCH) and radio and print advertisement in the Tulsa metropolitan area, Oklahoma. The study included 78 subjects who met DSM-IV-TR criteria for MDD (52 females, mean age = 33 ± 11) and 79 HCs who showed no history of any major psychiatric disorder, personally or in a first-degree relative (41 females,

* Corresponding author.
E-mail address: brett-mckinney@utulsa.edu (B.A. McKinney).

https://doi.org/10.1016/j.bbih.2022.100534
Received 21 May 2022; Received in revised form 8 September 2022; Accepted 9 October 2022
Available online 10 October 2022

mean age = 31 ± 10). The HCs were matched to the MDD cohort based on age and sex. See Table 1 of Ref (Le et al., 2018). for additional sample characteristics such as occupational status, educational status, smoking status, BMI, and MADRS. In the previous analysis, we used stranded RNA-Seq preprocessing, which enriches for antisense non-coding RNA, sometimes called Natural Antisense Transcripts (NATs). In the current study, we process the data with a sense alignment protocol (strandedness set to reverse), which enriches for gene expression. The RNA-Seq expression was derived from PBMCs isolated from morning blood samples. We normalize to counts per million reads followed by quantile normalization and log2 transformation. We filter genes lower than 0.045 coefficient of variation from 8923 genes, leaving 5587 genes. The preprocessed expression data and analysis code are available at https://github.com/insilico/DepressionGeneModules.

2.2. Feature selection

We use least absolute shrinkage and selection operator (LASSO) for multivariate feature selection. LASSO penalized logistic regression is a multivariate linear model that includes a penalty to efficiently select important features while mitigating correlation. We use the penalty hyperparameter that minimizes the cross-validation prediction error, which reduced the dimensionality to 49 genes from 5587 filtered genes. Prediction is based on the classification of MDD versus healthy controls using logistic regression. We compare LASSO selected genes with univariate logistic regression P-values corrected for multiple hypotheses with the BH adjustment. Additionally, we compare the top genes with lowest FDR-adjusted univariate P-value (45 genes), where each model is adjusted for age and sex. We explore the relationship between the top univariate genes using hierarchical clustering of the Euclidean distance matrix.

Table 1
Univariate logistic regression results of genes with adjusted P-values at 0.187 FDR (45 genes). Genes are sorted by raw P-value and separated by over (left) and under expression (right) in MDD versus healthy control. Hierarchical clustering of these genes in Fig. 1.

Over Expressed Genes in MDD			Under Expressed Genes in MDD		
gene	coefficient	raw P value	gene	coefficient	raw P value
MDGA1	3.272226	4.92E-05	NPFF	−4.93599	2.93E-04
ZDHHC20	3.447524	1.66E-04	FAM138A	−3.90726	4.31E-04
IRF2BPL	4.02807	3.99E-04	POGZ	−3.36093	6.20E-04
ARFGAP1	4.502432	4.08E-04	KANTR	−4.04004	6.57E-04
UBD	2.025948	4.76E-04	ZMYM6NB	−2.72513	6.66E-04
BCL2L12	3.550422	6.05E-04	ZNF658	−3.2259	7.14E-04
ZFP36L2	2.623329	6.23E-04	RCSD1	−4.11089	7.30E-04
ICOS	3.337997	7.09E-04	SS18L2	−3.22678	8.16E-04
CBL	4.308463	7.22E-04	FAM193A	−3.73778	8.29E-04
TSR3	4.027897	7.81E-04	CCAR2	−3.47723	8.30E-04
DYNLRB1	3.316966	8.48E-04	RARS2	−2.71883	8.88E-04
HNRNPUL2	5.406151	8.73E-04	PANX1	−3.6218	8.95E-04
TADA3	3.483415	9.39E-04	TSC2	−3.33729	9.33E-04
CD83	2.352797	1.21E-03	ISX	−5.30674	9.43E-04
C16orf74	3.829435	1.31E-03	SETDB2	−5.09759	9.95E-04
ATXN7L2	3.266239	1.35E-03	DPEP3	−2.34135	1.03E-03
			NDUFB1	−3.75476	1.13E-03
			SUPT7L	−5.36926	1.14E-03
			STX16-NPEPL1	−3.39406	1.18E-03
			CSPG4	−4.48842	1.19E-03
			CDK5RAP2	−3.66751	1.23E-03
			ABCE1	−2.82685	1.31E-03
			TRPC5	−4.58532	1.35E-03
			SRSF5	−3.19411	1.36E-03
			NGRN	−3.75049	1.37E-03
			ELP3	−3.32066	1.48E-03
			C1orf35	−3.03975	1.50E-03
			EIF2D	−3.64157	1.50E-03
			MCM3AP-AS1	−4.74524	1.51E-03

3. Results

The LASSO penalty term (lambda = 0.7855) that minimizes the cross-validation error results in a model containing 49 genes with non-zero coefficients and 78.98% cross-validation accuracy. MDGA1 (MAM domain–containing glycosylphosphatidylinositol anchor 1) has the highest LASSO coefficient (0.15, Fig. 1A) and the lowest nominal P-value (4.92e-05). Univariate logistic regression results in 45 genes with adjusted P-values at 0.187 FDR (Table 1). The value 0.187 is the lowest FDR threshold that yields significant associations for MDD. This adjusted P-value means approximately 18.7% of the 45 genes are false positives (8 of the 45 genes). The top gene, MDGA1, has been associated with other psychiatric disorders such as BD and SCZ (Kahler et al., 2008; Li et al., 2011). Hierarchical clustering of the top 45 univariate genes shows a cluster of genes correlated with MDGA1 (Fig. 1B), including ATXN7L2 (ataxin 7 like 2), which had a genetic association with MDD in a previous study (Shi et al., 2011). MDGA1 has both age and sex P-value lower than 0.05 in the logistic regression of gene expression as outcome and age and sex as predictors (Supplementary Fig. 2). However, MDGA1 remains significant for MDD when we adjust for age and sex. We find 31 intersecting genes between the top 45 MDD genes selected by logistic regression P-value without covariates and the top 45 genes when the model is adjusted for age and sex. MDGA1 is among the 31 intersecting genes (Supplementary Table 1).

4. Conclusion

Our univariate RNA-Seq analysis yields 45 genes with false discovery rate 0.187, which can be interpreted to mean that approximately 8 of the top 45 MDD genes are false positives. The top main effect, MDGA1, has not been implicated previously in MDD, but multiple studies have associated intronic single nucleotide polymorphisms in MDGA1 with SCZ and BD (Kahler et al., 2008) (Li et al., 2011). The identification of a known SCZ and BD gene in our MDD sample is not surprising due to the large degree of pleiotropy in psychiatric disorders. A recent GWAS meta-analysis of 8 psychiatric disorders found 109 significant loci shared by at least two disorders (Cross-Disorder Group of the Psychiatric Genomics Consortium, 2019). Using the shared genetics between the eight disorders, they found three clusters, where one of the clusters represents mood and psychotic disorders (MDD, BD, and SCZ). The genetic correlation using linkage disequilibrium score regression analyses was highest between BD and SCZ (0.7), and the genetic correlation between MDD and BD (0.36) was very similar to the genetic correlation between MDD and SCZ (0.34). The shared genetics with other disorders suggests heterogeneity within MDD, and MDGA1 could be investigated as a biological basis for MDD subtypes.

In mouse studies, MDGA1 is expressed in somatosensory areas of the brain and is important for forebrain development (Takeuchi et al., 2007). MDGA1 is a member of the immunoglobulin domain cell adhesion molecule subfamily, and neuronal cell adhesion molecules have been implicated in psychiatric disorders (Vawter, 2000). In our study, MDGA1 is highly correlated with ATXN7L2 (Fig. 1B), and although not genome-wide significant in Ref. (Shi et al., 2011), ATXN7L2 was one of the most significant associations in a GWAS of early-onset MDD. The cluster of genes that are over-expressed in MDD (Fig. 1B) includes MDGA1 and ATXN7L2. These two genes are in a subcluster next to IRF2BPL (Interferon regulatory factor 2 binding protein like), which is associated with neurological phenotypes (Marcogliese et al., 2018). IRF2BPL and IRF2BP2 are corepressors of IRF2, and IRF2BP2 is a regulator of microglial polarization (Vawter, 2000). Microglia are macrophages in the central nervous system (CNS) that can become polarized to an M1 pro-inflammatory state or to an M2 anti-inflammatory state. In a recent study, IRF2BP2-deficient macrophages showed elevated expression of IRF2BPL, likely to compensate for

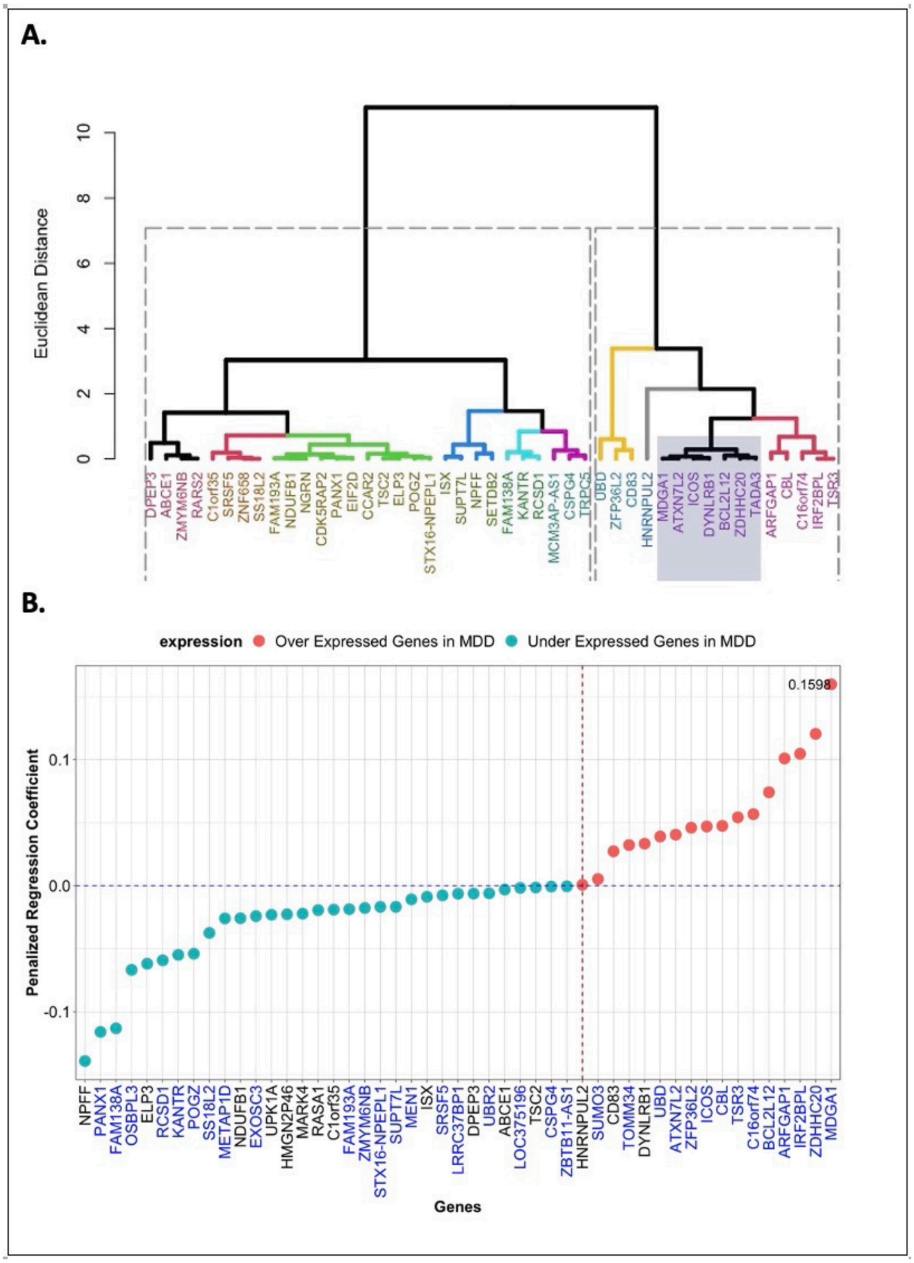

Fig. 1. Top MDD associated genes: Hierarchical clustering (Euclidean) of the 45 MDD-associated genes selected by univariate logistic regression with FDR (subplot **A**). The cluster containing the top gene, MDGA1, is shaded, and the IRF2BPL gene is in the adjacent cluster. Genes in the main left/right bifurcation, as separated by the dashed lines, are under/over expressed in MDD. Penalized regression coefficients of the 49 genes with non-zero LASSO coefficients (subplot **B**). The top over-expressed gene in MDD is MDGA1 with 0.1598 penalized coefficient value followed by IRF2BPL. The top under-expressed gene in MDD relative to controls is NPFF. The 35 overlapping genes between the LASSO and univariate analyses are colored blue on the horizontal axis of subplot B. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

the deficiency (Vawter, 2000).

The current study uses well-characterized and unmedicated MDD subjects and matched controls in an examination of gene expression differences using RNA-Seq. The top gene, MDGA1, is not significant at the traditional 0.05 adjusted P-value; however, we provide evidence for its differential expression between MDD and healthy controls based on the false discovery rate, prior studies in related disorder, and the biological function of this gene. MDGA1 is associated with sex and age (Supplementary Fig. 2); however, it remains a top MDD gene when adjusted for these covariates. Future characterization of the interactions between MDGA1, IRF2BPL and other genes in their clusters may help identify treatment targets and better understand MDD heterogeneity.

Declaration of competing interest

None of the authors have a conflict of interest.

Data availability

Data will be made available on request.

Acknowledgements

This study was supported by the National Institute of Mental Health NIH R01MH098099 award and National Institute of General Medical Sciences NIH P20GM121312 award, and in part by the William K. Warren Foundation.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.bbih.2022.100534.

References

Cole, J.J., McColl, A., Shaw, R., Lynall, M.E., Cowen, P.J., de Boer, P., Drevets, W.C., Harrison, N., Pariante, C., Pointon, L., et al., 2021. No evidence for differential gene expression in major depressive disorder PBMCs, but robust evidence of elevated biological ageing. Transl. Psychiatry 11 (1), 404.

Cross-Disorder Group of the Psychiatric Genomics Consortium, 2019. Electronic address pmhe, cross-disorder Group of the psychiatric Genomics C: **genomic relationships, novel loci, and pleiotropic mechanisms across eight psychiatric disorders**. Cell 179 (7), 1469–1482 e1411.

Hyde, C.L., Nagle, M.W., Tian, C., Chen, X., Paciga, S.A., Wendland, J.R., Tung, J.Y., Hinds, D.A., Perlis, R.H., Winslow, A.R., 2016. Identification of 15 genetic loci associated with risk of major depression in individuals of European descent. Nat. Genet. 48 (9), 1031–1036.

Kahler, A.K., Djurovic, S., Kulle, B., Jonsson, E.G., Agartz, I., Hall, H., Opjordsmoen, S., Jakobsen, K.D., Hansen, T., Melle, I., et al., 2008. Association analysis of schizophrenia on 18 genes involved in neuronal migration: MDGA1 as a new susceptibility gene. Am J Med Genet B Neuropsychiatr Genet 147B (7), 1089–1100.

Le, T.T., Savitz, J., Suzuki, H., Misaki, M., Teague, T.K., White, B.C., Marino, J.H., Wiley, G., Gaffney, P.M., Drevets, W.C., et al., 2018. Identification and replication of RNA-Seq gene network modules associated with depression severity. Transl. Psychiatry 8 (1), 180.

Le, T.T., Dawkins, B.A., McKinney, B.A., 2020. Nearest-neighbor Projected-Distance Regression (NPDR) for detecting network interactions with adjustments for multiple tests and confounding. Bioinformatics 36 (9), 2770–2777.

Li, J., Liu, J., Feng, G., Li, T., Zhao, Q., Li, Y., Hu, Z., Zheng, L., Zeng, Z., He, L., et al., 2011. The MDGA1 gene confers risk to schizophrenia and bipolar disorder. Schizophr. Res. 125 (2–3), 194–200.

Marcogliese, P.C., Shashi, V., Spillmann, R.C., Stong, N., Rosenfeld, J.A., Koenig, M.K., Martinez-Agosto, J.A., Herzog, M., Chen, A.H., Dickson, P.I., et al., 2018. IRF2BPL is associated with neurological phenotypes. Am. J. Hum. Genet. 103 (3), 456.

Mostafavi, S., Battle, A., Zhu, X., Potash, J.B., Weissman, M.M., Shi, J., Beckman, K., Haudenschild, C., McCormick, C., Mei, R., et al., 2014. Type I interferon signaling genes in recurrent major depression: increased expression detected by whole-blood RNA sequencing. Mol. Psychiatr. 19 (12), 1267–1274.

Rehm, J., Shield, K.D., 2019. Global burden of disease and the impact of mental and addictive disorders. Curr. Psychiatr. Rep. 21 (2), 10.

Shi, J., Potash, J.B., Knowles, J.A., Weissman, M.M., Coryell, W., Scheftner, W.A., Lawson, W.B., DePaulo Jr., J.R., Gejman, P.V., Sanders, A.R., et al., 2011. Genome-wide association study of recurrent early-onset major depressive disorder. Mol. Psychiatr. 16 (2), 193–201.

Takeuchi, A., Hamasaki, T., Litwack, E.D., O'Leary, D.D., 2007. Novel IgCAM, MDGA1, expressed in unique cortical area- and layer-specific patterns and transiently by distinct forebrain populations of Cajal-Retzius neurons. Cerebr. Cortex 17 (7), 1531–1541.

Vawter, M.P., 2000. Dysregulation of the neural cell adhesion molecule and neuropsychiatric disorders. Eur. J. Pharmacol. 405 (1–3), 385–395.

Wanowska, E., Kubiak, M.R., Rosikiewicz, W., Makalowska, I., Szczesniak, M.W., 2018. Natural antisense transcripts in diseases: from modes of action to targeted therapies. Wiley Interdiscip Rev RNA 9 (2).