RESEARCH ARTICLE

REVISED Predicting lithium treatment response in bipolar patients using gender-specific gene expression biomarkers and machine learning [version 2; peer review: 1 approved, 1 approved with reservations]

Andy R. Eugene ¹, Jolanta Masiak², Beata Eugene³

¹Department of Pharmacogenomics, Bernard J. Dunn School of Pharmacy, Inova Center for Personalized Health, Shenandoah University, Fairfax, VA, 22031, USA

²Independent Neurophysiology Laboratory, Department of Psychiatry, Medical University of Lublin, Lublin, 20-439, Poland

³Marie-Curie Skłodowska University, Lublin, 20-400, Poland

v2 First published: 18 Apr 2018, 7:474
<https://doi.org/10.12688/f1000research.14451.1>

Second version: 29 May 2018, 7:474
<https://doi.org/10.12688/f1000research.14451.2>

Latest published: 07 Dec 2018, 7:474
<https://doi.org/10.12688/f1000research.14451.3>

Abstract

Background: We sought to test the hypothesis that transcriptome-level genes signatures are differentially expressed between male and female bipolar patients, prior to lithium treatment, in a patient cohort who later were clinically classified as lithium treatment responders.

Methods: Gene expression study data was obtained from the Lithium Treatment-Moderate dose Use Study data accessed from the National Center for Biotechnology Information's Gene Expression Omnibus via accession number GSE4548. Differential gene expression analysis was conducted using the Linear Models for Microarray and RNA-Seq (limma) package and the Random Forests machine learning algorithm in R.

Results: In pre-treatment lithium responders, the following genes were found having a greater than 0.5 fold-change, and differentially expressed indicating a male bias: RBPM52, SIDT2, CDH23, LILRA5, and KIR2DS5; while the female-biased genes were: HLA-H, RPS23, FHL3, RPL10A, NBPF14, PSTPIP2, FAM117B, CHST7, and ABRACL.

Conclusions: Using machine learning, we developed a pre-treatment gender- and gene-expression-based predictive model selective for lithium responders with an ROC AUC of 0.92 for men and an ROC AUC of 1 for women.

Keywords

lithium, treatment response, gene expression, machine learning, microarray, transcriptome, precision medicine, pharmacogenomics

Open Peer Review

Approval Status

	1	2	3
version 3 (revision) 07 Dec 2018		 view	 view
version 2 (revision) 29 May 2018	 ↑	 view	
version 1 18 Apr 2018	 view		

1. **Ming-Fen Ho**, Mayo Clinic, Rochester, USA
2. **Sunil V. Kalmady** , University of Alberta, Alberta, Canada
Alberta Machine Intelligence Institute, Alberta, Canada
3. **Duan Liu** , Alberta Machine Intelligence Institute, Alberta, Canada

Any reports and responses or comments on the article can be found at the end of the article.

This article is included in the **Artificial Intelligence and Machine Learning** gateway.

Corresponding author: Andy R. Eugene (andyeugene.md@gmail.com)

Author roles: **Eugene AR:** Conceptualization, Data Curation, Formal Analysis, Investigation, Methodology, Project Administration, Resources, Software, Supervision, Validation, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Masiak J:** Project Administration, Writing – Review & Editing; **Eugene B:** Formal Analysis, Investigation, Methodology, Project Administration, Writing – Original Draft Preparation, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

Grant information: The author(s) declared that no grants were involved in supporting this work.

Copyright: © 2018 Eugene AR *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 work is properly cited.

How to cite this article: Eugene AR, Masiak J and Eugene B. **Predicting lithium treatment response in bipolar patients using gender-specific gene expression biomarkers and machine learning [version 2; peer review: 1 approved, 1 approved with reservations]** F1000Research 2018, 7:474 <https://doi.org/10.12688/f1000research.14451.2>

First published: 18 Apr 2018, 7:474 <https://doi.org/10.12688/f1000research.14451.1>

REVISED Amendments from Version 1

- For all differentially expressed results reported throughout tables within the manuscript, I changed the wording from genes up-regulated or down-regulated in males or females to a clearer description statement that of genes-associated with males or females.
- For the caption in [Figure 1](#), I updated the sample sizes for males and female patients. The updated [Figure 1](#) text reads: Males (n=20; with 40 pre- and post-treatment samples) and Females (n=40; with 80 pre- and post-treatment samples).
- I also updated the text to explain the variable importance plot shown in [Figure 2](#).
- Further, I added a more elaborative rationale in the study design found in the methods for the top 250 genes in the analysis.
- Lastly, I added the American Board of Clinical Pharmacology as opposed to simply the board-certified/eligible pharmacologist in the text.

See referee reports

Introduction

Lithium is the most well-established mood-stabilizer in the practice of psychiatry ([Jermain *et al.*, 1991](#); [Landersdorfer *et al.*, 2017](#)). A recent propensity-score adjusted and matched longitudinal cohort-study evaluating the effectiveness of the newer mood stabilizers: olanzapine (n=1477), quetiapine (n=1376), and valproate (n=1670), in comparison to lithium (n=2148), found that patients treated with lithium experienced reduced rates of both unintentional injury and self-harm ([Hayes *et al.*, 2016](#)). However, due to lithium's narrow window, 0.5–1.2 mEq/mL, of maximal effectiveness and safety (i.e. therapeutic index), Therapeutic Drug Monitoring is the standard-of-care to ensure patient safety in medical practice ([Hiemke *et al.*, 2011](#)). Further, divergent clinical response rates have been reported among male and female patients diagnosed with bipolar disorder and treated with lithium ([Viguera *et al.*, 2000](#)).

In a 1986, Zetin and colleagues published the results of a study that evaluated four methods for predicting lithium daily dosages, and the final equation resulted in a 147.8mg/day increased dosage-adjustment for male patients ([Zetin *et al.*, 1986](#)). Similarly, a later study by Lobeck and colleagues corroborated the 147.8 mg/day male increase dose requirement for the lithium maintenance dose in bipolar patients ([Lobeck *et al.*, 1987](#)). However, neither do the current dosing guidelines recommend a gender-based dose adjustment via clinical pharmacometrics, to avoid toxicity, nor are gender-specific gene expression screening panels available to predict lithium efficacy currently available and implemented.

A recent large-scale meta-analysis of human body-tissue gene expression reported that the body organ with the most abundant gender-biased gene expression is the anterior cingulate cortex within the frontal cortex of the brain ([Mayne *et al.*, 2016](#)). Thus, these findings suggest that therapeutic drug response may be influenced not only via drug absorption, distribution, metabolism, and elimination, but also within the underlying gene signatures

across the human transcriptome and mechanisms of gene-gene interactions that regulate physiology. Beech and colleagues conducted a study to identify gene expression differences from the peripheral blood in patients classified as lithium responders and non-responders ([Beech *et al.*, 2014](#)). However, the study reported that no significant gender-biased gene expression differences were found (p-value=0.941) in patients who were randomized to optimal therapy (control), defined as one FDA-approved mood stabilizer, versus patients treated with lithium plus optimal therapy ([Beech *et al.*, 2014](#)). Despite these initially reported findings, a recent study by Labonté and colleagues, which used RNA-Seq to evaluate the transcriptome in patients diagnosed with major depressive disorder (MDD), concluded that gender dimorphism exists at the transcriptome-level in MDD patients and that gender-specific treatments should be investigated ([Labonté *et al.*, 2017](#)).

Therefore, there is an urgent clinical need to improve behavioral healthcare by understanding gene expression variability that may lead to personalizing medicine in patients with mania. These findings may improve prediction of clinical drug response of lithium prior to initiating pharmacotherapy in patients with bipolar or schizoaffective disorders, who cannot risk drug inefficacy for obvious safety reasons. Therefore, the overall aim for our study is to define gender-specific transcriptional-level regulators of lithium treatment response that may influence treatment of bipolar or schizoaffective disorders. We will test the hypothesis that biologically plausible gene expression differences exist, prior to lithium treatment, in patients diagnosed with bipolar disorder in the following three patient subgroups: (1) male and female patients who were later clinically classified as lithium treatment responders; (2) male-responders versus male-non-responders; (3) female-responders versus female-non-responders.

Methods

Data

DNA microarray data analyzed in this study are originally referenced from the Lithium Treatment-Moderate dose Use Study placed in the National Center for Biotechnology Information (NCBI) Gene Expression Omnibus (GEO) via accession number [GSE45484](#) with the Illumina HumanHT12 V4.0 expression Beadchip GPL10558 platform file to associate gene names and descriptions. The original multisite clinical study recruited patients from Case Western Reserve University, Massachusetts General Hospital, Stanford University, Yale University, and the Universities of: Pittsburgh, Texas Health Science Center at San Antonio, and Pennsylvania ([Beech *et al.*, 2014](#)). From the original 120 peripheral blood samples used to generate probe and gene expression profiles, from patients diagnosed with bipolar disorder, the clinical phenotype of being either a treatment- responder or non-responder was assessed using the Clinical Global Impression Scale for Bipolar Disorder-Severity (CGI-BP-S) ([Spearing *et al.*, 1997](#)).

Study design

To assess for gender-specific differential gene signatures, in our first analysis we grouped patients based on gender alone and not on any other variables (i.e. optimal treatment versus lithium, or responder versus non-responder status). Then, we rationalized that from the results of the gender-specific transcriptome signatures from our

first analysis, we will set the top two-hundred and fifty genes as controls in an effort to identify pharmacologic treatment-response transcriptome biomarkers that are not directly linked to the X or Y chromosome. Therefore, we excluded the top two-hundred and fifty genes from all results that were reported in subsequent analyses to identify genes with lithium-specific transcriptional differences between genders associated with response treatment. In our second analysis, we only selected patients who were classified as lithium treatment-responders, at baseline, and the results from the gene expression differences are reported excluding the sex-specific control genes identified in the first experiment. In our third and fourth analyses, we compared: male-responders vs. male non-responders, and female-responders vs. female non-responders, respectively.

Gene expression analysis

Differential gene expression analysis of the microarray data was conducted using the Empirical Bayes method implemented within the *limma* package (version 3.34.5) and utilizes the Biobase package (version 2.38.0) which both run within the R for Statistical Programming environment (version 3.4.3; R Foundation for Statistical Computing, Vienna, Austria) (Ritchie *et al.*, 2015; Team, 2013). Due to multiple testing of the peripheral blood transcriptome, the False-Discovery Rate was adjusted using the Benjamini-Hochberg method. The Decision Tree and Random Forests machine learning algorithms were used to assess gender using transcriptional signatures and for predictive modeling using the discovered microarray genes to select for the gender-specific lithium responders. A p-value of less 0.05 was considered to be statistically significant and a differential gene expression threshold of 0.5 was used and reported during the machine learning process. Further methods detailing the Random Forests decision processes for male- and female-responders are located in [Supplementary File 1](#).

Results

Table 1 provides the patient age and sample sizes used during subgroup analyses. In our first analysis, which aimed to group patients based on gender alone and not based on clinical

variables detailed in the original study, data-driven gene analytics identified four female-labeled patient samples with gene expression levels similar to that found in male patients for the following Y-chromosome genes: *RPS4Y1*, *EIF1AY*, *KDM5D*, *RPS4Y2*; and the *XIST* gene located on the X-chromosome. Therefore, all subsequent hypothesis-testing were analyzed with the updated male-gender classification for the following NCBI GEO patient samples: GSM1105526 (baseline lithium-non-responder), GSM1105528 (1-month lithium-non-responder), GSM1105546 (baseline lithium-non-responder), and GSM1105548 (1-month lithium-non-responder). **Figure 1** illustrates the gene expression

Table 1. Patient age and sample sizes used during subgroup analyses.

Lithium treated patient population			
Baseline	Mean age	S.D.	Sample size (n)
Male-responder	36	8.1	3
Female-responder	31	11.8	6
Male-non-responder	40	10	7
Female-non-responder	44	9.2	12
*General mood stabilizers patient population			
Baseline	Mean age	S.D.	Sample size (n)
Male-responder	51	--	1
Female-responder	49	10.5	3
Male-non-responder	43	12.5	9
Female-non-responder	37	14.5	19
Total patient population			
Gender	Mean age	S.D.	Sample size (n)
Male	41	10.8	20
Female	39	13.1	40
Study population	40	12.3	60

Note: United States Food and Drug Administration approved Mood Stabilizers.

Figure 1. Gene expression levels for the Ribosomal protein S4, Y-linked 1 (RPS4Y1) gene illustrating 4 patient samples as labeled as female and were re-assigned to the male patient gender group. Males (n=20; with 40 pre- and post-treatment samples) and Females (n=40; with 80 pre- and post-treatment samples).

findings resulting in re-assignment for the aforementioned patient samples from females to males using a decision-tree approach that evaluated if the *RPS4Y1* gene had an expression level of greater than or equal to 9.6 resulting in: yes=male (31%) and no=female (69%). After proceeding with the machine learning analysis of both the 'training' and 'validation' datasets, the final 'test' dataset resulted in the following diagnostic test evaluation parameters: Sensitivity=100% (95% C.I. 66.37%-100.00%), Specificity=100% (95% C.I. 78.20%-100.00%), and an area under the receiver operator characteristic (ROC) curve of 1. **Figure 2** illustrates the variable importance plots used in the machine learning process for classifying patients as being a male-lithium-responder or female-lithium-responder relative to the full patient population. The results show, in descending order of predictive power, the genes selective for male lithium-responders versus the full patient population being *RBPM2S*, *CDH23*, and *SIDT2*. Similarly, in descending order of predictive power, for female lithium-responders versus the entire patient population, the *FHL3*, *ABRACL*, *RPL10A*, and *RPS23* genes are most selective.

Table 2 provides the results for the gender-specific differentially expressed genes from the entire study population using a fold-change (FC) threshold of 0.5. A total of five genes met the *a priori* FC requirements and were found to be *RPS4Y1*, *EIF1AY*, *KDM5D*, *RPS4Y2*, and *EIF1AY*. These five down-regulated male-biased genes were all found on the Y-chromosome. Contrastingly, a total of 10 upregulated female-biased genes were found to be: *XIST*, *S100P*, *IFIT3*, *TNFAIP6*, *IFITM3*, *IFIT2*, *CHURC1*, *ANXA3*, *ADM*, and *PROK2*. The *RPS4Y1* gene in males (FC= -4.9807, p=7.36E-47) and the *XIST* gene (FC=1.7615, p=2.98E-36), found on the X-chromosome, in females resulted in the greatest expression changes between genders. The male-favored genes resulted in a larger expression change than compared to the females.

Table 3 provides the results for the differentially expressed genes that were found between male and female responders prior to initiation of lithium and optimal therapy, meeting the FC criteria of at least 0.5. In male lithium responders, we found 5 differentially expressed while the RNA binding protein with multiple splicing 2 (*RBPM2S*) gene ranked with the greatest FC of -1.351 (unadjusted p=0.00111). Whereas, 9 genes were associated with female lithium responders, with greatest expression change being the major histocompatibility complex class-1-H (HLA-H) at 1.602 (unadjusted p-value=0.00099). The neuroblastoma breakpoint family member-14 (*NBPF14*) gene met the Benjamini-Hochberg adjusted p-value criteria and resulted with an expression change of 0.586 (adjusted p=0.0462). **Figure 3** illustrates the heat-map and dendrogram overview of the two-way unsupervised hierarchical cluster analysis of the reported differentially expressed genes among male and female responders to lithium therapy at baseline that correspond to values reported in **Table 3**.

Using the baseline blood sample microarray data, the predictive modeling results for identifying lithium-responders from the complete study population of male and female controls and treatment samples, resulted in a validation/test sample cohort for males of: Sensitivity=95.83% (95% C.I. 78.88%-99.89%), Specificity=not

calculated due sample size of test dataset, and an ROC curve AUC = 0.92 using the *RBPM2S* and *LILRA5* genes. Likewise, in the test dataset for females: Sensitivity=91.67% (95% C.I. 61.52%-99.79%), Specificity= not calculated due sample size of test dataset, and an ROC curve AUC = 1 with the *ABRACL* and *NBPF14* genes. Therefore, we developed a 2-gene predictive model for men and likewise for women predicting lithium response in bipolar patients from a general population of bipolar patients using transcriptional signatures at baseline.

Table 4 provides the list of 10 differentially expressed genes found in male lithium responders (5-genes) and male lithium-non-responders (5-genes). The RNA binding protein with multiple splicing 2 (*RBPM2S*) gene (FC= -1.326, unadjusted p=0.001358) in male lithium responders and the Ribosomal protein S23 (*RPS23*) gene (FC=1.521, unadjusted p=0.013306) were found to result in the largest expression change differences between subgroups. However, in female responders and female non-responders, the Family with Sequence Similarity 117 Member B (*FAM117B*) gene (FC=0.5257, unadjusted p=0.0048554) and the Golgin B1 (*GOLGB1*) gene (FC= -0.6536, unadjusted p=0.0003716) were differentially expressed, respectively and shown in **Table 5**.

Discussion

The purpose of this investigation was to define gender-specific transcriptome-level regulators of lithium treatment response prior to the initiation of lithium treatment. We first established the gender-relevant transcriptional control genes across all study-participant blood samples and specifically to male- and female-responders using a differential gene expression threshold of 0.5. We found this to be adequate and corroborated with similar studies that used a similar threshold for establishing gene transcription signatures (Jansen *et al.*, 2014; Mayne *et al.*, 2016). However, when comparing the male-responders to male non-responders, as well as, the female responders to female non-responders, we set an inclusion fold-change threshold to 0.3. This approach is not unusual, since it is already established that both large and subtle expression changes produce to significant biological and physiological processes (Wurmbach *et al.*, 2002). Our analysis is both hypothesis-generating, and establishes a computational methodology that provides insight to the importance of subgroup analysis in genomic medicine, irrespective of patient sample-sizes. The end-goal of such analyses serves as a testing methodology for establishing gene screening panels to improve personalized medicine in vulnerable and high-risk patient populations. In these patient populations, it is often not feasible to wait for weeks to determine whether a prescribed medication will work and in some cases manic patients are neither able to fully comprehend and be objectively assessed using the CGI-BP-S (Spearing *et al.*, 1997).

When reviewing the heat-map and dendrogram hierarchical cluster analysis patterns, specifically the numerous non-responders clinically-labeled and illustrated in **Figure 4**, they suggest that the underlying etiology resulting in clinical symptoms (e.g. mania) that led to the diagnosis of bipolar disorder may need re-classification. Further, the subsequent treatments may need to be tailored in data-driven computational psychiatry approaches.

Figure 2. Variable importance ratings of genes selective (above) male lithium responders versus the entire population of treated and untreated patient men and women; and (below) female lithium responders versus the entire population of treated and untreated men and women.

Table 2. Differentially expressed genes between genders across all study participants with a log fold-change threshold of 0.5.

Male-associated genes					
Gene	Adjusted P-value	P-value	Log fold change	Gene description	Location
RPS4Y1	7.36E-47	2.81E-51	-4.9807	Ribosomal Protein S4, Y-linked 1	Yp11.3
EIF1AY	1.02E-41	8.61E-46	-2.5861	Eukaryotic Translation Initiation Factor 1A, Y-linked	Yq11.223
KDM5D	7.36E-47	4.67E-51	-1.6658	Lysine Demethylase 5D	Yq11
HLA-DRB1	0.016	0.0000362	-1.7072	Major Histocompatibility Complex, Class II, DR Beta 1	
RPS4Y2	1.35E-40	1.43E-44	-1.5014	Ribosomal Protein S4, Y-linked 2	Yq11.223
EIF1AY	9.38E-31	1.98E-34	-0.9443	Eukaryotic Translation Initiation Factor 1A, Y-linked	Yq11.223
Female-associated genes					
Gene	Adjusted P-value	P-value	Log fold change	Gene description	Location
XIST	2.98E-36	5.03E-40	1.7615	X Inactive Specific Transcript (non-protein coding)	Xq13.2
S100P	1.70E-02	3.31E-05	1.028	S100 Calcium Binding Protein P	4p16
IFIT3	5.00E-03	5.82E-06	0.8765	Interferon Induced Protein with Tetratricopeptide Repeats 3	10q24
TNFAIP6	3.73E-04	2.52E-07	0.7304	TNF Alpha Induced Protein 6	2q23.3
IFITM3	4.51E-02	1.69E-04	0.7284	Interferon Induced Transmembrane Protein 3	11p15.5
IFIT2	4.91E-02	1.95E-04	0.6739	Interferon Induced Protein with Tetratricopeptide Repeats 2	10q23.31
CHURC1	6.30E-02	3.18E-04	0.6678	Churchill Domain Containing 1	14q23.3
ANXA3	2.33E-03	2.26E-06	0.6218	Annexin A3	4q21.21
ADM	8.69E-04	6.80E-07	0.5986	Adrenomedullin	11p15.4
PROK2	2.16E-02	4.79E-05	0.5189	Prokineticin 2	3p13

In [Figure 4](#), for the females, the samples in the center cluster illustrates that a group of patients are clear non-responders while the patients clustered in the far-right are partial-responders, from a molecular perspective. The natural questions that arise are: (1) How to best convert the non- and partial-responders to treatment-responders? (2) Is a behavioral intervention, in this select group of patients, for whom lithium is not effective, the best answer because the symptoms maybe of a different etiology? If indeed the symptoms are of a different etiology (e.g. inflammatory), from the lithium treatment-responders, then other diagnostic (e.g. electrophysiological neuroimaging) tools may be warranted and corresponding most efficacious treatments sought.

When differentiating between male and female patients, we found that the Ribosomal Protein S4, Y-linked 1 (*RPS4Y1*, adjusted p-value=7.36E-47) male-linked gene and the X Inactive Specific Transcript (*XIST*, adjusted p-value=2.98E-36) female-linked gene were the most differentially expressed among genders, which is consistent with previously published studies ([Guillén et al., 2014](#); [Jansen et al., 2014](#); [Mayne et al., 2016](#)). The genes that are specific

to male lithium responders, relative to female lithium responders, are *RBPM52*, *SIDT2*, *CDH23*, *LILRA5*, and *KIR2DS5*. Using the same methodology, genes identifying female lithium responders, relative to male lithium responders, are *HLA-H*, *RPS23*, *FHL3*, *RPL10A*, *NBPF14*, *PSTPIP2*, *FAM117B*, *CHST7*, and *ABRACL*. The Neuroblastoma Breakpoint Family Member 14 (*NBPF14*, adjusted p-value=0.0462, Fold-change=0.586) achieved the Benjamini-Hochberg adjusted p-value of 0.0462, and has been reported to be associated with cortical neurogenesis ([Suzuki et al., 2017](#)).

Computational psychiatry, as advocated by the National Institute of Mental Health's Research Domain Criteria (RDoC), may need data to drive the classification, diagnosis, and treatment response status, especially in patients with developmental delay, language difficulty, and condition of a potentially different etiology than traditionally taught ([Clark et al., 2017](#); [Eugene & Masiak, 2016](#)). Ideally, in such cases, alternative FDA-approved mood stabilizers may be initially selected prior to any pharmacological intervention by simply using a blood test. Perhaps, a gene expression screening panel at baseline, prior to the initiation of

Table 3. Differentially expressed genes between male and female responders prior to Lithium pharmacotherapy with a log fold-change threshold of 0.5.

Genes associated with male lithium responders					
Gene	Adjusted P-value	P-value	Log fold change	Gene description	Highest gene tissue expression
RBPM52	1	0.00111	-1.351	RNA Binding Protein with Multiple Splicing 2	Heart, Urinary Bladder
SIDT2	1	0.00932	-0.82	SID1 Transmembrane Family Member 2	Stomach, Prostate
CDH23	1	0.00388	-0.674	Cadherin-Related 23	Ovary, Fat
LILRA5	1	0.00359	-0.592	Leukocyte Immunoglobulin Like Receptor A5	Appendix, Bone Marrow
KIR2DS5	1	0.00431	-0.506	Killer Cell Immunoglobulin Like Receptor, Two Ig Domains and Short Cytoplasmic Tail 5	--
Genes associated with female lithium responders					
Gene	Adjusted P-value	P-value	Log fold change	Gene description	Highest gene tissue expression
HLA-H	1	0.000996	1.602	Major Histocompatibility Complex, Class I, H (pseudogene)	Lymph Node, Bone Marrow
RPS23	1	0.00308	1.471	Ribosomal Protein S23	Ovary, Bone Marrow
FHL3	1	0.000751	0.893	Four and a Half LIM Domains 3	Esophagus, Endometrium
RPL10A	1	0.00299	0.628	Ribosomal Protein L10a	Ovary, Bone Marrow
**NBPF14	**0.0462	0.00000782	0.586	Neuroblastoma Breakpoint Family Member 14	Skin, Ovary
PSTPIP2	1	0.000473	0.569	Proline-Serine-Threonine Phosphatase Interacting Protein 2	Bone Marrow, Spleen
FAM117B	1	0.00949	0.556	Family with Sequence Similarity 117 Member B	Testis, Adrenal
CHST7	1	0.00812	0.529	Carbohydrate Sulfotransferase 7	Spleen, Fat
ABRACL	1	0.00396	0.505	ABRA C-Terminal Like	Colon, Lymph Node

Notes: **The NBPF14 gene reached the Benjamini-Hochberg adjusted p-value.

lithium and/or other FDA-approved mood stabilizer, may be better in high-risk patient populations.

These findings suggest that when implementing genomic medicine, clinical research teams should move beyond the single-gene approach when screening for treatment responders or non-responders. This approach is currently the standard when screening for patient toxicity at standard doses in poor or ultra-rapid metabolizers; however, as more transcriptional factors are discovered that regulate the cytochrome (CYP) P-450 system of genes, multi-gene pharmacokinetic panels are inevitable and may be included in future Clinical Pharmacogenetics Implementation Consortium (CPIC) guidelines. Next, medical management of patients with mania and psychosis either with pharmacotherapy and/or behavioral intervention should be tailored to biological gender due to known neuronal circuitry differences in age-matched patients with psychosis (Eugene *et al.*, 2015). Further, as a result of lithium not being hepatically metabolized, but rather transported and renally excreted, as well as, the known myriad

drug-drug interactions, patient dose selection may benefit from clinical pharmacometrics modeling by American Board of Clinical Pharmacology certified physicians in applied clinical pharmacology/clinical pharmacology (Perera *et al.*, 2014; Zetin *et al.*, 1986). This approach may be implemented to ensure drug pharmacokinetic safety.

The limitations of our analysis and in most genetic studies are understandably due to multiple-comparison p-value adjustments and patient sample size (Dudoit *et al.*, 2003). The fundamental aims of our research questions were designed to answer biological questions of gender and clinical response to lithium and not meant to be driven exclusively by multiple comparisons adjusted p-values. This approach has led to various successes in genomic medicine, specifically, in genome-wide association studies; however, understandably, the limitations are thoroughly acknowledged. In reference to patient sample sizes, 9 out of the 28 patients who received lithium and optimal therapy were classified as lithium treatment responders. Further, 30% of men and

Figure 3. Heat-map and dendrogram overview of the two-way unsupervised hierarchical cluster analysis of differentially expressed genes in male (n=3) and female (n=6) lithium responders after filtering out the top 250 differentially expressed genes found gender biased genes.

Table 4. Differentially expressed genes between Male Responders and Male Non-Responders at baseline with a log fold-change threshold of 0.3.

Genes associated with male lithium responders					
Gene	Adjusted P-value	P-value	Log fold change	Gene description	Highest gene tissue expression
RBPMS2	1	0.001358	-1.326	RNA Binding Protein with Multiple Splicing 2	Heart, Urinary Bladder
SVBP	1	0.01366	-0.76	Small Vasohibin Binding Protein	Testis, Fat
LILRA5	1	0.011739	-0.714	Leukocyte Immunoglobulin Like Receptor A5	Appendix, Bone Marrow
CPA3	1	0.008048	-0.592	Carboxypeptidase A3	Gall Bladder, Lung
SLC45A3	1	0.016508	-0.455	Solute Carrier Family 45 member 3	Prostate, Stomach
ZNF234	1	0.003254	-0.41	Zinc Finger Protein 234	Spleen, Thyroid
DIDO1	1	0.008232	-0.385	Death Inducer-Obliterator 1	Ovary, Spleen
TPP2	1	0.013053	-0.385	Tripeptidyl Peptidase 2	Testis, Thyroid
KRT73	1	0.007333	-0.373	Keratin 73	Skin, Lymph Nodes
ZMYM3	1	0.00363	-0.372	Zinc Finger MYM-type Containing 3	Ovary, Testis
NOTCH2 NL	1	0.009657	-0.348	Notch 2 N-terminal Like	Testis, Skin
TIPRL	1	0.007794	-0.34	TOR Signaling Pathway Regulator	Endometrium, Brain
CAMK1D	1	0.005376	-0.333	Calcium/Calmodulin dependent Protein Kinase ID	Brain, Skin
EFNA1	1	0.00632	-0.324	Ephrin A1	Placenta, Lung

Genes associated with male lithium non-responders					
Gene	Adjusted P-value	P-value	Log fold change	Gene description	Highest gene tissue expression
RPS23	1	0.013306	1.521	Ribosomal Protein S23	Ovary, Bone Marrow
IRF2BPL	1	0.010952	1.005	Interferon Regulatory Factor 2 Binding Protein Like	--
HLA-C	1	0.003461	0.997	Major Histocompatibility Complex, Class I, C	Lung, Bone Marrow
RGPD1	1	0.001745	0.76	RANBP2-like and GRIP Domain Containing 1	Testis, Liver
ASGR2	1	0.019947	0.598	Asialoglycoprotein Receptor 2	Liver, Gall Bladder
LPAR1	1	0.01374	0.453	Lysophosphatidic Acid Receptor 1	Brain, Placenta
RRN3P1	1	0.017025	0.42	RRN3 homolog, RNA Polymerase I Transcription Factor Pseudogene 1	Thyroid, Lymph Node
TOMM34	1	0.016655	0.416	Translocase of Outer Mitochondrial Membrane 34	Testis, Adrenal
ACAD11	1	0.015882	0.405	Acyl-CoA Dehydrogenase Family Member 11	Kidney, Liver
CEBPE	1	0.00269	0.404	CCAAT/enhancer Binding Protein Epsilon	Bone Marrow, Small Intestine
CMIP	1	0.017203	0.394	C-Maf Inducing Protein	Brain, Small Intestine
IGSF6	1	0.011786	0.38	Immunoglobulin Superfamily Member 6	Spleen, Appendix
HDHD2	1	0.01764	0.361	Haloacid Dehalogenase Like Hydrolase Domain Containing 2	Brain, Thyroid
LMO4	1	0.012872	0.359	LIM Domain Only 4	Brain, Stomach
BACE2	1	0.000711	0.353	Beta-site APP-Cleaving Enzyme 2	Stomach, Gall Bladder
TPP1	1	0.00061	0.341	Tripeptidyl Peptidase 1	Spleen, Appendix
GALNS	1	0.007613	0.341	Galactosamine (N-acetyl)-6-Sulfatase	Bone Marrow, Testis
SYNM	1	0.019042	0.322	Synemin	Esophagus, Prostate

Table 5. Differentially expressed genes between Female Responders and Female Non-Responders at baseline with a log fold-change threshold of 0.3.

Genes associated with female lithium responders					
Gene	Adjusted P-value	P-value	Log fold change	Gene description	Highest gene tissue expression
FAM117B	0.998	0.0048554	0.5257	Family with Sequence Similarity 117 Member B	Testis, Adrenal
STAMBPL1	0.998	0.0074433	0.39	STAM Binding Protein Like 1	Adrenal, Testis
CD248	0.998	0.0038199	0.3626	CD248 Molecule	--
IFIH1	0.998	0.0075822	0.3453	Interferon Induced with Helicase C domain 1	Spleen, Appendix
GPR160	0.998	0.0071723	0.3394	G Protein-coupled Receptor 160	Small Intestine, Duodenum
STAP1	0.998	0.0053096	0.3222	Signal Transducing Adaptor Family Member 1	Lymph Node, Appendix
YEATS4	0.998	0.0089003	0.3103	YEATS Domain Containing 4	Testis, Bone Marrow
CD83	0.998	0.0004367	0.3014	CD83 Molecule	Bone Marrow, Lymph Node
TMOD2	0.998	0.0081514	0.3012	Tropomodulin 2	Brain, Appendix

Genes associated with female lithium non-responders					
Gene	Adjusted P-value	P-value	Log fold change	Gene description	Highest gene tissue expression
GOLGB1	0.998	0.0003716	-0.6536	Golgin B1	
RASA4CP	0.998	0.0030349	-0.4554	RAS p21 Protein Activator 4C, Pseudogene	Spleen, Endometrium
NACC2	0.998	0.0061286	-0.3803	NACC Family Member 2	Brain, Fat
EDARADD	0.998	0.0021425	-0.3553	EDAR Associated Death Domain	Urinary Bladder, Kidney
ZNF573	0.998	0.0058465	-0.3463	Zinc Finger Protein 573	Thyroid, Spleen
ALDH2	0.998	0.0031872	-0.335	Aldehyde Dehydrogenase 2 Family (mitochondrial)	Fat, Liver
TAPBPL	0.998	0.0032596	-0.3206	TAP Binding Protein Like	Duodenum, Small Intestine

Figure 4. Heat-map and dendrogram overview of the two-way unsupervised hierarchical cluster analysis of differentially expressed genes prior to lithium treatment in (above) male lithium responders (n=3, RESP_Male) and male lithium non-responders (n=7, NR_Male); and (below) female responders (n=6, RESP_Fem) and female non-responders (n=14, NR_Fem).

33% of women, who were treated with lithium, were found to be responders at the respective gender categories (Beech *et al.*, 2014). However, the strengths of our findings are in the gender-gene screening ability for lithium treatment-responders in the general population of 60 patients at baseline, minus the tested responder group. Opportunities exist for prospective clinical trials and application of the methods outlined in this text for other therapeutic agents across several medical specialties.

Conclusion

We explored the Lithium Treatment-Moderate dose Use Study clinical trial gene expression data with the aim of identifying gender-specific transcriptome-level regulators of lithium treatment response. Using machine learning, we successfully developed a pre-treatment gender- and gene-expression-specific predictive model selective for lithium responders with an ROC AUC of 0.92 for men and an ROC AUC of 1 for women. Further, by using well-established Bayesian statistical methods, to identify differentially expressed genes and then machine learning, we discovered 5-genes selective for men and 9-genes that are selective for women that will inform the physicians and clinical staff of whether the patient will respond to lithium prior to being prescribed the drug. With the small number of

patient responders from the clinical trial, our results should be confirmed. Lastly, in an overall context, our results suggest that the methodology used in this analysis may be extended to other therapeutic drug classes and provides insight to the gender-based gene transcriptome differences influencing lithium pharmacodynamics.

Data availability

Data used in this study are available from <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE45484>

Competing interests

No competing interests were disclosed.

Grant information

The author(s) declared that no grants were involved in supporting this work.

Acknowledgements

The authors gratefully acknowledge the patients in the original clinical trial, the medical staff, and the NCBI GEO database accession GSE4548.

Supplementary material

Supplementary File 1: Supplementary methods.

[Click here to access the data.](#)

References

- Beech RD, Leffert JJ, Lin A, *et al.*: **Gene-expression differences in peripheral blood between lithium responders and non-responders in the Lithium Treatment-Moderate dose Use Study (LITMUS).** *Pharmacogenomics J.* 2014; 14(2): 182–91.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Clark LA, Cuthbert B, Lewis-Fernández R, *et al.*: **Three Approaches to Understanding and Classifying Mental Disorder: ICD-11, DSM-5, and the National Institute of Mental Health's Research Domain Criteria (RDoC).** *Psychol Sci Public Interest.* 2017; 18(2): 72–145.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Dudoit S, Shaffer JP, Boldrick JC: **Multiple Hypothesis Testing in Microarray Experiments.** *Statist Sci.* 2003; 18(1): 71–103.
[Publisher Full Text](#)
- Eugene AR, Masiak J: **Identifying Treatment Response of Sertraline in a Teenager with Selective Mutism Using Electrophysiological Neuroimaging.** *Int J Clin Pharmacol Toxicol.* 2016; 5(4): 216–19.
[PubMed Abstract](#) | [Free Full Text](#)
- Eugene AR, Masiak J, Kapica J, *et al.*: **Electrophysiological Neuroimaging Using sLORETA Comparing 22 Age Matched Male and Female Schizophrenia Patients.** *Hosp Chron.* 2015; 10(2): 91–98.
[PubMed Abstract](#) | [Free Full Text](#)
- Guillén IA, Fernández JR, Palenzuela DO, *et al.*: **Analysis of Gene Expression Profile for Gender in Human Blood Samples.** *International Journal of Innovation and Applied Studies.* 2014; 7(1): 329–42.
[Reference Source](#)
- Hayes JF, Pitman A, Marston L, *et al.*: **Self-harm, Unintentional Injury, and Suicide in Bipolar Disorder During Maintenance Mood Stabilizer Treatment: A UK Population-Based Electronic Health Records Study.** *JAMA Psychiatry.* 2016; 73(6): 630–7.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Hiemke C, Baumann P, Bergemann N, *et al.*: **AGNP Consensus Guidelines for Therapeutic Drug Monitoring in Psychiatry: Update 2011.** *Pharmacopsychiatry.* 2011; 44(6): 195–235.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Jansen R, Batista S, Brooks AI, *et al.*: **Sex differences in the human peripheral blood transcriptome.** *BMC Genomics.* 2014; 15(1): 33.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Jermain DM, Crismon ML, Martin ES 3rd: **Population pharmacokinetics of lithium.** *Clin Pharm.* 1991; 10(5): 376–81.
[PubMed Abstract](#)
- Labonté B, Engmann O, Purushothaman I, *et al.*: **Sex-specific transcriptional signatures in human depression.** *Nat Med.* 2017; 23(9): 1102–11.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Landersdorfer CB, Findling RL, Frazier JA, *et al.*: **Lithium in Paediatric Patients with Bipolar Disorder: Implications for Selection of Dosage Regimens via Population Pharmacokinetics/Pharmacodynamics.** *Clin Pharmacokinet.* 2017; 56(1): 77–90.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Lobeck F, Nelson MV, Evans RL, *et al.*: **Evaluation of Four Methods for Predicting Lithium Dosage.** *Clin Pharm.* 1987; 6(3): 230–33.
[PubMed Abstract](#)
- Mayne BT, Bianco-Miotto T, Buckberry S, *et al.*: **Large Scale Gene Expression Meta-Analysis Reveals Tissue-Specific, Sex-Biased Gene Expression in Humans.** *Front Genet.* 2016; 7: 183.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Perera V, Bies RR, Mo G, *et al.*: **Optimal sampling of antipsychotic medicines: a pharmacometric approach for clinical practice.** *Br J Clin Pharmacol.* 2014; 78(4): 800–814.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Ritchie ME, Phipson B, Wu D, *et al.*: **Limma Powers Differential Expression Analyses for RNA-Sequencing and Microarray Studies.** *Nucleic Acids Res.* 2015; **43**(7): e47.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Spearing MK, Post RM, Leverich GS, *et al.*: **Modification of the Clinical Global Impressions (CGI) Scale for Use in Bipolar Illness (BP): The CGI-BP.** *Psychiatry Res.* 1997; **73**(3): 159–71.
[PubMed Abstract](#) | [Publisher Full Text](#)

Suzuki IK, Gacquer D, Van Heurck R, *et al.*: **Hominin-Specific NOTCH2 Paralogs Expand Human Cortical Neurogenesis through Regulation of Delta/Notch Interactions.** *bioRxiv.* Cold Spring Harbor Laboratory, 2017; 221358.
[Publisher Full Text](#)

Team R: **R Development Core Team. R: A Language and Environment for**

Statistical Computing. 2013.
Reference Source

Viguera AC, Tondo L, Baldessarini RJ: **Sex differences in response to lithium treatment.** *Am J Psychiatry.* 2000; **157**(9): 1509–11.
[PubMed Abstract](#) | [Publisher Full Text](#)

Wurmbach E, González-Maeso J, Yuen T, *et al.*: **Validated genomic approach to study differentially expressed genes in complex tissues.** *Neurochem Res.* 2002; **27**(10): 1027–33.
[PubMed Abstract](#) | [Publisher Full Text](#)

Zetin M, Garber D, De Antonio M, *et al.*: **Prediction of lithium dose: a mathematical alternative to the test-dose method.** *J Clin Psychiatry.* 1986; **47**(4): 175–78.
[PubMed Abstract](#)

Open Peer Review

Current Peer Review Status:

Version 2

Reviewer Report 12 November 2018

<https://doi.org/10.5256/f1000research.16407.r39873>

© 2018 Kalmady S. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Sunil V. Kalmady

¹ Department of Psychiatry, University of Alberta, Alberta, Canada

² Alberta Machine Intelligence Institute, Alberta, Canada

Eugene et al. tackles the hard question on the transcriptome predictors of clinical outcome in lithium treatment of Bipolar disorder. This research question is of clinical relevance and importance, however, there might be some issues with how the data was analyzed and presented in the manuscript.

- The study design and methods do not seem to be coherent to a single unifying goal. Whether the goal is to predict the responder using the 'gender' and 'transcriptome' data as features? If so, such a model can be learned using standard machine learning methods. On the other hand, whether the goal is to identify are genes with statistically significant group differences in their expression? If so, this can be achieved by biostatistical inference tests of association.

Please note that task of 'association' and 'prediction' are quite distinct in their formulation and desired objective. Authors have to be really careful about that they trying to test and claim while using both of approaches in conjunction.

- Study design is a bit unconventional, and hence needs to be motivated and explained better. For example: Performing successive sub-group analyses partitioned on factors like gender have less power, and should be generally restricted to post-hoc tests. Why not simply use standard biostatistical tools such as factorial ANOVA with 'sex' and 'response' as between-subjects factors of interest?

I agree with Reviewer #1 that flowchart of analysis pipeline will help the understanding. Steps of sub-group selection and variable/feature selection can be indicated in this flowchart. Care should be taken to avoid the circularity that can arise from selection because statistical inference can be invalid whenever the results statistics are not inherently independent of the selection criteria under the null hypothesis.

- Authors might be asking too many questions with limited data in hand. Sample size might not enough to study individual effects of multiple factors - such as treatment-type, response and then, the gender. Cell-wise sample sizes resulting from this 8-way split is less than 10 for all but two cells (less than 5 for 3 cells). Suitability of applied statistical tests and

generalizability of their claims are questionable here. Authors should also think about 1:2 skew in male:female ratio, which makes this issue worse. Study can greatly benefit from asking specific and limited hypothesized questions.

Also, since multiple objectives are stated, methods and results section can describe each objective separately for sake of better clarity.

- Machine learning methods are not described. The methods used for learning of model and its evaluation process needs be specified. Example: How was training and test splits performed? How was feature selection performed? How were the hyper-parameters optimized? Whether the reported performance metrics are for training or testing sets? Whether the discovery dataset used for identifying '250 genes' disjoint from validation set? etc. Without these details, it is hard to comment on validity of a predictive study.

Is the work clearly and accurately presented and does it cite the current literature?

Partly

Is the study design appropriate and is the work technically sound?

Partly

Are sufficient details of methods and analysis provided to allow replication by others?

No

If applicable, is the statistical analysis and its interpretation appropriate?

Partly

Are all the source data underlying the results available to ensure full reproducibility?

Yes

Are the conclusions drawn adequately supported by the results?

Partly

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Psychiatric research, Biostatistics, Machine learning.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Author Response 27 Nov 2018

Andy Eugene

Please see our responses in bold-face to the comments from Referee #2. We thank you for reviewing this article.

Eugene et al. tackles the hard question on the transcriptome predictors of clinical outcome

in lithium treatment of Bipolar disorder. This research question is of clinical relevance and importance, however, there might be some issues with how the data was analyzed and presented in the manuscript.

- The study design and methods do not seem to be coherent to a single unifying goal. Whether the goal is to predict the responder using the 'gender' and 'transcriptome' data as features? If so, such a model can be learned using standard machine learning methods. On the other hand, whether the goal is to identify are genes with statistically significant group differences in their expression? If so, this can be achieved by biostatistical inference tests of association.
- **Response: First and foremost, I sincerely appreciate you taking time out to review this article and provide constructive feedback. In answering these points, this paper addressed questions that were not addressed in the original clinical study and attempted not to duplicate work previously published. Hence, is why the methods are creative and not employ standard biostatistical inference tests of association. For example, we aimed to test the hypothesis if gender influenced selecting transcriptome signatures associated with lithium efficacy. Therefore, in order to accomplish this task, we conducted a quality control test that identifies known gender-biased transcriptome signatures. This proved to be essential and is why we identified that several patients were misclassified as male/female. Results are shown, documented, and corroborated with other studies identifying the same genes as being gender-specific. Next, standard statistical tests were indeed used, however, we chose to use those adapted for Gene Expression analysis as has been developed using the “limma” package and for machine learning methods, we used the Random Forest algorithm. We then aimed to develop a predictive model to see if this may be a first-step for other medical research teams to validate with further clinical studies and for Clinical Pharmacology laboratories to pursue and develop novel experiments to determine lithium’s effect in cells, tissues, laboratory animals, and later in humans.**

Please note that task of 'association' and 'prediction' are quite distinct in their formulation and desired objective. Authors have to be really careful about that they trying to test and claim while using both of approaches in conjunction.

Response: Well said and this is duly noted. We first aimed to identify genes associated with responders and hoped to identify if using those genes would help in creating hypothesis-generating predictors treatment response, only to be later validated by other studies. Thank you for the comment.

- Study design is a bit unconventional, and hence needs to be motivated and explained better. For example: Performing successive sub-group analyses partitioned on factors like gender have less power, and should be generally restricted to post-hoc tests. Why not simply use standard biostatistical tools such as factorial ANOVA with 'sex' and 'response' as between-subjects factors of interest?
- **Response: While your point with “gender” should be restricted to post-hoc analysis, we are looking to make gender the primary point of our analysis in patients responding to lithium treatment and specifically not repeat the analysis from the original publication in Nature. This was attempted on the**

original article, however, we clearly identified that there was patient-gender misclassification in the original study. So, we sought another route of analysis to ensure that gender was a primary point of analysis and not side-lined to the post-hoc analyses. However, we do understand that this compromises statistical power and therefore, sought not to analyze the data with ANOVA, because we are addressing minor gene expression level changes that might have real-world clinical insight. This is a hypothesis-generating analysis, however, we do thank you as well for this comment.

I agree with Reviewer #1 that flowchart of analysis pipeline will help the understanding. Steps of sub-group selection and variable/feature selection can be indicated in this flowchart. Care should be taken to avoid the circularity that can arise from selection because statistical inference can be invalid whenever the results statistics are not inherently independent of the selection criteria under the null hypothesis.

Response: We appreciate the request for having a graphical flowchart depicting the analysis pipeline and have included the figure in the updated version of the manuscript. We are confident that circularity is not an issue, given the specific aims of our analysis. These methods are determined to seeking to identify the influence of linked gender-drug-response to genes at baseline and not after treating with the mood stabilizer. The gender differences in clinical practice are a well-documented reality and we literally sought to identify any signal, on the gene expression level, to address the clinical question rather than entirely use traditional statistical methods which did not necessarily translate to clinical translation. With our results, we are expecting laboratories having strengths in gene knock-down/out and gene over-expression experiments to identify the mechanisms to lithium's efficacy. This is an old drug and until this day, most textbooks lack knowledge of the drug's mechanism. These mechanisms may be attributable to biological, biochemical, gene expression, hormonal, and proteomic differences that we are aiming to identify here in this article. Please see the new figure showing the data analysis pipeline used in this paper.

- Authors might be asking too many questions with limited data in hand. Sample size might not enough to study individual effects of multiple factors - such as treatment-type, response and then, the gender. Cell-wise sample sizes resulting from this 8-way split is less than 10 for all but two cells (less than 5 for 3 cells). Suitability of applied statistical tests and generalizability of their claims are questionable here. Authors should also think about 1:2 skew in male:female ratio, which makes this issue worse. Study can greatly benefit from asking specific and limited hypothesized questions.
- **Response:** We do appreciate your robustness in identifying the obvious study limitations due to sample-size, however, we are limited to the feasibility, cost of research, patient population, and all of the work accomplished from the original study team stemming from Case Western Reserve University, Massachusetts General Hospital (Harvard University), Stanford University, Yale University, the University of Pittsburgh, Texas Health Science Center at San Antonio, and the University of Pennsylvania that uploaded this data into the Gene Expression Omnibus (GEO) database maintained by the National Institutes of Health. This was a massive undertaking in a multi-site trial. The sample-size limitations are classically used to not have results generalized, however, we ask you realize

that this work not necessarily straight-forward to accomplish with in the real-world of medical care in mental health. Nevertheless, your points are well noted.

- **Response:** The 1:2 male:female skew you are referring to is exactly what is seen in clinical medicine. Females tend to respond more so than males and I clearly stated this in the introduction of the paper. We are working with the data that has been uploaded and have not found any other datasets in GEO. However, please understand that our efforts are indeed, as you clearly pointed in the beginning of this review, that this is a difficult clinical question to answer. Rather than saying we do not have enough samples, we aimed to do 'something' rather than let the dataset sit in GEO while patients are in need and laboratories have the capability and funding to seek follow-up studies. We appreciate your clear expertise and concern. However, we are working to create hypothesis-generating results to be later confirmed, expanded-upon, and validated for sick patients. Therefore, the generalizability of our claims are clearly limited to the dataset we obtained from clinical study of the aforementioned university hospitals. Of the 60 (sixty) patients treated with Lithium, literally only 9 responded in follow-up my expert medical teams and high quality care. Hence the need to find 'some signal' with the data at-hand in the form of expression patterns of lithium treatment responders. Thank you for the statement and again these are well-noted points you stated here.

Also, since multiple objectives are stated, methods and results section can describe each objective separately for sake of better clarity.

- Machine learning methods are not described. The methods used for learning of model and its evaluation process needs be specified. Example: How was training and test splits performed? How was feature selection performed? How were the hyper-parameters optimized? Whether the reported performance metrics are for training or testing sets? Whether the discovery dataset used for identifying '250 genes' disjoint from validation set? etc. Without these details, it is hard to comment on validity of a predictive study.
- **Response:** Thank you for these questions and comments. We have updated the methods to better explain the approach used in the study. The new graphical analysis pipeline will help in explaining the approach. We also added the final Decision Tree diagrams to identify male- and female-treatment responders. Thank you for the review and we have made considerable updates to this version of the paper to address these concerns and improve this research manuscript.

Competing Interests: None.

Reviewer Report 27 June 2018

<https://doi.org/10.5256/f1000research.16407.r34437>

© 2018 Ho M. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Ming-Fen Ho

Division of Clinical Pharmacology, Department of Molecular Pharmacology and Experimental Therapeutics, Mayo Clinic, Rochester, MN, USA

Is the work clearly and accurately presented and does it cite the current literature?

Partly

Is the study design appropriate and is the work technically sound?

Partly

Are sufficient details of methods and analysis provided to allow replication by others?

Partly

If applicable, is the statistical analysis and its interpretation appropriate?

Partly

Are all the source data underlying the results available to ensure full reproducibility?

Partly

Are the conclusions drawn adequately supported by the results?

Partly

Competing Interests: No competing interests were disclosed.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Version 1

Reviewer Report 21 May 2018

<https://doi.org/10.5256/f1000research.15730.r33941>

© 2018 Ho M. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Ming-Fen Ho

¹ Division of Clinical Pharmacology, Department of Molecular Pharmacology and Experimental

Therapeutics, Mayo Clinic, Rochester, MN, USA

² Division of Clinical Pharmacology, Department of Molecular Pharmacology and Experimental Therapeutics, Mayo Clinic, Rochester, MN, USA

The authors demonstrated that sex-differences gene expression might contribute to lithium treatment response using microarray expression data.

Major comments:

1. A samples size of 60 might be too small to determine the sex effects. Can the sample size n=60 provide adequate power for data interpretation, especially separated men and women for study sex-effect on gene expression?
2. The authors stated that their predictive model for lithium responders with an ROC AUC 0.92 for men, and 1 for women. If the prediction accuracy is so significant, what are the potential biological mechanisms beyond these genes? More discussion regarding the biology of those genes should be included in the paper. Once again, if the prediction accuracy is so significant, it is needed a replication study using different data sets? In summary, the authors claimed the prediction model with very high accuracy; it should be included either functional validation of those genes or a replication study population.

Specific comments:

1. Methods - study design, it might be better to use a flow chart to demonstrate the study design.
2. Methods - study design, please clarify the rationale of filtering out "250" genes.
3. Table 1 shows total study population n=60, but figure 1 legend shows male: n=41, female: n=39?
4. Figure 2: please elaborate the data presented in Figure 2. The key results for each of the four panels should be summarized in Results.
5. Table 2 and Table 4, the log FC threshold of 0.5 or 0.3 might be too low. The changes in gene expression are very subtle in Table 4.
6. Table 2, are there any gene up-regulated in males? /downregulated in females?
7. Limitations of the study should be addressed in Discussion.

Is the work clearly and accurately presented and does it cite the current literature?

Partly

Is the study design appropriate and is the work technically sound?

Partly

Are sufficient details of methods and analysis provided to allow replication by others?

Partly

If applicable, is the statistical analysis and its interpretation appropriate?

Partly

Are all the source data underlying the results available to ensure full reproducibility?

Partly

Are the conclusions drawn adequately supported by the results?

Partly

Competing Interests: No competing interests were disclosed.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Author Response 21 May 2018

Andy Eugene

Major Comment Responses:

Response 1:

This point is well noted; however, it is important to realize that our gender-effects of gene expression is consistent with other studies noted within the paper and shown below:

Jansen, Rick, et al. "Sex differences in the human peripheral blood transcriptome." *BMC genomics* 15.1 (2014): 33.

Mayne, Benjamin T., et al. "Large scale gene expression meta-analysis reveals tissue-specific, sex-biased gene expression in humans." *Frontiers in genetics* 7 (2016): 183.
Further, our gender-specific results met the Benjamini-Hochberg multiple comparisons criteria adjustment due to multiple comparisons.

Comment Response 2:

We welcome and thank the reviewer's comments on the biological mechanisms beyond these genes. Clearly, it is well noted and cited in the paper that in clinical practice there is a wide inter-individual variability in the treatment and response to treatments of bipolar disorder. Moreover, these patients were not treated with lithium monotherapy, alone, and therefore further insight into the biological mechanisms were left out due to these patients were treated with an "Optimal Therapy" that includes a variety of other FDA-approved mood stabilizers.

In reference to the comment regarding the prediction accuracy, we agree that the study may warrant functional validation in a laboratory; however, it is beyond the scope of our computational psychiatry study and we will leave the functional genomics characterization of the genes to investigators seeking to pursue the findings from our results.

Competing Interests: No competing interests were disclosed.

Author Response 21 May 2018

Andy Eugene

Specific Comment Responses:

We thank you for your specific comments and have addressed several of the pertinent points in your review. For all differentially expressed results reported throughout tables within the manuscript, we changed the wording from genes up-regulated or down-regulated in males or females to a clearer description statement that of genes-associated with males or females. However, we thought not necessary to include an extra figure, but rather encourage the reader to (1) review the study design section within the methods to better understand the computational approach used in our analysis and (2) read the systematic tabular reporting of the results in the manuscript text as well to understand that study approach.

For the caption in Figure 1, we thank you for the comment and have updated the sample sizes for males and female patients. The updated Figure 1 text reads: Males (n=20; with 40 pre- and post-treatment samples) and Females (n=40; with 80 pre- and post-treatment samples).

The comments regarding: (1) the fold-change of 0.5 and 0.3 being subtle and (2) the study limitations, are already specifically addressed within the original version of the manuscript. Again, it is well established and referenced within the text that small changes in gene expression have already been reported to result in major functional outcomes in human physiology.

We will update the variable importance illustration shown in Figure 2 and that will be added to the updated version of the manuscript.

Competing Interests: No competing interests were disclosed.

The benefits of publishing with F1000Research:

- Your article is published within days, with no editorial bias
- You can publish traditional articles, null/negative results, case reports, data notes and more
- The peer review process is transparent and collaborative
- Your article is indexed in PubMed after passing peer review
- Dedicated customer support at every stage

For pre-submission enquiries, contact research@f1000.com

F1000Research