Reviewer's report

Title: Genetic interaction of GSH metabolic pathway genes in Cystic Fibrosis

Version: 3 Date: 4 February 2013

Reviewer: Frauke Stanke

Reviewer's report:

The manuscript “Genetic interaction of GSH metabolic pathway genes in Cystic Fibrosis” by de Lima Marson and colleagues describes the association of polymorphisms in four different genes from the GST metabolic pathway in patients with cystic fibrosis.

The reviewer thanks the authors for the extensive modifications of their manuscript made in this revised version. However, a few points still remain to be clarified. The reviewer is convinced that most of these points, some of which were raised before, got lost or were insufficiently addressed due to the language barrier that makes communication between two non-native speaking parties so difficult. If this is the case, the reviewer apologizes for raising points repetitively. However, for the sake of the readers of The Journal, these should be clarified in the final version to enable other researchers to fully appreciate the data.

Comment 1: - a non-confidential comment to the editorial office: I have not assessed the manuscript for typographical or grammatical errors in order to keep the deadline, trusting that BMC will check for these. Example: page four : “In our group, as previously publish, …” must read “In our group, as previously publishED, …”; or: page 5: “Polymorphisms - 129C>T and -3506A>G, located in GCLC gene promoter region, have been associated to GSH reduced production [15;25].” Is correctly phrased as “Polymorphisms - 129C>T and -3506A>G, located in GCLC gene promoter region, have been associated with reduced GSH production [15;25].”

Comment 2: Principle of the GST deletion genotyping assay – text currently reads: “For the GSTM1 and GSTT1 genes deletion, a multiplex PCR reaction was performed and the CYP1A1 gene was included as an internal reaction control [34]. The multiplex PCR performed target the wildtype and the deleted allele at GSTM1 and GSTT1 simultaneously. As GSTM1 and GSTT1 deletion analyze not provide complete genotypes, minor allele frequencies and Hardy-Weinberg Equilibrium cannot be evaluated for these loci in our study. By the method realized we have two groups of GSTM1 and GSTT1 gene: (i) patients homozygous to gene deletion; (2) patients with at least an allele expressed.” The reviewer assumes that the assay tests for the wild type allele, hence samples for which no signal is obtained are homozygous for the deletion. Heterozygotes and homozygotes for the wild-type allele cannot be differentiated. If this is technically correct, the phrasing “The multiplex PCR performed target the wildtype and the deleted allele at GSTM1 and GSTT1 simultaneously.” is wrong – the assay only
targets the wild-type allele, but it does so for the GSTM1 and GSTT1 locus at the same time. Please reword accordingly.

Comment 3: Patient population – text currently reads: “Other genes were studied in the same reference center with similar population as previously published [11; 13; 14]. The GSTM1 and GSTT1 genes were studied in our center considering other cystic fibrosis patient group as recent data showed [12].” The wording is unclear – are the patient populations from references 11, 13, 14 dissimilar from the population in 12? If so, how are the populations different? Please note that it is a strong argument for the power of the sample and the power of the phenotype to describe a modifier, if this has been successfully done before.

Comment 4: data display with respect to the observed genotype-phenotype-association – text on page 10 reads: “Multifactor Dimensionality Reduction analysis showed evidence of interaction of GSTM1 and GSTT1 genes deletion, GSTP1*+313A>G, and mutations in CFTR gene (p=0.008) and BS clinical score (Table 3). All data was previously associated with a point-point analysis considering which gene polymorphism in association with one clinical manifestation at a time by CFTR groups, and after Bonferroni correction any pvalue was positive (data not showed).” The data need to be shown – firstly, because they substantiate the major finding of the article, secondly because the description in the text is very hard to understand. The reviewer assumes that the authors have followed a multiple step approach to their data. The reviewer assumes that each of the genes was tested independently for association with each of the phenotypes. The reviewer furthermore assumes that the association of the genes GSTM1, GSTT1, GSTP1 and CFTR passed the significance threshold after Bonferroni-correction, hence the authors choose to display this finding in a single step. If this is correct, it would be extremely helpful if the primary data, including raw & corrected p-values is provided. It is likely that figure 1 already shows this data, albeit without the primary raw p-values, in a more complex manner (see comment below). Maybe the information can be provided in a separate table? Maybe figure 1 becomes redundant as a consequence? Please check carefully which format will display the data in an easy-to-grasp manner, irrespectively of the suggestion of the software used for evaluation.

Comment 5: table 2 – data on GSTM1 and GSTT1 lists no genotypes. However, even though wt/del and wt/wt cannot be discriminated, the primary data should be given here, summarizing the appropriate columns.

Comment 6: table 2 – data on GCLC, rs17883901. This shows deviation from HWE, which should not be ignored. As this has been addressed before, here is the previous comment and the author’s answer:

Previous comment with respect to deviation from HWE: “….. the lack of HWE for the GCLC polymorphism emphasizes the point 6. – how do the authors interpret the deviation from HWE in their sample if not by a survivor effect?” ANSWER “The Brazilian population is admixed being a factor associated with different values for the Hardy-Weinberg
equilibrium. In addition, the sample size despite being representative for statistical analyzes may impair the description associated with genes that exhibit some minor allele frequency in the population as the GCLC rs17883901 which has the minor frequency to MAF in our study. The polymorphism can have a historical importance in the population, but to the clinical markers included and in the actual context, the polymorphism did not show association in our sample study, being a difficult problem to understand and describe.” The authors explain the deviation from HWE with population admixture, which can be the case. However, alternatively the CF sample population might be enriched for carriers of a mild allele at GCLC rs17883901, and at the same time the CF sample population might be depleted from a risk allele at GCLC rs17883901. As a consequence, no genotype-phenotype-correlation is observed: the risk allele, assumed to be associated with the severe phenotype, will be depleted from the entire population sample in case of a survivor bias. This might be distinguished when the allele frequencies at GCLC rs17883901 in other population samples (e.g. from the 1000genomes project or from HapMap) are compared. Is the MAF equally low in other non-CF populations? That would be an argument for the population admixture scenario. Is the MAF in the investigated Brasilian CF population lower than in all other populations? That would be an argument for the selection bias / risk allele depleted theory. To clarify: The reviewer assumes that GCLC, rs17883901 is a modifier for cystic fibrosis that cannot be detected by genotype-phenotype-association as the risk allele is missing from the studied population. Or still in other words: p < 0.005 is significant enough to be discussed in more detail.

Comment 7: table 3 – text reads “CFTR*GCLC(129C/T)*GSTP1*GSTT1*GSTM1” – this is the only line which reads GCLC(129C/T) and not GCLC. Why?

Comment 8: Throughout the manuscript, it is best to stick to one nomenclature for the SNPs (either rs or 129C/T).

Comment 9: Figure 1 and legend to figure 1 – several comments. As it is this figure cannot be interpreted.
A – shading in light grey and dark grey are very hard to discriminate. Suggestion: change to dark grey vs. white.
B – the following phrasing is awkward: “…only the gene with positive interaction with the BS was showed.” (is shown?) and “with the GSTP1*+313A>G polymorphism in a short association.” (There is no such thing as a short association, the opposites are “strong” and “weak”)
C – general: please explain in a case example what is visualized. For instance, the strongest difference between the left and the right half of the figure are 2 vs 8 (left side) and 13 vs 3 (right side) individuals. These are ….?

Comment 10: Figure 2 and legend to figure 2 – the scale for A is missing.
Level of interest: An article of importance in its field

Quality of written English: Needs some language corrections before being published

Statistical review: Yes, and I have assessed the statistics in my report.

Declaration of competing interests:

I declare that I have no competing interests