Author's response to reviews

**Title:** Batch correction of microarray data substantially improves the identification of genes differentially expressed in Rheumatoid Arthritis and Osteoarthritis

**Authors:**

Peter Kupfer (peter.kupfer@hki-jena.de)
Reinhard Guthke (reinhard.guthke@hki-jena.de)
Dirk Pohlers (d.pohlers@skc.de)
René Huber (huber.rene@mh-hannover.de)
Drik Koczan (dirk.koczan@med.uni-rostock.de)
Raimund W. Kinne (raimund.w.kinne@med.uni-jena.de)

**Version:** 3  **Date:** 2 May 2012

**Author's response to reviews:** see over
Dear Tim Sands

Please find enclosed our revised manuscript entitled “Batch correction of microarray data substantially improves the identification of genes differentially expressed in Rheumatoid Arthritis and Osteoarthritis.”

We appreciate the thoughtful comments of the referees and have introduced the changes highlighted in the text.

Point-by-point response to the reviewer comments:

Reviewer#1
We thank Reviewer#1 for his/her encouraging statements and valuable comments.

Ad 1.) There seems to be a typo on page 10 (‘GPower: Literatur (Uni Düsseldorf)’)

Unfortunately, we missed this error and we want to apologize for the mistake. We have now included the information about the tool “GPower” in the methods section (page 10, paragraph 3), including the applied values and 2 new references.

Reviewer#2
We thank Reviewer#2 for his/her positive evaluation and valuable comments.

Ad 1.) The top of page 10 contains the single line “GPower: Literatur (Uni Dusseldorf)”. I guess that there should be more here? One of my original criticisms of this paper was a lack of power calculations. It seems that these have been done (as contained in table S2) however the methods do not give any information on how these were performed using GPower. I guess this single line was an oversight and there should be more here?

We want to apologize for this mistake once again (see also response to reviewer #1). We have now included the information about the tool “GPower” in the methods section (page 10, paragraph 3), including the applied values and 2 new references. In addition, we have extended Table S2 to now contain the means and standard deviations for the genes of interest. These values were used to calculate the power before and after the batch correction with the tool GPower.

Ad 2.) Reference has now been made to multiple testing [beginning of page 10] however the False Discovery rate (FDR) used has not been declared? This is not the same as an adjusted p-value, as it is a procedure (the rate is set in advance and a set of variables are declared as having passed at this rate). If the FDR has been used the authors should be clear about what rate was set. It could be that q-values has been taken and a threshold of 0.05 set, but this has not been stated.
At the beginning of page 10, we have now attempted to clarify this issue by describing the thresholds for the changes in gene expression (2-fold) and for the q-value (0.05). In addition, the q-value has now been clearly described as the ‘FDR adjusted p-value’.

I hope that the present changes will now render the manuscript acceptable for publication in BMC Medical Genomics and remain yours sincerely, on behalf of the authors.

Raimund W. Kinne