Author's response to reviews

Title: A combined blood based gene expression and plasma protein abundance signature for diagnosis of epithelial ovarian cancer - A study of the OVCAD consortium

Authors:

Dietmar Pils (dietmar.pils@univie.ac.at)
Dan Tong (dan.cacsire-castillo@meduniwien.ac.at)
Gudrun Hager (gudrunhager@ymail.com)
Eva Obermayr (eva.obermayr@meduniwien.ac.at)
Stefanie Aust (stefanie.aust@meduniwien.ac.at)
Georg Heinze (georg.heinze@meduniwien.ac.at)
Maria Kohl (maria.kohl@meduniwien.ac.at)
Eva Schuster (eva.schuster@meduniwien.ac.at)
Andrea Wolf (andrea.wolf@meduniwien.ac.at)
Jalid Sehouli (jalid.sehouli@charite.de)
Ioana Braicu (ioana@braicu.de)
Ignace Vergote (ignace.vergote@uz.kuleuven.ac.be)
Toon Van Gorp (toon.vangorp@hotmail.com)
Sven Mahner (s.mahner@uke.de)
Nicole Concin (Nicole.Concin@i-med.ac.at)
Paul Speiser (paul.speiser@meduniwien.ac.at)
Robert Zeillinger (Robert.Zeillinger@meduniwien.ac.at)

Version: 3 Date: 17 December 2012

Author's response to reviews: see over
Answer to reviewers:

Many thanks to the reviewer for their reasonable objections and comments. We are pleased to follow each of them.

Reviewer: Sergei Moshkovskii

1. One general problem of the paper which cannot be fixed in the submission is the ratio between FIGO I-II and III-IV stage patients recruited for tests. It is 33 vs. 430, less than 8% early stage. I understand that collection of early ovarian cancer is very difficult. However, reporting data of such pooled set means a substantial bias toward III-IV stage. Despite data are of novelty and importance, the ratio between FIGO I-II and III-IV should be clearly announced to readers elsewhere, inter alia, in the abstract.

The FIGO numbers are now announced in the abstract, the methods section and Table 1, which was merged with Supplementary Table 1. In addition, this apparent shortcoming of this manuscript is now mentioned in the discussion.

2. The separate statistical analysis of FIGO I-II stage group is done, but it is described too briefly. It is interesting to know what specific RNA and protein entities significantly changed in early cancer.

There are now gene expression changes for FIGO I/II and FIGO III/IV samples shown separately in Table 4A and protein abundance differences for FIGO I/II and FIGO III/IV samples compared to control samples in Table 4B and corresponding boxplots in Figure 4.

3. Protein analysis is described purely. In order to know the list of proteins analyzed readers should go to the external references. If we have only 6 protein variables, the descriptive statistics on each of them should be shown, at least in the supporting data. If authors exclude some protein variables from the study, as I could hardly understand, they should note what exact protein(s) is(are) excluded.
In Table 6 is the complete information of all discriminative models built for this manuscript, including the information which proteins are missing in each concrete model. Descriptive statistics of the six proteins are shown in Table 4B and boxplots in Figure 4. Some proteins were excluded from some discriminative models by the model building process (L1 penalized logistic regressions), which uses internal cross-validation to find the best model.
Reviewer: Ian Cree

- The supplementary files are tables that should be part of the main paper - this is an online journal and the few kilobytes this would take are insignificant. The histological make up of the group should be stated in the paper and whether histology was centrally reviewed.

  Both Suppl. Figures are now included in the main manuscript. Histology was not reviewed centrally, but by the specialized pathologists of the five participating hospitals (which are all university hospitals with ample experience in gynecologic cancer) according to reviewed OVCAD criteria. This is now stated in the manuscript.

- The list of proteins tested should be clearly stated earlier in the paper - it was not until opening the supplementary table 2 that they were obvious. Reference in the methods section to a table of the genes and proteins tested would help.

  A list of genes and proteins is now provided in the abstract and the methods section.

- The authors correctly state that they did not look at other disease states and this could be discussed further: it is important in my view.

  This is now discussed more deeply in the discussion section.

- Figure 3 is the only place in the paper which I could begin to work out whether protein or genes should be tested - and the difference between protein alone v. protein + genes seems small. Is this truly the case, are the tests completely independent, and which would you test first in patients? Could one develop a clinical decision rule for testing to avoid the costs involved in testing large numbers of women with both methods?

  We think that the difference between proteins alone (AUC: 0.973) and the models with either 5 gene expression plus 5 protein values or the model with 13 gene expression and 6 protein values is tremendous (each AUC: 0.998), given the demand to have very high sensitivities and specificities for early detection of ovarian cancer, a rather rare disease. We would recommend doing both types of tests (gene expression and plasma protein abundance) together, because the high discriminative power comes from the combined model of gene expression values and plasma protein abundances. We think that the high sensitivity comes from the gene expression changes in the leukocytes fraction and specificity from the proteins (which are partly ovarian tissue specific).

  The idea to begin with the 13 gene expression test (which will be the cheaper test, given the availability of a RT-qPCR machine) and test only the persons with clear or questionable results further with the protein test would be a possibility, but this should be evaluated in an independent patient and control cohort, including patients with other (malignant) diseases.
- There are occasional minor typographical errors - for instance, 'seeked' instead of 'sought' in the first paragraph of the abstract.

  Corrected, as far as we found them.