Article Title: Neural markers in pediatric bipolar disorder and familial risk for bipolar disorder

Authors: Jillian Lee Wiggins, Ph.D.; Melissa A Brotman, Ph.D.; Nancy E Adleman, Ph.D.; Pilyoung Kim, Ph.D.; Caroline G Wambach, BS; Richard C Reynolds, MS; Gang Chen, Ph.D.; Kenneth Towbin, M.D.; Daniel S Pine, MD; Ellen Leibenluft, MD

Background: The authors compared youth at high familial risk for BD (i.e., those with a first-degree relative with the disorder; "high-risk", HR) to low-risk (LR) youth (i.e., those without a first-degree BD relative) and to patients with BD during an fMRI experiment focusing on face emotion labelling. The authors found three patterns of results: regions where (a) BD and HR share deficits (risk endophenotypes), (b) HR show unique alterations (resilience markers), and (c) BD show unique alterations (disorder sequelae). Risk endophenotypes included alterations in higher-order face processing (e.g., middle temporal gyrus, dorsolateral prefrontal cortex). Resilience markers and disorder sequelae included different patterns of alterations across other face processing (e.g., fusiform), executive function (e.g., inferior frontal gyrus), and mentalizing (e.g., default network, superior temporal sulcus, temporo-parietal junction) regions. The authors concluded that neural patterns that may be risk endophenotypes could be used to identify individuals at risk for BD for prevention and treatment measures.

Overall feedback: I appreciate the opportunity to review this very well-written paper focusing on the functional correlates of BD in patients and high/low risk individuals. I commend the authors for a number of strengths including the comparison of high/low risk individuals and BD and the use of a task measuring face labelling recognition as it has implications for the identification of neural markers of BD, “hot cognition” and “affective processing”. Considering these strengths, though, as I read the manuscript I found some areas in which I would have appreciated greater clarity. I believe the paper could be further strengthened by addressing the following points:

1. I appreciate the fact the introduction needs to be concise but I would recommend that the authors write a short paragraph on affective/emotional regulation in BD (patients and relatives), and related networks that usually activate. This could help the reader better compare the results and relevance of the findings.
2. Could the authors explain whether LR individuals had relatives (other than first-degree relatives) who suffered from mental illness?
3. Could you please clarify if participants had comorbidities, in particular ADD/ADHD, MDD NOS or BD NOS?
4. Did the authors collect both accuracy and response latencies to the emotion labelling task? If not do they think this could contributed to the interpretation of the fMRI data or not?
5. The fact that the omnibus F for emotion labelling accuracy was significant and the posthoc analyses were not is intriguing. How do the authors explain this? I assume it’s due to the correction for multiple comparisons in post-hoc tests, but reading the conclusions it becomes unclear if the authors think that there are or there are not differences in accuracy across groups.
6. Did the authors attempt to relate accuracy rates to the functional measures (e.g. correlations, covariates) to see if there were trends?
7. Also did the authors correlate functional measures with mood state at the time of testing? I would like to know how much mood may have modulated functional activation during this task.
8. The authors talk about neural markers of resilience, risk factors etc. But if the behavioral “outcome” is almost the same would the authors suggest that there are mechanisms of neural efficiency or overcompensation? Please clarify how you view the interaction between functional and behavioral results in this study.
9. In one of your previous paper you reported results related to the relation between Emotion and Intensity in terms of labelling accuracy. Could you please clarify what you found in the HR group compared to LR/BD?
10. Could you please explain and interpret the “more exaggerated cubic response” in anger and happiness in the inferior frontal gyrus and precentral gyrus in your discussion/
11. Also please provide a rationale for showing emotional faces of “different” intensity? The rationale is intuitive but needs to be explicitly stated.
12. Which measure of irritability did you use?
13. And overall what kind of clinical measures did you administer? I see a list of clinical measures in Table 1 but they are not mentioned in the manuscript.
14. Were some of HR and BD biologically related?
15. Did the authors consider assessing what kind of emotions the images elicited in the participants? Do they think this may affected the current findings?
16. Did the authors consider “task duration” as one of the confounding effects? (I assume it’s still 8.5 minutes x 4?) Did they for instance compare results at the beginning and at the end of the task? Related to this, did the authors check for attentional deficits?
17. In terms of clinical relevance of the findings I wonder if the authors could discuss implications for cognitive and psychosocial functioning of BD participants and high-risk individuals.
18. Tables/figures: could you please define where you would place these in the manuscript.
19. Overall, the methods section did not provide sufficient details were provided for a proper replication of the study. Reading your paper published in Soc Cogn Affect Neurosci clarified some of my queries. You should perhaps consider referring to this paper one or two more times if additional methodological details (besides the task description) are provided in that paper.