Author's response to reviews

Title: Integrated mutation, copy number and expression profiling in resectable non-small cell lung cancer

Authors:

Genni M Newnham (Genni.Newnham@svhm.org.au)
Matthew Conron (Matthew.Conron@svhm.org.au)
Sue Anne McLachlan (SueAnneMcLachlan@svhm.org.au)
Alexander Dobrovic (Alexander.Dobrovic@petermac.org)
HongDo Do (Hongdo.Do@petermac.org)
Jason Li (Jason.Li@petermac.org)
Kenneth Opeskin (Kenneth.Opeskin@svhm.org.au)
Natalie J Thompson (natalie.j.thompson@gmail.com)
Gavin Wright (Gavin.Wright@svhm.org.au)
David M Thomas (David.Thomas@petermac.org)

Version: 3 Date: 11 February 2011

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

Title: Integrated mutation, copy number and expression profiling in resectable non-small cell lung cancer

Authors:

Genni M Newnham (Genni.Newnham@svhm.org.au) Matthew Conron (Matthew.Conron@svhm.org.au) Sue Anne McLachlan (SueAnneMcLachlan@svhm.org.au) Alexander Dobrovic (Alexander.Dobrovic@petermac.org) HongDo Do (Hongdo.Do@petermac.org) Jason Li (Jason.Li@petermac.org) Kenneth Opeskin (Kenneth.Opeskin@svhm.org.au) Natalie J Thompson (natalie.j.thompson@gmail.com) Gavin Wright (Gavin.Wright@svhm.org.au) David M Thomas (David.Thomas@petermac.org)

Version: 2 Date: 12 November 2010

Author's response to reviews: see over
Integrated mutation, copy number and expression profiling in resectable non-small cell lung cancer

Please accept this revised manuscript with changes made according to reviewer reports. Each reviewer comment has been specifically addressed below.

**Response to reviewer reports**

**Reviewer 1**

1. Data submitted to GEO (GEO ID to be provided to us within 1 week).
2. T-statistic: LIMMA was used (now described in methods section).
3. Figure 1 removed. Table 1 referenced in results section.
4. Description of breakpoints included in results (aneuploidy) section.
5. It is not considered appropriate to list all instances of CNV due to large numbers.
6. Incorrect use of the term ‘correlation’ has been altered in headings.
7. Group comparisons: Sample sizes have been entered for each group comparison discussed, if they were not already included. We have stated in the manuscript that due to low sample numbers specific comparisons were not made (eg. EGFR mutant vs. Wild-type tumours).

8. External dataset comparisons: The manuscript has been altered to include a more detailed description of the GSE11117 dataset and the methods used (Correlation with external datasets sub-section in Results section).

9. Integration of datasets: We accept that integration was in the form of comparing expression levels and copy numbers of genes at specific loci. We believe this approach is valid and has identified several regions of interest.

10. Validation experiments would constitute a further body of work.

**Reviewer 2**

**Major**

1. P value is corrected for multiple testing using LIMMA method as now described in methods section.
2. Paragraph describing the integration of genomic and transcriptional profiles has been added to Methods section.

**Minor**

1. Incorrect use of the term ‘correlation’ has been altered in headings.
Major

1. Method of predicting survival in external dataset comparisons: The reviewer has suggested that a machine learning algorithm would be a preferable method of testing the ability of our predictive gene signature to classify an external dataset. However, we were comparing data generated on different platforms and intensity levels of expression are known to differ among platforms. For this reason creating a classifier based on actual data from one platform is unlikely to generalise to others. A clustering approach, on the other hand, relies only the gene lists created - the ability of separating two groups is unaffected by cross-platform issues. As long as the observed intensities are consistent within a platform, the resulting clusters will truly represent a separation of transcriptional profile across the given gene list. We understand evaluating different clustering methods and parameters can help optimise the observed significance of our gene list in separating between recurrence/nonrecurrence. However, we have already demonstrated significance using a straightforward correlation-based clustering approach, and we believe that is sufficient for the current work as the key message here is the reliability of the gene lists.

2. Data submitted to GEO (GEO ID to be provided to us within 1 week).

Minor

1. Annotation of genome in Figures 2 and 3 (now figures 1 and 2) has been altered so as to allow better identification of chromosomes.

2. Figure and table legends included.

3. Fold changes and p values added to table 3. Fold changes already in table 4 (log2 difference in expression between groups). P values added to table 4.

4. P values added to table 5.

5. Incorrect reference to figure 1 removed.

6. Tables and figures have been included in landscape orientation as it allows viewing and printing at full size. If desired by the editor this can be altered.