# Corrigendum

**Pediatric Crohn disease patients exhibit specific ileal transcriptome and microbiome signature**

Yael Haberman, Timothy L. Tickle, Phillip J. Dexheimer, Mi-Ok Kim, Dora Tang, Rebekah Karns, Robert N. Baldassano, Joshua D. Noe, Joel Rosh, James Markowitz, Melvin B. Heyman, Anne M. Griffiths, Wallace V. Crandall, David R. Mack, Susan S. Baker, Curtis Huttenhower, David J. Keljo, Jeffrey S. Hyams, Subra Kugathasan, Thomas D. Walters, Bruce Aronow, Ramnik J. Xavier, Dirk Gevers, and Lee A. Denson

Original citation: *J Clin Invest*. 2014;124(8):3617–3633. doi:10.1172/JCI75436.

Citation for this corrigendum: *J Clin Invest*. 2015;125(3):1363. doi:10.1172/JCI79657.

The authors recently completed an additional sample quality control process for the RNA-sequencing (RNA-seq) data reported in the manuscript that involved matching base calling from the RNA-seq data to the associated genotyping data and clinical metadata for each subject. This process identified 55 RNA-seq data files that were not linked to the correct clinical metadata. For 41 RNA-seq data files, the relabeling occurred within Crohn disease (CD) cases. For 6 files, the relabeling resulted in a change in diagnosis from ulcerative colitis (UC) to CD, while for 6 files, the relabeling resulted in a change in diagnosis from CD to UC. One file was relabeled within the non–inflammatory bowel diseases (non-IBD) control group, and for 1 file, relabeling resulted in a change from CD to non-IBD control. In addition, the authors identified 13 cases of sample duplication, 7 cases of African American race or IBD-undefined (IBD-U) diagnosis, and 20 cases in which the link between the RNA-seq data, genotype data, and clinical metadata could not be confirmed because of uncertain base calling. This included 26 CD cases, 12 UC cases, and 2 non-IBD control cases that must be excluded from the analysis. The data set for NCBI's Gene Expression Omnibus (GEO GSE57945) has been updated to reflect these changes. In addition, revised versions of Supplemental Table 3, Supplemental Table 12, and Supplemental Table 13 have been updated online. A revised version of Table 1 is below.

Correction of the mislabeling did not result in a significant difference in the clinical and demographic features of the cohort (Table 1). Following the correction, 1,159 of 1,281 (90%) genes from the original publication for the core ileal CD (iCD) gene list (Supplemental Table 3) were still determined to be significantly and differentially expressed between 2 independent groups of iCD versus control; 155 of 179 (87%) genes from the publication for the colon-only CD (cCD) versus UC gene list (Supplemental Table 12) were still determined to be significantly and differentially expressed between cCD and UC; and 342 of 345 (99%) genes from the publication for the iCD without deep ulcers (iCD-noDU) versus iCD with deep ulcers (iCD-DU) gene list (Supplemental Table 13) were still determined to be significantly and differentially expressed between iCD-noDU and iCD-DU. This included upregulation of *DUOX2* and suppression of *APOA1* within the top 5 up- and downregulated genes within the core ileal CD gene list and maximal suppression of *APOA1* within the cCD versus UC comparison and the iCD-DU versus iCD-noDU comparison.

The authors regret the errors.

**Table 1. RISK RNA-seq cohort clinical and demographic characteristics**

| | Ctl *n* = 42 | UC1 *n* = 38 | cCD1 *n* = 33 | UC2 *n* = 24 | cCD2 *n* = 23 | iCD1 *n* = 81 | iCD2 *n* = 81 | All iCD *n* = 162 | iCD-DU *n* = 76 | iCD-noDU *n* = 86 |
|---|---|---|---|---|---|---|---|---|---|---|
| Mean (SD) age (yr) | 11 (3) | 12 (3) | 12 (3) | 14 (3) | 13 (3) | 12 (3) | 12 (3) | 12 (3) | 12 (3) | 12 (3) |
| Male sex (%) | 62 | 47 | 52 | 71 | 61 | 60 | 57 | 59 | 58 | 60 |
| MED ethnicity (3 of 4 grandparents) (%) | 95 | 88 | 91 | 88 | 83 | 85 | 88 | 86 | 89 | 84 |
| Perianal involvement (%) | 0 | 0 | 18 | 0 | 26 | 20 | 14 | 17 | 17 | 17 |
| Ileal deep ulcers (%) | 0 | 0 | 0 | 0 | 0 | 52 | 42 | 47 | 100 | 0 |
| Body mass index Z<-2 (%) | 0 | 3 | 18 | 4 | 17 | 20 | 17 | 19 | 23 | 15 |
| **PCDAI at diagnosis** | | | | | | | | | | |
| ≤10 (inactive, %) | NA | NA | 13 | NA | 5 | 11 | 9 | 10 | 13 | 7 |
| 11 to 30 (mild, %) | NA | NA | 32 | NA | 33 | 35 | 47 | 41 | 32 | 49[A] |
| >30 (moderate-severe, %) | NA | NA | 55 | NA | 62 | 54 | 44 | 49 | 55 | |

Differences between selected groups were tested by ANOVA for continuous variables and $\chi^2$ for dichotomous variables. Ctl, control; MED, mixed European descent. [A]$P = 0.03$ vs. iCD-DU.