Review of: "Proteogenomics analysis to identify acquired resistance-specific alterations in melanoma PDXs on MAPKi therapy"

Teck Yew Low[1]

1 Universiti Kebangsaan Malaysia

Potential competing interests: The author(s) declared that no potential competing interests exist.

This is a very interesting manuscript. I have the following comments:

1. The authors might want to define Resistance-Specific Alterations (RSA) more clearly. Are differences in the expression level of a protein or transcript between resistant and vehicle samples sufficient to define them? Or expression differences at both protein and transcript in the same direction is needed? Or that variant sequence is needed to be present?
2. The authors mentioned: "The RNA-seq reads at $NRAS^{Q61L}$ the following percentages of T and A: R1- 76% A, 24% T; R2-87% A, 12% T; R4- 84% A, 16% T; R5- 79% A 21% T; V1- 60% A, 40% T.

>>Does that mean in the original PDX, the Q61L mutation happens only in one of the NRAS alleles? And in the resistant tissues, either there were conversion of T>A; or simply the allele containing A rather than T, is overexpressed ?

3. Proteomic Analysis Method:

(i) Custom Reference Database not provided

(ii) No info as to Raw MS data/ analyzed data not were deposited in Proteomic Repository

(iii) There are no details on the custom database (how many entries etc)

(iv) MaxQuant search parameters not provided

(v) FDR for at peptide and protein levels not provided?

4. From Figure 1.

(i) What is inside the Custom Reference Database? Only variant sequences, or also those canonical protein sequences?

(ii) The workflow in Figure 1A seems to imply two separate MS database search, i.e.(i) Normal and Custom Reference Databases; then combining the search results via multi-omics integration. I wonder if it is more appropriate to search an integrated database which containing both canonical and variant sequences? Would that help to arrive at a more realistic estimation of FDR than performing two separate searches and later combine the results? For example, the integrated database can be tailored to contain two different alleles of NRAS i.e. NRAS and NRAS Q61L.

(iii) Would it be better if the PDX sample on Day 42 (before MEKi treatment starts) was also included in the proteogenomics analysis?

(iv) There was only 1 vehicle sample, could that affect the statistics?

(v) Why is the end-point for Vehicle sample earlier that the resistant samples?