



Plasma sterols and depressive symptom severity in a population-based cohort

Basar Cenik^{1®}, Can Cenik^{2®}, Michael P. Snyder², E. Sherwood Brown¹*

- 1 Department of Psychiatry, The University of Texas Southwestern Medical Center, Dallas, Texas, United States of America, 2 Department of Genetics, Stanford University School of Medicine, Stanford, California, United States of America
- These authors contributed equally to this work.
- * sherwood.brown@utsouthwestern.edu



Convergent evidence strongly suggests major depressive disorder is heterogeneous in its etiology and clinical characteristics. Depression biomarkers hold potential for identifying etiological subtypes, improving diagnostic accuracy, predicting treatment response, and personalization of treatment. Human plasma contains numerous sterols that have not been systematically studied. Changes in cholesterol concentrations have been implicated in suicide and depression, suggesting plasma sterols may be depression biomarkers. Here, we investigated associations between plasma levels of 34 sterols (measured by mass spectrometry) and scores on the Quick Inventory of Depressive Symptomatology-Self Report (QIDS-SR₁₆) scale in 3117 adult participants in the Dallas Heart Study, an ethnically diverse, population-based cohort. We built a random forest model using feature selection from a pool of 43 variables including demographics, general health indicators, and sterol concentrations. This model comprised 19 variables, 13 of which were sterol concentrations, and explained 15.5% of the variation in depressive symptoms. Desmosterol concentrations below the fifth percentile (1.9 ng/mL, OR 1.9, 95% CI 1.2-2.9) were significantly associated with depressive symptoms of at least moderate severity (QIDS-SR₁₆ score ≥10.5). This is the first study reporting a novel association between plasma concentrations cholesterol precursors and depressive symptom severity.



G OPEN ACCESS

Citation: Cenik B, Cenik C, Snyder MP, Brown ES (2017) Plasma sterols and depressive symptom severity in a population-based cohort. PLoS ONE 12(9): e0184382. https://doi.org/10.1371/journal.pone.0184382

Editor: David Meyre, McMaster University, CANADA

Received: June 20, 2017

Accepted: August 22, 2017

Published: September 8, 2017

Copyright: © 2017 Cenik et al. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: The Executive Committee of the Dallas Heart Study has imposed restrictions on sharing the de-identified data set due to the potential loss of anonymity of participants. Requests for access to the data can be addressed to Dr. Helen H. Hobbs, Director, Donald W. Reynolds Cardiovascular Clinical Research Center, The University of Texas Southwestern Medical Center, 5323 Harry Hines Boulevard NB10.204A MS 8591, Dallas, Texas 75930.

Introduction

Major depressive disorder (MDD) is a highly prevalent disease and a leading cause of disability worldwide [1]. MDD is heterogeneous in its etiology and clinical characteristics [2]. Depression biomarkers hold potential for identifying etiological subtypes, improving diagnostic accuracy, predicting treatment response, and personalization of treatment. Changes in various biochemical pathways have been associated with MDD, including inflammatory, neurotrophic, and hypothalamic-pituitary-adrenal (HPA) axis alterations [3]. In contrast, plasma lipids have received relatively little attention as depression biomarkers despite a long-recognized association between low plasma cholesterol, suicidal behavior, and depression [4–11].



Funding: This work was supported in part by grant UL1TR001105 from the National Center for Advancing Translational Science (https://ncats.nih. gov/), National Institutes of Health (NIH, https:// www.nih.gov/), in part by the National Institute of Mental Health (https://www.nimh.nih.gov/index. shtml) of the National Institutes of Health under Award Number R25MH101078, and in part by NIH grant 1K99CA20452201 to CC. The content is solely the responsibility of the authors and does not necessarily represent the official views of the Center for Translational Medicine, The University of Texas Southwestern Medical Center and its affiliated academic and health care centers, the National Center for Advancing Translational Sciences, or the National Institutes of Health. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: I have read the journal's policy and the authors of this manuscript have the following competing interests: Dr. Brown has been funded by the NIH, the Stanley Medical Research Institute and Dreyfus Health Foundation and received an honorarium from Genentech. Dr. Snyder is co-founder and a member of the scientific advisory board (SAB) of Personalis and a member of the SAB of Genapsys. Dr. B. Cenik and Dr. C. Cenik report no financial conflicts of interests. This does not alter our adherence to PLOS ONE policies on sharing data and materials.

Human plasma contains a large number of lipids that have not been systematically investigated [12]. Twenty-nine of these lipids, as well as commonly assessed lipoprotein fractions, were recently quantified with mass spectrometry in a large, ethnically diverse, phenotypically well-characterized, population-based sample in the Dallas Heart Study (DHS) [13]. Between 2001 and 2012, more than 115 reports have been published based on the DHS data, mostly on cardiovascular outcomes and genetics of intermediary metabolism. Most relevant to our study, Stiles et al. [13] screened more than 60 sterol species and identified 29 that are consistently detected in human plasma. Of these 29, plasma concentrations of several sterols varied interindividually, and by age, gender, and ethnicity. Some of these sterols, namely sitosterol, campesterol, stigmasterol, and stigmastanol, are plant sterols thought to be dietary in origin. Stiles et al. also identified 16 genetic loci correlated with concentrations of 19 sterols in a genome wide association study in this sample. Remarkably, many of these sterols are under-studied despite the central role played by cholesterol and its metabolites in human metabolism and disease.

There is ample reason to hypothesize that sterols may be associated with depression. Low plasma and postmortem brain cholesterol concentrations have been found to correlate with increased risk for depression and suicidal behavior. [4, 5, 8, 10, 11]. Oxysterols, oxygenated derivatives of cholesterol, modulate the Sonic Hedgehog pathway, NMDA receptors, and the expression of nuclear receptors in retinal neurons, in addition to being intermediates in cholesterol catabolism, potential regulators of cholesterol homeostasis, and transport forms of cholesterol. Some oxysterols accumulate in postmortem brains from Alzheimer disease patients [14–16]. Vitamin D, a secosteroid with potential roles in brain development, adult brain function, and neuropsychiatric illness has also been linked to depression [17].

In this study, we hypothesized that concentrations of plasma lipids, including oxysterols, sterol and secosteroid metabolism intermediaries, lipoproteins and dietary sterols are predictive of depressive symptom severity scores in the DHS sample. This is the first DHS report examining plasma sterols in relation to depression.

Materials and methods

Study population

This investigation was conducted as part of the Dallas Heart Study (DHS). DHS was undertaken in 2000 as a single-site (Dallas County, Texas, USA), multiethnic, cross-sectional, population-based study of cardiovascular health. The Dallas Heart Study 2 (DHS-2) is a longitudinal follow-up study of a subset of participants who returned from 2007 onwards for a second clinical examination, an extensive health survey, laboratory testing and imaging studies. Of note, DHS was a probability-based, epidemiological sample representative of Dallas County, except for intentional over-sampling of African American ethnicity (~52% of subjects who provided a plasma sample); the subjects were not selected based on presence of heart disease or any cardiovascular risk factors. Sampling design, recruitment procedures and other details of the study population have been published [13, 18]. Out of the 3402 subjects included in DHS-2 sample, plasma sterol concentrations were available for 3228. Of those, we included data from the 3117 participants who had completed at least one item on the QIDS-SR₁₆ scale.

This study was approved by the Institutional Review Board of The University of Texas Southwestern Medical Center as part of the Dallas Heart Study and all subjects provided written informed consent after receiving a complete description of the study.

The Executive Committee of the Dallas Heart Study has imposed restrictions on sharing the de-identified data set due to the potential loss of anonymity of participants. Requests for access to the data can be addressed to Dr. Helen H. Hobbs, Director, Donald W. Reynolds



Cardiovascular Clinical Research Center, The University of Texas Southwestern Medical Center, 5323 Harry Hines Boulevard NB10.204A MS 8591, Dallas, Texas 75930.

Measurement of sterol and lipoprotein concentrations

Data on sterol and lipoprotein concentrations was obtained from the DHS executive committee. The mean, median, and range for all sterol concentrations have been previously published and can be found in Table 1 of Stiles et al. [13].

Briefly, blood samples were collected in the fasting state; deuterated standards were added to plasma; lipids were extracted from saponified plasma; the sample was hydrolyzed, followed by solid phase extraction; compounds were resolved with high performance liquid chromatography and gas chromatography (GC) and quantified by quadrupole mass spectrometry (MS) or GC-MS. The complete methods and assay characteristics have been published [13, 19, 20]. The 34 plasma sterols and lipoproteins shown in Table 1 were included in this study.

Table 1. 34 plasma sterols and lipoproteins investigated in this study.

14-Desmethyl lanosterol
22R-Hydroxycholesterol
24,25-Epoxycholesterol
24-Dihydrolanosterol
24-Oxocholesterol
24S-Hydroxycholesterol
25-Hydroxycholesterol
25-Hydroxyvitamin D₂
25-Hydroxyvitamin D ₃
27-Hydroxycholesterol
4β-Hydroxycholesterol
5,6α-Epoxycholesterol
5,6β-Epoxycholesterol
5α-Hydroxycholesterol
7-Dehydrocholesterol
7-Oxocholesterol
7α,27-Dihydroxycholesterol
7α-Hydroxycholesterol
8-Dehydrocholesterol
Campesterol
Cholestanol
Cholestenone
Desmosterol
HDL Cholesterol
Lanosterol
Lathosterol
LDL Cholesterol
Sitosterol
Stigmastanol
Stigmasterol
Total Cholesterol
Triglycerides
VLDL Cholesterol
Zymosterol
https://doi.org/10.1971/journal.none.019/399.4001



Measurement of depressive symptom severity

Depressive symptom severity was measured using QIDS-SR₁₆, a validated, widely-used, self-report instrument [21–25]. The scale is freely available online at http://www.ids-qids.org/tr-english.html. This scale is composed of 16 questions, rated 0 to 3. For questions 1 to 4, 6 to 9, and 15 to 16, only the highest rating is taken into account for the final score, resulting in a scale composed of 9 items corresponding to the 9 criteria for major depressive episodes, scored 0 to 3, for a maximum possible score of 27. The mean scores and standard deviations for all 9 items are shown in Table 2.

86.3% of subjects had completed all items on QIDS-SR $_{16}$, while 9.5% had left exactly one item blank. The remaining 4.1% had two or more missing items. To address the missing values, we employed a k-nearest neighbor method for imputation [26]. Eleven subjects had answered less than half the items. For these eleven subjects, we assigned the mean score for that item instead of the missing value. After the imputation step, we calculated the total QIDS-SR $_{16}$ scores as previously described [21]. Although scores on QIDS-SR $_{16}$ are normally integers, the imputation method assigns decimals; therefore, we considered scores \geq 10.5 as indicative of depressive symptoms of at least moderate severity. Depressive symptom severity was modeled continuously for primary analysis and dichotomized only for secondary analysis.

Statistical analysis and random forest classifier training

All statistical analyses were carried out in R software (version 3.2.5 [2016-04-14]). As predictor variables for QIDS-SR₁₆ score, we used 34 sterol and lipoprotein concentrations and nine demographic and general health related parameters. All sterol concentrations, as well as age, formal education in years, and number of chronic illnesses were treated as quantitative variables. Number of chronic illnesses was based on self-report and defined as the number of positive responses to the question "Has a doctor ever told you that you have X?" where X is one of the following: "heart attack", "stroke", "heart failure", "diabetes", "hypertension", "emphysema", "hepatitis", "kidney disease", "lupus", "rheumatoid arthritis", "inflammatory bowel disease", "pulmonary sclerosis", "Alzheimer disease", "obstructive sleep apnea", or "cancer". Yearly household income was categorized into 10 income brackets, 1 denoting the lowest income and 10 denoting the highest income, and used as a quantitative variable. Gender, ethnicity (Hispanic, non-Hispanic white, non-Hispanic black, or other) and marital status (married, living as married/living with partner, separated/divorced, widowed, or never married) were treated as categorical variables. Drinking status was classified into four categories, "lifelong abstainer", "recent abstainer", "low risk drinking", or "high risk drinking". High risk drinking was defined as more than 14 standard drinks per week for men and more than 7

Table 2. Mean scores and standard deviations for 9 items comprising the QIDS-SR₁₆.

Item (Symptom)	Mean	Standard Deviation
1–4 (Sleep)	2.034	0.865
5 (Sadness)	0.523	0.743
6–9 (Appetite)	1.051	1.099
10 (Concentration)	0.402	0.649
11 (View of self)	0.243	0.688
12 (Suicidality)	0.082	0.356
13 (Interest)	0.258	0.609
14 (Energy)	0.494	0.729
15–16 (Psychomotor)	0.534	0.887



standard drinks per week for women, based on Centers for Disease Control and Prevention recommendations. Smoking status was analyzed categorically: "never smoker", "past smoker", or "current smoker".

The sterol levels are correlated with each other [13]. Furthermore, given the biochemical synthesis pathways, the levels of certain sterols are conditionally dependent on each other. Therefore, a simple regression framework is unsuitable and we instead adopted a random forest approach for modeling [27]. Random forest models were constructed using "randomForest" library in R [28]. We assessed the prediction performance using explained variation as defined by

$$1 - \sum (y - \hat{y})^2 / \sum (y - \mu_y)^2$$

where y is the QIDS-SR $_{16}$ score, μ_y is the mean QIDS-SR $_{16}$ score, and \hat{y} is the predicted score from the random forest model.

To assess feature importance, we used a permutation based approach. Specifically, values of each variable is permuted while holding the other variables constant and the percent increase in mean squared error for out-of-bag samples is calculated. Note that by definition, this approach can underestimate the importance of correlated variables [27].

For feature selection, we used the 'rfe' function in the "caret" R package [29]. Specifically, we employed a recursive feature elimination approach using a 5-fold repeated cross-validation scheme. We used the variation explained as the metric to be maximized. We allowed for either 1% or 5% tolerance to ensure a balance between performance and the number of selected variables. We utilized the "doMC" R package to parallelize the feature selection [30]. In this validation approach, the scoring for a subject as a function of sterol measurements was performed using a model that had not been trained on that subject, enabling the separation of training and validation samples in an unbiased way.

In secondary analysis, we defined depressive symptoms of at least moderate severity as corresponding to QIDS-SR $_{16}$ scores ≥ 10.5 . We used Fisher's exact test to calculate the odds ratios and p-values of association between selected sterol concentrations and depressive symptoms of at least moderate severity. Spearman's method was used to calculate correlation coefficients. p values were corrected for multiple comparisons with Holm's method. Fig 1 summarizes the pre-specified data analysis plan.

Results

We analyzed data from 3117 adults (1874 females, 1243 males) from the Dallas Heart Study (DHS)-2 cohort. First, we determined that 13.7% of the participants had not answered one or more questions on the QIDS-SR₁₆. To assess any potential systematic differences between subjects that completed all questions and those that left at least one question blank, we carried out two complementary analyses. First, for each question, we compared the score distribution of subjects with no missing values and at least one missing value. We found that, except for item 14 (Energy level: 0: There is no change in my usual level of energy. 1: I get tired more easily than usual. 2: I have to make a big effort to start or finish my usual daily activities. 3: I really cannot carry out most of my usual daily activities because I just don't have the energy.), there was no statistically significant difference between the two groups (Bonferroni-corrected chisquared test, p > 0.05). Second, we analyzed the mean score on questions that were completed. For individuals that completed the entire questionnaire, this is simply their total score divided by nine. Subjects who left more than three questions blank had higher per question severity scores (median score 0.56 for subjects with no missing items versus 1.5 for subjects with more than 3 missing items; Wilcoxon rank sum test, $p = 1 \times 10^{-10}$). These systematic differences are consistent with more depressed subjects being less likely to complete the entire questionnaire.



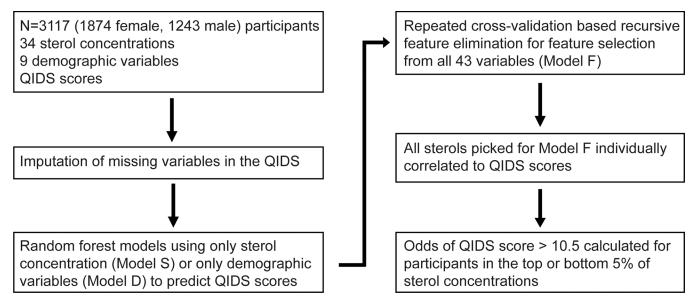


Fig 1. Pre-specified data analysis flowchart. The models S, D, and F are further specified in Table 4. QIDS, Quick Inventory of Depressive Symptomatology-Self Report.

Hence, we decided to impute the missing values since excluding these potentially more severely depressed subjects would have biased our results and reduced the power of our analysis. Nonetheless, we repeated all analyses presented in the manuscript using only subjects that completed the entire questionnaire (i.e., without imputation) and observed qualitatively consistent results for all reported findings (data not shown).

Median score on the QIDS-SR₁₆ was 5 (range 0–24, S1 Fig) and 11.6% of subjects had scores greater than 10.5 (corresponding to depressive symptoms of at least moderate severity) after imputation of missing values. This figure is roughly consistent with the reported 12-month prevalence of major depression (5–10%) in North America [31, 32]. Other characteristics of the subjects are summarized in Table 3.

Next, we investigated whether plasma sterol concentrations could be used to predict depressive symptom severity as assessed by the QIDS-SR₁₆. We treated QIDS-SR₁₆ scores as a quantitative variable. Plasma sterol concentrations are known to be highly correlated with each other [13]. Given the dependencies between the sterol concentrations, we decided to avoid standard regression-based approaches and used a machine learning approach to model the relationship between depression symptom severity and plasma sterol concentrations. We constructed a random forest model (RFM) using concentrations of 34 different sterols and lipoproteins. This model explained 4.2% of the variation in QIDS-SR₁₆ scores in the study population (Model S in Table 4). Table 5 shows all 34 variables used in the model ranked by importance according to percent increase in mean squared error metric.

Plasma concentrations of several sterols have been reported to correlate with gender and ethnicity [13], both of which are associated with depression prevalence [31]. Hence, it is possible that the predictive power of sterols is solely driven by their relationship to gender and ethnicity. To test this hypothesis, we first built a model using only age, gender, and ethnicity as variables (Model A in Table 4). These three variables explained only 2.2% of the variation in QIDS-SR₁₆ scores. To find out whether sterol concentrations provide any further predictive information on top of age, gender, and ethnicity, we combined these three demographic variables with the sterol concentrations to build an RFM with 37 variables (Model S+A in Table 4);



	Number of subjects	3117 60	
Gender	Female (%1)		
Age	Median (range)	50 (18–85)	
Ethnicity (%1)	Black	51	
	Non-Hispanic White	33	
	Hispanic	14	
	Other	2	
Annual household income	25 th percentile	\$20,000-\$25,000	
	Median	\$35,000-\$40,000	
	75 th percentile	\$50,000-\$75,000	
Formal education in years	Median (range)	13 (0–16)	
High-risk alcohol use	Percentage ¹	7	
Current smoker	Percentage ¹	22	
Number of chronic illnesses	Median (range)	1 (0-7)	
QIDS-SR ₁₆ score ²	Median (range)	5 (0–24)	
	≥ 10.5 (%)	11.6	

^{1,} rounded to the nearest integer

this model explained 7.4% of the variation in QIDS-SR $_{16}$ scores, an apparently additive improvement in predictive accuracy. The variable with the highest importance by percent increase in mean squared error metric in this model was gender, followed by desmosterol concentration.

For plasma biomarkers to be clinically useful, they should have predictive value over and above known risk factors for depression. To test this, we constructed Model D from nine demographic and general health variables previously suggested to be correlated with depression risk and available in the DHS-2 dataset: gender, ethnicity, age, household income, education in years, alcohol use, smoking status, number of chronic illnesses, and marital status [32, 33]. The RFM constructed from these nine variables explained 12.8% of the variation in QIDS-SR₁₆ scores. Next, we wanted to test whether inclusion of a subset of sterols may further improve the predictive power of a model that uses the known demographic risk factors. Hence, we employed a repeated cross-validation based recursive feature elimination scheme to determine a subset of these 43 variables (34 sterols + 9 demographics). We adopted a random forest validation approach, such that the scoring for subject as a function of sterol measurements was performed using a model that had not been trained on that subject. Hence, our method enables the separation of training and validation samples in an unbiased way. We set a tolerance threshold of 5% increase in variation explained to balance parsimony with explained variation. This model yielded 19 variables, 13 of which were sterol concentrations. For

Table 4. Summary of random forest models.

Model	Variables used	Number of Variables	% Variation in QIDS-SR ₁₆ Scores explained
S	Sterol concentrations	34	4.2
A	Age, gender, and ethnicity	3	2.2
S+A	Sterol concentrations, age, gender, and ethnicity	37	7.4
D	Demographics and general health indicators	9	12.8
F	Selected features from models S and D	19	15.5

², after imputation of missing items



Table 5. Variables used in random forest model s, ranked in order of importance.

Sterol	% Increase in mean squared error		
Desmosterol	37.19		
14-Desmethyl lanosterol	27.35		
Sitosterol	27.24		
Stigmasterol	25.41		
Lathosterol	25.03		
Lanosterol	24.79		
7-Dehydrocholesterol	23.93		
Cholestanol	22.36		
7-Oxocholesterol	22.11		
25-Hydroxycholesterol	20.97		
Campesterol	20.96		
25-Hydroxyvitamin D ₃	19.94		
Cholestenone	19.63		
5α-Hydroxycholesterol	18.87		
5,6β-Epoxycholesterol	18.43		
Total Cholesterol	17.91		
4β-Hydroxycholesterol	16.55		
24-Dihydrolanosterol	16.23		
22R-Hydroxycholesterol	15.67		
Zymosterol	15.16		
24,25-Epoxycholesterol	14.40		
5,6α-Epoxycholesterol	13.39		
LDL Cholesterol	12.99		
7α-Hydroxycholesterol	12.91		
8-Dehydrocholesterol	12.73		
HDL Cholesterol	11.78		
24S-Hydroxycholesterol	10.74		
24-Oxocholesterol	9.85		
27-Hydroxycholesterol	9.49		
VLDL Cholesterol	9.41		
Triglycerides	8.68		
7α,27-Dihydroxycholesterol	6.97		
Stigmastanol	5.88		
25-Hydroxyvitamin D ₂	0.96		

comparison, setting the threshold at 1% would have yielded 22 variables (Fig 2). An RFM constructed with these 19 variables explained 15.5% of the variation in QIDS-SR₁₆ scores, better than all the other models tested (Model F, Table 4). Household income was the most important variable predicting depressive symptom severity scores, followed by gender and the concentrations of 14-desmethyl lanosterol, sitosterol, and desmosterol (Fig 3). Next, we applied Model F to female-only and male-only subsets of the dataset. Model F explained 13.7% of the variation in QIDS scores in the female-only sample and 6.9% of the variation in the male only sample. Few male participants had QIDS scores >10.5, likely limiting the power of the analysis. The most important variables in the female-only model were income category, smoking status, age, education, 14-desmethyl-lanosterol, 7-dehydrocholesterol (7-DHC), sitosterol, lathosterol, 5-6- β -hydroxycholesterol and desmosterol, which is qualitatively similar to the results from the full dataset.

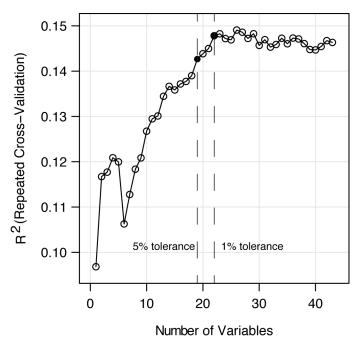


Fig 2. Unbiased feature selection using a recursive elimination method. A total of 34 sterol concentration measurements and 9 demographic parameters were used to predict QIDS-SR16 scores from 3117 subjects. A repeated cross-validation approach is used such that data is partitioned into test and training sets via resampling. Performance as measured by percent variation explained as a function of the number of features used in training the RFM is shown.

Finally, we wanted to explore the interpretability of our model. For this purpose, we tested whether any of the 13 sterol concentrations picked by the feature selection algorithm were individually correlated with QIDS-SR₁₆ scores. Concentrations of desmosterol, 5,6β-epoxycholesterol, and 25-hydroxycholesterol were significantly correlated with QIDS-SR₁₆ scores after Holm's correction for multiple comparisons, while adjusted p values for cholestanol and 14-desmethyl lanosterol were close to significance (Table 6). For each of these 13 sterols, we also calculated odds ratios for having a QIDS-SR₁₆ score greater than 10.5 by setting a threshold at either the 5th, or the 95th percentile of sterol concentrations based on the sign of the correlation coefficient. Concentrations of 7-DHC above the 95th percentile (5.4 ng/mL, OR 2.4, 95% CI 1.6–3.6, adjusted-p < 0.001) and desmosterol concentrations below the 5th percentile (1.9 ng/mL, OR 1.9, 95% CI 1.2-2.9, adjusted-p 0.053) were associated with depressive symptoms of at least moderate severity, albeit the multiple hypothesis testing corrected p value for desmosterol was on the margin of significance. An association between higher cholestanol concentrations and depressive symptom severity was not statistically significant following correction (Table 6). Fig 4 depicts the correlations between QIDS-SR₁₆ scores and 7-DHC (panel a) and desmosterol (panel b) concentrations.

Discussion

The contribution of derangements in sterol metabolism to depression is unknown. In this study, we used a machine learning approach to investigate whether plasma sterols concentrations are associated with depressive symptoms in an epidemiological cohort. We found novel associations between several sterols and depressive symptom severity.



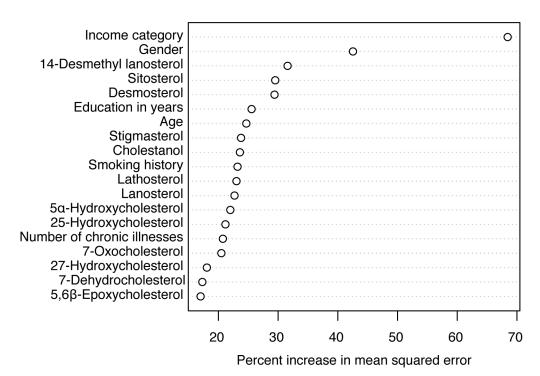


Fig 3. Nineteen variables comprising Model F. Parameters selected from a pool of 43 demographic or general health variables and sterol concentrations by a feature selection algorithm based on explained variation in QIDS-SR16 scores is shown, ranked by importance judged by the percent increase in mean squared error metric.

Desmosterol and 7-DHC are the immediate precursors in the two main cholesterol synthesis pathways in mammals: 7-DHC in the Kandutsch-Russell pathway and desmosterol in the Bloch pathway (Fig 5). In the adult mammalian brain, the Kandutsch-Russell pathway is predominant in neurons, while the Bloch pathway is predominant in astrocytes. Cholesterol is thought to be synthesized largely in astrocytes and transported to the neurons via apolipoprotein E [34]. We found that increased 7-DHC (precursor in the neuronal pathway) and decreased desmosterol (precursor in the glial pathway) are associated with higher depressive symptom severity scores, suggesting an imbalance of cholesterol synthesis via the Kandutsch-

Table 6. Odds ratios and correlation coefficients for six of the 13 sterols picked by feature selection.

	Odds of QIDS-SR ₁₆ score ≥ 10.5				Correlation with QIDS-SR ₁₆ Score	
Sterol	Cut-off	OR	95% CI ¹	p-adjusted ¹	Spearman's rho	p-adjusted1
7-Dehydrocholesterol	> 5.4 ng/mL ²	2.4	1.6–3.6	<0.001	0.03	0.639
Desmosterol	< 1.9 ng/mL ³	1.9	1.2-2.9	0.053	-0.11	<0.001
Cholestanol	> 23.5 ng/mL ²	1.6	1.0-2.5	0.316	0.05	0.058
5,6β-Epoxycholesterol	> 2.7 ng/mL ²	1	0.6–1.7	1	0.05	0.028
25-Hydroxycholesterol	> 0.1 ng/mL ²	1.1	0.7–1.8	1	0.06	0.006
14-Desmethyl lanosterol	< 1.2 ng/mL ³	1.3	0.8–2.1	1	-0.05	0.064

¹, p values are corrected for 13 comparisons using Holm's method. Confidence intervals are based on Fisher's exact test.

OR, odds ratio; CI, confidence interval

², 95th percentile

³, 5th percentile



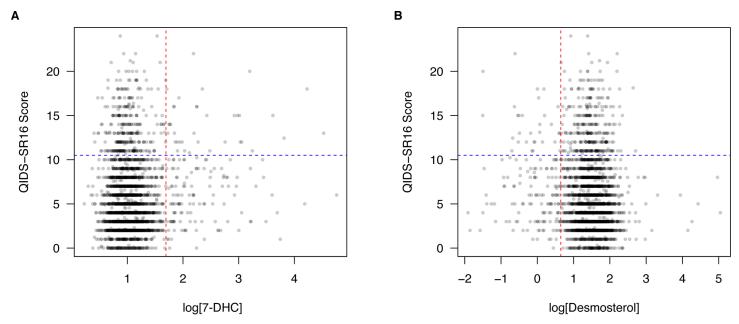


Fig 4. Correlation of QIDS-SR₁₆ scores with 7-DHC and desmosterol concentrations. (a), 7-DHC; (b), desmosterol. Blue dashed line indicates QIDS-SR₁₆ score of 10.5, red dashed line indicates 95th percentile for 7-DHC and 5th percentile for desmosterol concentrations. Each gray circle indicates one data point.

Russell versus Bloch pathways. Additionally, an earlier precursor in the Bloch pathway, 14-desmethyl lanosterol, also trended towards a negative correlation with depression scores, similar to desmosterol concentrations, further strengthening the hypothesis that decreased concentrations of the Bloch pathway intermediaries correlate with higher depressive scores. Accumulation of 7-DHC is the hallmark of Smith-Lemli-Opitz Syndrome (SLOS), an autosomal recessive disorder of cholesterol synthesis that results from mutations in the gene coding for the enzyme, 7-DHC reductase, and manifests with aggressive and self-injurious behaviors [35].

Vitamin D is synthesized from 7-DHC in the human skin in a reaction that requires UV light [36]. Vitamin D deficiency has been associated with depression [17]. We included 25-Hydroxyvitamins D_2 and D_3 in the modelling. While 25-Hydroxyvitamin D_3 was one of the predictive variables in Model S, it was not picked by feature selection when we included other demographic and health related variables.

Low cholesterol levels have long been associated with suicidal behavior [5, 6, 10]. In one study, low post-mortem brain cholesterol was associated with violent suicides [11]. An association between low plasma [7, 9] or brain [4] cholesterol and depression has also been reported. The underlying mechanism remains unknown. One hypothesis is that the link between brain cholesterol and impulsive-suicidal behavior is modulation of neurotransmitter signaling by cholesterol in lipid rafts, membrane microdomains rich in cholesterol [37, 38]: Key neurotransmitter receptors including 5-HT_{1A} , 5-HT_{2} , and D_{1} localize to lipid rafts and cholesterol content of neural membranes and localization of receptors to lipid rafts may modulate the function of these receptors [39]. Singh et al., [39] have suggested that either cholesterol or desmosterol is required for optimal ligand binding to the 5-HT_{1A} receptor. Similarly, lipid rafts may be necessary for the brain derived neurotropic factor (BDNF) signaling on nerve growth cones [40], intriguing given BDNF's well-known association with depression [41]. Another potential link between cholesterol metabolism and brain function is neurosterols and oxysterols, both of which are synthesized from cholesterol [42]. Oxysterols, as well as 7-DHC,



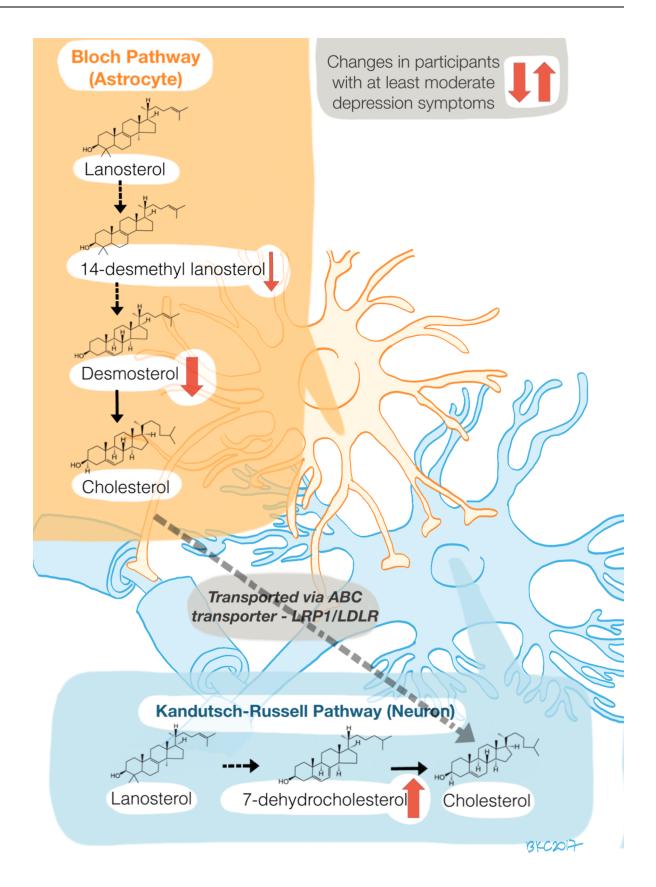




Fig 5. Hypothetical model of sterol derangements in depression. Arrows summarize main findings of the study (Table 6): Desmosterol concentrations are negatively correlated with depressive symptoms and low (<5th percentile) desmosterol concentrations predict moderate to severe depressive symptoms. 7-DHC concentrations above the 95th percentile also predict moderate to severe depressive symptoms. 14-Desmethyl lanosterol concentrations trend towards a negative correlation with depressive symptoms (not statistically significant after correction of *p* values for multiple comparisons).

modulate hedgehog signaling, a signaling pathway involved in brain development and maintenance of hippocampal neurogenesis and may be modulated by antidepressants and antipsychotics [43]. It remains to be elucidated whether modulation of this pathway by cholesterol content of brain membranes plays a role in depression.

The changes in sterol concentrations that we correlated with depressive symptoms may be either state or trait markers. While it is not possible to rule out either based on a cross-sectional study, the latter might be more likely. Plasma sterol measurements are unlikely to directly reflect brain cholesterol metabolism, given that cholesterol does not ordinarily pass the bloodbrain barrier [44]. The 7-DHC/desmosterol bio-signature probably hints at a metabolic disposition, perhaps genetically determined, that predisposes to depression. Indeed, in their analysis of this same DHS cohort, Stiles et al. identified genetic loci significantly linked to desmosterol and 7-DHC levels [13]. Future studies may explore the potential connection between these genetic loci and depression.

This study has several limitations. Firstly, this was an unbiased, hypothesis-generating analysis of epidemiological level data. We did not intend to develop a clinically useful biomarker panel at this stage and the population we studied was not a clinical sample. Rather, we have discovered previously unknown associations between several sterols and depression that may, in the future, contribute to development of such a panel, but also is hinting at novel biology on its own. Given the cross-sectional nature of our study, the mechanistic and diagnostic/clinical impact of our findings will have to be clarified by future clinical and experimental studies. Furthermore, correlational results from a cross-sectional study obviously do not prove causality, hence the changes we observed in sterol concentrations could be a result of depression, changes in diet, or medications.

Several psychotropic medications have been shown to alter 7-DHC levels and metabolism. Most relevant to our findings, Korade et al., recently reported that the use of aripiprazole, haloperidol and trazodone in psychiatric patients and controls is correlated with increased 7-DHC levels in plasma [45]. Aripiprazole is a second generation antipsychotic that is sometimes used as an augmentation agent in depression resistant to first line therapies. Trazodone, a selective serotonin reuptake inhibitor at high doses, is rarely used as an antidepressant currently, but is commonly prescribed as a sleep aid. Therefore, the elevated 7-DHC concentrations we observed in depressed participants could be explained by this confounder. On the other hand, Lauth et al., had reported that imipramine, clozapine, haloperidol and chlorpromazine induce the expression of 7-DHC reductase, the enzyme that converts 7-DHC to cholesterol, leading to modulation of hedgehog signaling [43]. Another study reported sterol regulatory elementbinding protein induction by antidepressants [46]. It remains to be determined whether the effects of psychotropic medications on lipid metabolism are pleiotropic or related to their therapeutic effects. It would be interesting to test whether sterol concentrations predict response to different antidepressants in future studies. Finally, peripheral measurements of biomarkers, while advantageous in terms of feasibility and acceptability, are nevertheless one step removed from concentrations in the brain.

In summary, our study underscores the promise of unbiased machine learning approaches to biomarker discovery in depression. We identified novel leads for biomarker development in depression, under-studied sterols that are independently associated with depressive symptoms



and, when deranged, confer risk for depressive symptoms of at least moderate severity in a large epidemiological cohort. Our findings suggest that an imbalance between glial versus neuronal cholesterol synthesis may be implicated in depression. Further studies may expand the role of lipid biosynthetic pathways in psychiatric illness from the known association of rare genetic defects causing severe behavioral disturbances to potential determinants of population-level risk for moderate depression. Future prospective studies of these sterols as depression biomarkers, as well as basic science investigations into their biology and genetic determinants seem warranted. Particularly, it would be very informative to conduct prospective studies investigating the change in plasma sterol concentrations versus depressive scores over time.

Supporting information

S1 Fig. Distribution of QIDS scores. (TIF)

Acknowledgments

Fig 5 was drawn by Dr. Bercin Kutluk Cenik.

Author Contributions

Conceptualization: Basar Cenik, E. Sherwood Brown.

Formal analysis: Basar Cenik, Can Cenik.

Investigation: Basar Cenik, Can Cenik.

Supervision: Michael P. Snyder, E. Sherwood Brown.

Writing - original draft: Basar Cenik.

Writing - review & editing: Basar Cenik, Can Cenik.

References

- Moussavi S, Chatterji S, Verdes E, Tandon A, Patel V, Ustun B. Depression, chronic diseases, and decrements in health: results from the World Health Surveys. Lancet. 2007; 370(9590):851–8. https://doi.org/10.1016/S0140-6736(07)61415-9 PMID: 17826170.
- 2. Hasler G. Pathophysiology of depression: do we have any solid evidence of interest to clinicians? World Psychiatry. 2010; 9(3):155–61. PMID: 20975857; PubMed Central PMCID: PMCPMC2950973.
- Huang TL, Lin CC. Advances in biomarkers of major depressive disorder. Advances in clinical chemistry. 2015; 68:177–204. https://doi.org/10.1016/bs.acc.2014.11.003 PMID: 25858873.
- Beasley CL, Honer WG, Bergmann K, Falkai P, Lutjohann D, Bayer TA. Reductions in cholesterol and synaptic markers in association cortex in mood disorders. Bipolar disorders. 2005; 7(5):449–55. https://doi.org/10.1111/j.1399-5618.2005.00239.x PMID: 16176438.
- Boscarino JA, Erlich PM, Hoffman SN. Low serum cholesterol and external-cause mortality: potential implications for research and surveillance. Journal of psychiatric research. 2009; 43(9):848–54. https://doi.org/10.1016/j.jpsychires.2008.11.007 PMID: 19135214.
- Fiedorowicz JG, Coryell WH. Cholesterol and suicide attempts: a prospective study of depressed inpatients. Psychiatry research. 2007; 152(1):11–20. https://doi.org/10.1016/j.psychres.2006.09.003 PMID: 17360043; PubMed Central PMCID: PMC1965495.
- Morgan RE, Palinkas LA, Barrett-Connor EL, Wingard DL. Plasma cholesterol and depressive symptoms in older men. Lancet. 1993; 341(8837):75–9. PMID: 8093404.
- Partonen T, Haukka J, Virtamo J, Taylor PR, Lonnqvist J. Association of low serum total cholesterol with major depression and suicide. The British journal of psychiatry: the journal of mental science. 1999; 175:259

 –62. PMID: 10645328.



- Steegmans PH, Hoes AW, Bak AA, van der Does E, Grobbee DE. Higher prevalence of depressive symptoms in middle-aged men with low serum cholesterol levels. Psychosomatic medicine. 2000; 62 (2):205–11. PMID: 10772398.
- Wu S, Ding Y, Wu F, Xie G, Hou J, Mao P. Serum lipid levels and suicidality: a meta-analysis of 65 epidemiological studies. Journal of psychiatry & neuroscience: JPN. 2016; 41(1):56–69. https://doi.org/10.1503/jpn.150079 PMID: 26505144; PubMed Central PMCID: PMC4688029.
- Lalovic A, Levy E, Luheshi G, Canetti L, Grenier E, Sequeira A, et al. Cholesterol content in brains of suicide completers. The international journal of neuropsychopharmacology / official scientific journal of the Collegium Internationale Neuropsychopharmacologicum. 2007; 10(2):159–66. https://doi.org/10. 1017/S1461145706006663 PMID: 16707033.
- Quehenberger O, Armando AM, Brown AH, Milne SB, Myers DS, Merrill AH, et al. Lipidomics reveals a remarkable diversity of lipids in human plasma. Journal of lipid research. 2010; 51(11):3299–305. https://doi.org/10.1194/jlr.M009449 PMID: 20671299; PubMed Central PMCID: PMC2952570.
- 13. Stiles AR, Kozlitina J, Thompson BM, McDonald JG, King KS, Russell DW. Genetic, anatomic, and clinical determinants of human serum sterol and vitamin D levels. Proceedings of the National Academy of Sciences of the United States of America. 2014; 111(38):E4006–14. https://doi.org/10.1073/pnas. 1413561111 PMID: 25201972; PubMed Central PMCID: PMC4183318.
- Javitt NB. Oxysterols: novel biologic roles for the 21st century. Steroids. 2008; 73(2):149–57. https://doi.org/10.1016/j.steroids.2007.10.004 PMID: 18068744.
- Linsenbardt AJ, Taylor A, Emnett CM, Doherty JJ, Krishnan K, Covey DF, et al. Different oxysterols have opposing actions at N-methyl-D-aspartate receptors. Neuropharmacology. 2014; 85:232–42. https://doi.org/10.1016/j.neuropharm.2014.05.027 PMID: 24878244; PubMed Central PMCID: PMC4107067.
- Paul SM, Doherty JJ, Robichaud AJ, Belfort GM, Chow BY, Hammond RS, et al. The major brain cholesterol metabolite 24(S)-hydroxycholesterol is a potent allosteric modulator of N-methyl-D-aspartate receptors. The Journal of neuroscience: the official journal of the Society for Neuroscience. 2013; 33 (44):17290–300. https://doi.org/10.1523/JNEUROSCI.2619-13.2013 PMID: 24174662; PubMed Central PMCID: PMC3812502.
- Anglin RE, Samaan Z, Walter SD, McDonald SD. Vitamin D deficiency and depression in adults: systematic review and meta-analysis. The British journal of psychiatry: the journal of mental science. 2013; 202:100–7. https://doi.org/10.1192/bjp.bp.111.106666 PMID: 23377209.
- Victor RG, Haley RW, Willett DL, Peshock RM, Vaeth PC, Leonard D, et al. The Dallas Heart Study: a
 population-based probability sample for the multidisciplinary study of ethnic differences in cardiovascular health. The American journal of cardiology. 2004; 93(12):1473

 –80. https://doi.org/10.1016/j.
 amjcard.2004.02.058 PMID: 15194016.
- McDonald JG, Smith DD, Stiles AR, Russell DW. A comprehensive method for extraction and quantitative analysis of sterols and secosteroids from human plasma. Journal of lipid research. 2012; 53

 (7):1399–409. https://doi.org/10.1194/jir.D022285 PMID: 22517925; PubMed Central PMCID: PMC3371252
- Vega GL, Adams-Huet B, Peshock R, Willett D, Shah B, Grundy SM. Influence of body fat content and distribution on variation in metabolic risk. The Journal of clinical endocrinology and metabolism. 2006; 91(11):4459–66. https://doi.org/10.1210/jc.2006-0814 PMID: 16926254.
- Rush AJ, Trivedi MH, Ibrahim HM, Carmody TJ, Arnow B, Klein DN, et al. The 16-Item Quick Inventory
 of Depressive Symptomatology (QIDS), clinician rating (QIDS-C), and self-report (QIDS-SR): a psychometric evaluation in patients with chronic major depression. Biological psychiatry. 2003; 54(5):573

 –83.
 PMID: 12946886.
- Rush AJ, Gullion CM, Basco MR, Jarrett RB, Trivedi MH. The Inventory of Depressive Symptomatology (IDS): psychometric properties. Psychol Med. 1996; 26(3):477–86. PMID: 8733206.
- 23. Trivedi MH, Rush AJ, Ibrahim HM, Carmody TJ, Biggs MM, Suppes T, et al. The Inventory of Depressive Symptomatology, Clinician Rating (IDS-C) and Self-Report (IDS-SR), and the Quick Inventory of Depressive Symptomatology, Clinician Rating (QIDS-C) and Self-Report (QIDS-SR) in public sector patients with mood disorders: a psychometric evaluation. Psychol Med. 2004; 34(1):73–82. PMID: 14971628
- Reilly TJ, MacGillivray SA, Reid IC, Cameron IM. Psychometric properties of the 16-item Quick Inventory of Depressive Symptomatology: a systematic review and meta-analysis. Journal of psychiatric research. 2015; 60:132–40. https://doi.org/10.1016/j.jpsychires.2014.09.008 PMID: 25300442.
- Rush AJ, Fava M, Wisniewski SR, Lavori PW, Trivedi MH, Sackeim HA, et al. Sequenced treatment alternatives to relieve depression (STAR*D): rationale and design. Control Clin Trials. 2004; 25(1):119– 42. PMID: 15061154.



- **26.** Hastie T, Tibshirani R, Narasimhan B, Chu G. impute: Imputation for microarray data. R package version 1.46.0. 2016.
- Breiman L. Random Forests. Machine Learning. 2001; 45(1):5–32. https://doi.org/10.1023/a:1010933404324
- 28. Liaw A, Wiener M. Classification and Regression by randomForest. R News. 2002; 2(3):18–22.
- Kuhn M. caret: Classification and Regression Training. R package version 6.0–70. https://CRAN.R-project.org/package=caret. 2016.
- Revolution Analytics, Weston S. doMC: Foreach Parallel Adaptor for 'parallel'. R package version 1.3.4. https://CRAN.R-project.org/package=doMC. 2015.
- Weaver A, Himle JA, Taylor RJ, Matusko NN, Abelson JM. Urban vs Rural Residence and the Prevalence of Depression and Mood Disorder Among African American Women and Non-Hispanic White Women. JAMA psychiatry. 2015; 72(6):576–83. https://doi.org/10.1001/jamapsychiatry.2015.10 PMID: 25853939; PubMed Central PMCID: PMC4456259.
- Hasin DS, Goodwin RD, Stinson FS, Grant BF. Epidemiology of major depressive disorder: results from the National Epidemiologic Survey on Alcoholism and Related Conditions. Archives of general psychiatry. 2005; 62(10):1097–106. https://doi.org/10.1001/archpsyc.62.10.1097 PMID: 16203955.
- Patten SB, Wang JL, Williams JV, Currie S, Beck CA, Maxwell CJ, et al. Descriptive epidemiology of major depression in Canada. Canadian journal of psychiatry Revue canadienne de psychiatrie. 2006; 51(2):84–90. https://doi.org/10.1177/070674370605100204 PMID: 16989107.
- Zhang J, Liu Q. Cholesterol metabolism and homeostasis in the brain. Protein & cell. 2015; 6(4):254–64. https://doi.org/10.1007/s13238-014-0131-3 PMID: 25682154; PubMed Central PMCID: PMC4383754.
- Nowaczyk MJ, Irons MB. Smith-Lemli-Opitz syndrome: phenotype, natural history, and epidemiology. American journal of medical genetics Part C, Seminars in medical genetics. 2012; 160C(4):250–62. https://doi.org/10.1002/ajmg.c.31343 PMID: 23059950.
- Bikle DD. Vitamin D metabolism, mechanism of action, and clinical applications. Chem Biol. 2014; 21 (3):319–29. https://doi.org/10.1016/j.chembiol.2013.12.016 PMID: 24529992; PubMed Central PMCID: PMCPMC3968073.
- Cantarelli Mda G, Tramontina AC, Leite MC, Goncalves CA. Potential neurochemical links between cholesterol and suicidal behavior. Psychiatry research. 2014; 220(3):745–51. https://doi.org/10.1016/j. psychres.2014.10.017 PMID: 25457283.
- Tsui-Pierchala BA, Encinas M, Milbrandt J, Johnson EM Jr. Lipid rafts in neuronal signaling and function. Trends in neurosciences. 2002; 25(8):412–7. PMID: 12127758.
- Singh P, Jafurulla M, Paila YD, Chattopadhyay A. Desmosterol replaces cholesterol for ligand binding function of the serotonin(1A) receptor in solubilized hippocampal membranes: support for nonannular binding sites for cholesterol? Biochimica et biophysica acta. 2011; 1808(10):2428–34. https://doi.org/ 10.1016/j.bbamem.2011.06.022 PMID: 21763272.
- **40.** Guirland C, Suzuki S, Kojima M, Lu B, Zheng JQ. Lipid rafts mediate chemotropic guidance of nerve growth cones. Neuron. 2004; 42(1):51–62. PMID: 15066264.
- Bjorkholm C, Monteggia LM. BDNF—a key transducer of antidepressant effects. Neuropharmacology. 2016; 102:72–9. https://doi.org/10.1016/j.neuropharm.2015.10.034 PMID: 26519901; PubMed Central PMCID: PMC4763983.
- Dubrovsky BO. Steroids, neuroactive steroids and neurosteroids in psychopathology. Progress in neuro-psychopharmacology & biological psychiatry. 2005; 29(2):169–92. https://doi.org/10.1016/j. pnpbp.2004.11.001 PMID: 15694225.
- Lauth M, Rohnalter V, Bergstrom A, Kooshesh M, Svenningsson P, Toftgard R. Antipsychotic drugs regulate hedgehog signaling by modulation of 7-dehydrocholesterol reductase levels. Molecular pharmacology. 2010; 78(3):486–96. https://doi.org/10.1124/mol.110.066431 PMID: 20558592.
- Wellington CL, Frikke-Schmidt R. Relation between plasma and brain lipids. Current opinion in lipidology. 2016; 27(3):225–32. https://doi.org/10.1097/MOL.000000000000291 PMID: 27149391.
- Korade Z, Liu W, Warren EB, Armstrong K, Porter NA, Konradi C. Effect of psychotropic drug treatment on sterol metabolism. Schizophr Res. 2017. https://doi.org/10.1016/j.schres.2017.02.001 PMID: 28202290.
- 46. Raeder MB, Ferno J, Glambek M, Stansberg C, Steen VM. Antidepressant drugs activate SREBP and up-regulate cholesterol and fatty acid biosynthesis in human glial cells. Neuroscience letters. 2006; 395 (3):185–90. https://doi.org/10.1016/j.neulet.2005.10.096 PMID: 16324787.