Response to Comment on “A subset of HLA-I peptides are not genomically templated: Evidence for cis- and trans-spliced peptide ligands”

Pouya Faridi1, Chen Li1,2, Sri H. Ramarathinam1, Patricia T. Illing1, Nicole A. Mifsud1, Rochelle Ayala1, Jiangning Song1,3, Linden J. Gearing4, Nathan P. Croft1*, Anthony W. Purcell1*

This is our response to the Technical Comment by Rolfs et al. where we point out errors in their reanalysis of our data.

The Technical Comment by Rolfs et al. in response to our publication (1) uses additional approaches to evaluate the accuracy of the sequences of cis- and trans-spliced peptides reported in our study. To clarify, the message and conclusions in our paper are that (i) nongenomically encoded peptides, which we believe are best explained as cis- and trans-spliced peptides, comprise a considerable, yet varying, proportion of the immunopeptidome of a range of human leukocyte antigen–A (HLA-A) and HLA-B alleles [combined, these proportions are close to that previously published by Liepe et al. (2) in Science] and (ii) we describe an accessible informatics workflow for their identification from mass spectrometry data of peptide ligands extracted from purified HLA complexes. (iii) The unique nature of HLA-bound peptide diversity is also reflected by the absence of spliced peptide identification in trypsin and elastase (a broad specificity protease) digests of whole-cell protein digests. Our examination of the reanalysis of our data by Rolfs et al. suggest that they did not completely understand the context of the experiments and the analysis we performed, and thus, their secondhand analysis of the data has led them to incorrect assertions as detailed below.

HYDROPHOBICITY ANALYSIS

Rolfs et al. used hydrophobicity index (HI) prediction to compare nonspliced and spliced peptide populations. They report that nonspliced peptide sequence identifications had tighter correlation with HI prediction values than the spliced sequences reported in our study. We present our analysis of HI prediction in Fig. 1B for the HLA-A*01:01 full dataset and in Fig. 1C for HLA-B*57:01, our largest dataset, both of which show a similar, albeit broader, correlation of peptide HI prediction for both linear and spliced peptides.

REANALYSIS OF OUR DATA USING AN ALTERNATIVE PEAKS WORKFLOW

Rolfs et al. compared the PEAKS score for linear and spliced peptides to conclude that PEAKS scores for spliced peptides are overall lower to those for contiguous peptide identifications. However, Rolfs et al. state that they “extracted all peptides reported by Faridi et al. for the HLA-A*01:01 sample,” but as discussed above, not all data were extracted. Further, it is not clear which score is referred to, as the text states “PEAKS score” (which is the score of the PEAKS DB search and reported in the pep.xml file), but their Fig. 1B legend states the use of the “PEAKS de novo score,” which gives an average local confidence (ALC) of the peptide sequence independent of a database search. If we extract PEAKS score as shown in Fig. 1D, then the ratio of the distribution of linear and spliced peptides varies across the 1% false discovery rate (FDR) PEAKS score. Clearly, the subset of data analyzed by Rolfs et al. does not accurately represent the distribution in the entire dataset.
ROLFS ET AL. ALSO PROPOSED USING AN “ALTERNATIVE MEANS TO CALCULATE FALSE DISCOVERY RATE”

The calculation of FDRs for nontryptic peptides has been the subject of much debate (4). We used the target-decoy method, which is a generally accepted approach (5).

Rolfs et al. suggest an alternate means to calculate FDR that is inconsistent with our careful analysis of assignments even for nonspliced peptide sequences. The quality of a de novo sequence can be assessed using an ALC scoring function. In our study, we determined a threshold for the ALC score (per dataset) to allow inclusion of a sequence from de novo sequencing into our database for final researching of the data. This was rationalized using the ALC score of nonspliced peptides that were also identified using a standard database search at a 1% FDR threshold for the de novo only candidates. As noted above, the HLA-A*01:01 dataset was incompletely analyzed by Rolfs et al., so we cannot evaluate the results for this dataset in more depth. If we focus on Rolfs et al.’s reanalysis of the HLA-A*02:03 dataset, they reported an FDR of 2.53% even for the highest possible scoring de novo sequences (i.e., ALC score of 99). Such overly stringent treatment of the data would also invalidate all of the linear sequences that we reported and that have been corroborated in previous independent studies.

To further address these concerns, we have calculated FDR by an alternate means to assess its impact on our own data. We have extracted the PEAKS confidence score (−10logP) that was used for the calculation of 1% FDR in our first PEAKS search (the database for this search only contains the reference human proteome and no spliced peptides—i.e., what Rolfs et al. refer to as the “gold standard”). If we then use this PEAKS score for extraction of peptides at 1% FDR in the second search (i.e., against the conventional human proteome database and spliced peptides), then this does not affect the proportion of spliced peptides any of the datasets (Fig. 1E).

ALTERNATE SEARCH ENGINE ANALYSIS

Rolfs et al. reanalyzed our data using the alternate search engine “Comet.” This is not a new approach; in fact, this was one of the validation steps we performed in our paper, using the SEQUEST search engine in addition to PEAKS. In our hands, using different search engines, we did not observe any notable effect on the ratio between linear and spliced peptides in any of the datasets analyzed. A critical point for comparing the results of two search engines is to use the same settings for the search. Even the authors noted that “Comet was run with the search settings reported by Faridi et al.”
that up to 6% (8) of cis-spliced and potentially much higher proportions of trans-spliced peptides (9) may be present in the immunopeptidome. We believe that spliced peptides exist at a high prevalence, and we have recently shown their role in tumor antigen recognition (10).

J. Vivian, P. Hertzog, N. Ternette, and J. Rossjohn felt that because of the technical nature of this response, their co-authorship was not warranted. They all stand by the conclusions made in the paper.

REFERENCES


Submitted 30 January 2019
Accepted 15 July 2019
Published 16 August 2019
10.1126/sciimmunol.aaw8457

Response to Comment on "A subset of HLA-I peptides are not genomically templated: Evidence for cis- and trans-spliced peptide ligands"

Pouya Faridi, Chen Li, Sri H. Ramarathinam, Patricia T. Illing, Nicole A. Mifsud, Rochelle Ayala, Jiangning Song, Linden J. Gearing, Nathan P. Croft and Anthony W. Purcell

DOI: 10.1126/sciimmunol.aaw8457