Author's response to reviews

Title: Meta-analysis derived atopic dermatitis (MADAD) transcriptome confirms core AD characteristics and presents novel pathogenic insights

Authors:

David Adrian Ewald (adrian@e-wald.com)
Dana Malajian (dana.malajian@gmail.com)
James G Krueger (kruegej@rockefeller.edu)
Christopher T Workman (workman@cbs.dtu.dk)
Tianjiao Wang (jiaolx@qq.com)
Suyan Tian (windytian@hotmail.com)
Thomas Litman (thomas.litman@leo-pharma.com)
Emma Guttman-Yassky (eguttman@rockefeller.edu)
Mayte Suarez-Farinas (farinam@rockefeller.edu)

Version: 2 Date: 8 June 2015

Author's response to reviews: see over
Dear Prof. Huang,

Enclosed please find our point-by-point responses to the comments from the reviewer for our manuscript entitled "Meta-analysis derived atopic dermatitis (MADAD) transcriptome confirms core AD characteristics and presents novel pathogenic insights". We thank the editor and the reviewers for their helpful comments. Revised line numbers appear in parentheses (numbers refer to revised copy).

COMMENTS FROM EDITOR:

In addition, the authors should point out in the paper that all the four selected datasets were from the authors' previous studies.

Response to Comment from Editor:

We would like to emphasize that we performed a large database search to establish the set of studies to be used. This data mining process as described in the manuscript, was carried out in accordance with established principles for meta-analyses (PRISMA) and was unbiased. The fact that the studies are all related to our group was due to the lack of relevant other studies satisfying the inclusion criteria defined prior to the data selection process (biopsy samples with no treatment, etc.). We have added a line clarifying this in the manuscript (line 135).
COMMENTS FROM REVIEWER #1:

**General Response:** We thank the reviewer for the considered and positive evaluation of our manuscript.

**Minor Comment 1:** Fig. 1A, while colorful, does not belong to the main text, but in the supplement. A simple statement that only 25 genes were common to the 4 studies will suffice.

**Response:** We thank the reviewer for pointing this out, and do agree that a simple statement regarding the 25 common genes could generally suffice. However, as the core of our work is the lack of overlap between the datasets considered, and hence the need for a proper meta-analysis, we want to emphasize this fact visually by this figure, making it immediately obvious to the reader the limitation of intercepting gene lists and how our work can contribute to the field.

**Minor Comment 2:** p.11 line 218 KRT 16 should be KRT16.

**Response:** We thank the reviewer for capturing this, and have corrected it accordingly (line 225).

**Minor Comment 3:** 'lipid genes' should be 'lipid metabolism genes'

**Response:** We have changed ‘lipid genes’ to ‘lipid metabolism genes’ in the manuscript (lines 177, 281, 288, 291, 294, 297,298, 302-303,306).

**Minor Comment 4:** 'S. aureus' should be italicized.

**Response:** We have italicized 'S. aureus' in the manuscript (line 462).

**Minor Comment 5:** Legends for Fig 1B and C are inverted
Response: We thank the reviewer for this comment. We have switched the figure legends 1B and 1C, to reflect the figure set (lines 1035-1039).

Minor Comment 6: Legend for Fig 1B is missing (or C).
Response: We double-checked carefully and the legends seem to be present (page 53).

Minor Comment 7: In fig 3C a representation of the 19 genes in this metaanalysis, comparing it to the 4 individual analyses, would be illustrative.
Response: We thank the reviewer for this suggestion. We have added the original MTGDR output table to the supplementary section (Table E13), which illustrates the four individual analyses. The figure represents the information on this table highlighting that the coefficients of the classifier do vary within each study.

Minor Comment 8: In Supplement 12 'Molecules" should be 'Genes'
Response: We thank the reviewer for noticing this, and have changed Molecules to Genes in supplement 12 (now supplement Table E5).

Minor Comment 9: Column widths in supplement tables should be congruent.
Response: We have adjusted the column widths in the supplement tables to be congruent.

Minor Comment 10: Table numbers in the supplement seem random.
Response: We thank the reviewer for pointing this out. We have thoroughly
evaluated the supplement numbers with respect to first appearance in the manuscript, and adjusted the order if needed.

COMMENTS FROM REVIEWER #2:

**General Response:**
We thank the reviewer for the well-considered in-depth evaluation and valuable constructive comments. In this study we considered four studies, of which three do not differ significantly in SCORAD, whereas the fourth study, which uses EASI for disease scoring, was considered under the assumption of comparable patient selection.

**Major Comment 1:**
Reference 9 describes matched non-lesional, acute and chronic lesions from 10 AD patients from their previous study, in which they reported intensification of pathway expression between non-lesional, to acute and chronic lesions. When assessing individual study effects in this manuscript, were acute lesions considered? If so, what were the methods employed? If hierarchical models were utilized and non-lesional, acute, and chronic lesions were analyzed, were acute and chronic lesions considered as separate levels of the tissue factor? See related Minor comment 3.

**Response:**
We thank the reviewer for this valuable suggestive question. We have indeed considered the possibility of distinguishing between acute and chronic lesions in this manuscript. Unfortunately reference 9 (GSE 32924) was the only (of the selected) studies that explicitly made this distinction and included acute lesions in additions to the commonly considered chronic lesions (as was the case for the other three datasets). We therefore included solely chronic lesions from this study. We have now clarified this in the methods section (line 193-196).

**Major Comment 2:**
Potentially important differences in clinical characteristics of the studies included
this meta-analysis are not clearly described in this manuscript, nor is it obvious from the comparison of the original published papers. Differences in severity and current therapies may be important sources of gene expression variation with implications for the meta-analysis. The authors need to more thoroughly address this to facilitate interpretation of these analyses. For example, the Methods section describes that data “subject to treatments” were not included in these analyses; does this include TCS and calcineurin inhibitors at time of sample collection? For example, were samples described by reference 10 only from studies M4A, M4B, and M12 which excluded “Treatment with topical glucocorticosteroids, tacrolimus, and/or pimecromilus within 1 week before baseline visit”?

Response:

We appreciate this insightful comment and question. We only choose baseline data of samples from patients that were not in treatment at the time of sample collection, generally patients were out of topical treatments for two weeks and biological for four weeks. For GSE59294 (reference 10, Beck et al. 2014) in particular, this indeed included M4A, M4B, and M12.

The four studies included only patients with moderate (SCORAD 25-50) to severe (SCORAD 50-103) disease symptoms. Furthermore, as indicated in our introductory general response the distribution of the disease severity scores for the individual studies do not differ significantly among the 3 studies with SCORAD values (p=0.133, Anova). To compare with the 4th study where only EASI was used, we used the following transformation

\[
\text{SCORAD} = 22.9734 + 1.16 \times \text{EASI}
\]

(which was obtained from a dataset with 134 patients where both EASI/SCORAD was obtained). No significant differences were found when the 4 studies were compared (p=0.12). Additionally no differences in IgE were found across studies (p=0.33) (see lines 193-196).
**Major Comment 3:**
We commend the authors for addressing dataset level effects, i.e. batch-effects that are potentially significant technical sources of variance in preparing the data for meta-analysis. The authors employed the ComBat method from the sva package to address this. Related to major comment 2, please address the implication of this method in the context of bona fide (if any) biological, i.e. transcriptional study-level differences, which are potentially confounded by batch.

**Response:**

We thank the reviewer for acknowledging our attempt to address the dataset specific sources of superficial discordance. As discussed in response to comment 2, we feel confident, that the majority of the observed discordance is due to other sources than biological differences. We added a general disclosure about the possibility of confounded true biological differences by our analysis pipeline to the discussion (lines 379-384).

**Major Comment 4:**
The authors analyze the molecular phenotype of AD skin in the context of therapeutic treatment, Dupilumab, Cyclosporin, and UVB having potentially distinct mechanisms of action, utilizing 19 discriminating genes. In this important and interesting analysis, the authors describe the extent of molecular phenotype recovery by treatment in terms of an average gene-based metric. Upon inspection of Table 2, it appears that there is substantial heterogeneity of individual gene responses to specific therapies, which may have implications for these therapies and the importance of these discriminating genes. For example, MUC7, HSD11B1, and MMP3 appear affected by these treatments. Please discuss the potential implications of these important observations more thoroughly.

**Response:**

We thank the reviewer for taking notice of this important point. Table 2 highlights the diverse mechanistic effects of these therapies, and there are several interesting observations that can be taken from the data that are worth mentioning. First, MMP3, a marker of general inflammation, displays an impressive recovery with the targeted therapy dupilumab compared to the more nonspecific immune suppressants. This inflammatory gene improved over 300% in only 4 weeks of treatment with dupilumab, while long-term treatment with
cyclosporine and UVB did not reach 100% recovery. This may point to dupilumab's ability to modulate immune dysregulation in AD in a shorter timeframe than less specific agents. Another noteworthy gene is Selectin E (SELE), encoding a protein involved in leukocyte extravasation, which also showed higher levels of recovery in dupilumab compared to CsA and UVB therapy. The difference in these markers may be related to the mechanism of action of each of the drugs; UVB has direct effects on keratinocytes and thus mainly mediates signals originating there, while cyclosporine is a nonspecific inhibitor of B-cells and T-cells and all of their related processes. Dupilumab more specifically targets inflammatory processes related to the IL-4/IL-13 pathway, which has been implicated in the pathogenesis of AD, and this may account for the impressive recovery seen in genes central to the inflammatory response generated by AD. We have included a discussion of these points into the manuscript (lines 469-486).

**Minor Comment 1:**
Why weren’t down-regulated genes included in the analysis underlying Figure 1A?

**Response:**
We thank the reviewer for capturing this. In Figure 1A both up- and down-regulated DEGs were included. We changed the figure caption of Figure 1A to reflect this (|FCH|≥2, FDR≤0.05), in line 1035.

**Minor Comment 2:**
Figure 3 appears to be erroneously labeled as Figure 4.

**Response:**
We thank the reviewer for capturing this mislabeling. Figure 3 is now labeled correctly.

**Minor Comment 3:**
The supplementary methods state that hierarchical linear models were utilized to assess individual study effects while the main text methods stated paired t-tests were employed. Could you clarify?

**Response:**
We thank the reviewer for the thorough review even of this part. We use mixed-effect models to take into account the study/patients dependency structure. The
LS vs NL differences are estimated using contrasts with the parameters estimated in the model (which will be somewhat equivalent to a within study paired t-test). Sometimes we add the ‘paired t test description as a way to deal with readers that do not understand that mixed models take into account the paired structure (Yes, we got a heavy review recently criticizing our ‘lack of stats rigor’ for not using the pairdness of the data when we said we used mix models). We agree with you, this is confusing so we have changed the text accordingly (lines 148-149).

**Minor Comment 4:**
Figure 2 legend erroneously refers to panel C, which appears to describe panel B. Related to figure panel B, are the significance levels adjusted for multiple testing? A description in the figure legend or text cannot be found in reference of the right section of panel B which appears to be a schematic of gene set analysis correlation analysis of Th2 and Lipid gene sets – please clarify.

**Response:**
We thank the reviewer for this helpful comment. We have adjusted the figure legend to link to the correct panels (line 1044). The significance levels for the panels are, due to the limited sample size, not adjusted for multiple testing. We have added a supplementary table with the Benjamini-Hochberg adjusted P-values (Table E10). The schematic representation of our Th2 and Lipid gene set correlation analysis on the right of panel B is intended as an additional explanation of the heatmap on the left section of panel B. We have added a description of the schematic representation to clarify this (lines 1049-1052).

**Minor Comment 5:**
Figure 3A reports correlations of WGCNA with age, SCORAD, and IgE. Are P-values corrected for multiple testing?
Response:
We thank the reviewer for this question. The p-values in Figure 3 A) and B) are due to the minimal amount of tests, not adjusted for multiple testing. We have added Table E11 and E12 with Benjamini-Hochberg adjusted p-values for Figure 3 panel A) and B) respectively.

Furthermore, we have adjusted the title to be “Meta-analysis derived atopic dermatitis (MADAD) transcriptome defines a robust AD signature highlighting the involvement of atherosclerosis and lipid metabolism pathways”.

We hope that you will find this revised manuscript suitable for publication in *BMC Medical Genomics*, and we thank you for your consideration.

Yours sincerely,

Mayte Suárez-Farías PhD

Laboratory of Investigative Dermatology
Biostatistician, Center for Clinical and Translational Science
The Rockefeller University
1230 York Ave, Box 178
New York, NY 10065
Phone: +1(212) 327-8213

farinam@mail.rockefeller.edu
Currently at:

Dept. of Population Health Science and Policy
Dept. of Genetics and Genomics Science
Dept. of Dermatology
Icahn Institute for Genomics and Multiscale Biology
Icahn School of Medicine at Mount Sinai

1425 Madison Ave, L2-70C, Box 1077,
New York, NY 10029
P (212) 659-9678
F (212) 423-2998
mayte.suarezfarinas@mssm.edu