

ANTIGEN PRESENTATION

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 Li^{1,2}, Sri H. Ramarathinam¹, Patricia T. Illing¹, Nicole A. Mifsud¹, Rochelle Ayala¹, Jiangning Song^{1,3}, Linden J. Gearing⁴, Nathan P. Croft^{1*}, Anthony W. Purcell^{1*}

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) digestions 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 non-spliced sequence identifications had tighter correlation with HI prediction values than the spliced sequences reported in our study. Our experience is that these predictions are not always accurate and are quite HLA allotype dependent (see Fig. 1, A to C).

We have two comments to this analysis. First, HI predictions (SSRCalc) have been designed for proteomics studies and trained almost exclusively on tryptic peptides (bearing Lys or Arg at the C terminus) with carbamidomethylation of cysteine residues. These sequence features and fixed modifications are not relevant for HLA peptides, which rarely have basic termini and are not routinely re-

duced and alkylated during their isolation and analysis (3). The authors have in fact assumed that all cysteine residues are carbamidomethylated in our data, which is not the case. Thus, the HI predictions are incorrect for all peptides containing cysteine (>20% of spliced peptides for some allomorphs).

Second, Rolfs *et al.* did not extract all relevant data for the HLA allomorphs in question because they did not realize that there were several fractions and liquid chromatography–tandem mass spectrometry (LC-MS/MS) files per sample. As we noted in our methods, we used high-performance liquid chromatography for prefractionation of our samples before analysis by LC-MS/MS, resulting in several LC-MS/MS acquisition runs per HLA allomorph. This means that the scan number is not unique and can be assigned to a different peptide sequence in different fractions (Fig. 1A). As the fraction number was not taken into account by Rolfs *et al.*, some degree of inaccuracy of the ranking of peptide sequences can be expected. Furthermore, Rolfs *et al.* extracted only a subset of HLA allomorph specific data from the pep.xml files, missing over half of the scans, which will have not only a direct effect on visual inspection of the data but also consequences for subsequent analysis of the number of spliced peptide sequences in each dataset. 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.

¹Infection and Immunity Program and Department of Biochemistry and Molecular Biology, Biomedicine Discovery Institute, Monash University, Clayton, Victoria 3800, Australia. ²Department of Biology, Institute of Molecular Systems Biology, ETH-Zürich, Zürich 8093, Switzerland. ³Centre for Data Science, Faculty of Information Technology, Monash University, Melbourne, Victoria 3800, Australia. ⁴Centre for Innate Immunity and Infectious Diseases, Hudson Institute of Medical Research and Department of Molecular and Translational Science, School of Clinical Science, Monash University, Clayton, Victoria 3168, Australia.

*Corresponding author. Email: anthony.purcell@monash.edu (A.W.P.); nathan.croft@monash.edu (N.P.C.)

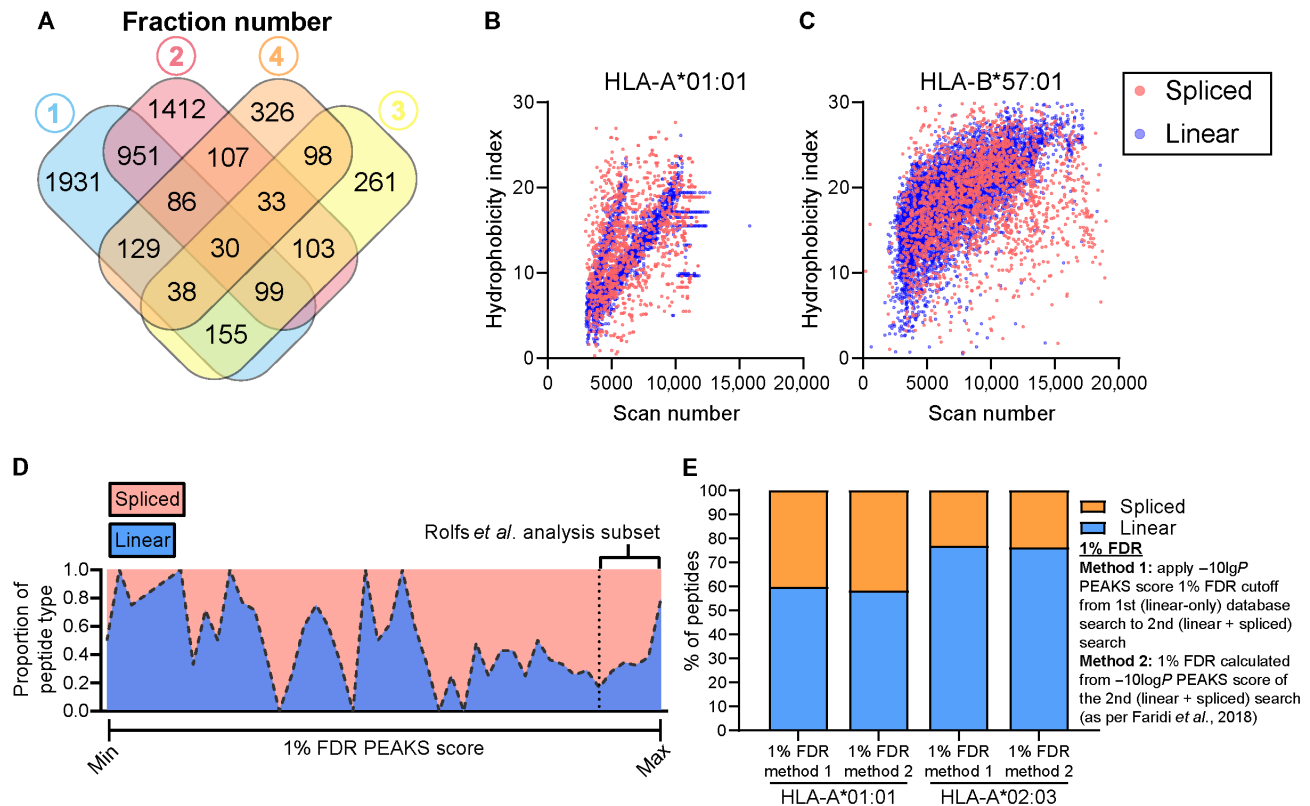


Fig. 1. Selected charts associated with our response to Rolfs *et al.*'s reanalysis of our data. (A) Venn diagram representation of the overlap (indicated in each segment) of scan numbers across four fractions from the same sample. This highlights the potential for incorrect sequence assignments if a fraction number is not associated with each scan number. (B) Correlation between scan number and HI in the complete HLA-A*01:01 dataset. As shown in this figure, there is not a clear correlation between scan number and HI for linear or spliced peptides, and (C) the performance of HI prediction is variable depending on the HLA allomorph selected, with HLA-B*57:01 peptides showing a higher correlation. (D) Ratio of the number of linear and spliced peptides across different PEAKS scores within the 1% FDR cutoff. The subset chosen by Rolfs *et al.* is therefore biased toward linear peptides, and it is not representative of the overall distribution. (E) Effect of applying two different FDR calculation methods to the distribution of linear and spliced peptides from the HLA-A*01:01 and HLA-A*02:03 datasets. In method 1, the $-10\log P$ PEAKS score, corresponding to a 1% FDR cutoff from the first database search (i.e., against the conventional human proteome database), was applied to the output from the second database search (i.e., containing the conventional human proteome database and spliced peptides). Method 2 is the original method we used in (7).

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 non-spliced 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 and adopting an equivalent cutoff score 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 ($-10\log P$) 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 proteome plus 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.*,"

but the “Comet search settings” document provided by Rolfs *et al.* indicated several discrepancies. First, a 20 parts per million (ppm) mass tolerance for precursors was used compared with 15 ppm in our own study. Second, an additional PTM (N-terminal acetylation) was included, and last, the incorrect fragmentation modality of higher-energy collisional dissociation (HCD) instead of collision-induced dissociation (CID) was defined in the engine. It is not clear from the Technical Comment and provided details which database was used for searching the data and how it was constructed. However, even using these incorrect settings, when we exported peptides from Rolfs *et al.*'s pep.xml files for the HLA-A*01:01 dataset, we found around 13% spliced peptides (154 of 1185 peptides) at 1% FDR, so it remains unclear how Rolfs *et al.* cannot find any spliced peptides in their own search. Again, to address this concern, we have reanalyzed our data with Comet with the same settings and database as we used for PEAKS 8.5 in our original paper. We have applied this approach for the HLA-A*01:01 and HLA-B*07:02 datasets, and we found 3036 (25.2% spliced) and 2774 (34.1% spliced) 8- to 12-amino acid peptides, respectively (data supplied for review purposes).

MOTIF ANALYSIS AND SYNTHETIC PEPTIDE VALIDATIONS

To clarify, we did not use motif analysis in our paper to validate spliced peptides. Actually, that linear and spliced peptides share similar motifs is one of the findings in our paper.

Rolfs *et al.* claim “significant differences” between the experimental spectrum and validation spectrum using retrospectively synthesized peptide. We disagree—in addition to visual inspection, in our study, we more objectively calculated Pearson correlations between the sets of log intensities of the observed *b* and *y* ions of eluted peptides and their retrospectively synthesized counterparts. Our analysis showed that paired spectra are statistically similar with matching retention time and fragmentation spectra.

In general, we do not believe that the reanalysis of our data by Rolfs *et al.* had been performed with the necessary accuracy and that, therefore, their comments are not substantiated. Their conclusions are not only in contrast with that of new studies from various independent groups (6, 7) but also in contrast with their own findings 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

1. P. Faridi, C. Li, S. H. Ramarathinam, J. P. Vivian, P. T. Illing, N. A. Mifsud, R. Ayala, J. Song, L. J. Gearing, P. J. Hertzog, N. Ternette, J. Rossjohn, N. P. Croft, A. W. Purcell, A subset of HLA-I peptides are not genomically templated: Evidence for cis- and trans-spliced peptide ligands. *Sci. Immunol.* **3**, eaar3947 (2018).
2. J. Liepe, F. Marino, J. Sidney, A. Jeko, D. E. Bunting, A. Sette, P. M. Klotzel, M. P. H. Stumpf, A. J. R. Heck, M. Mishto, A large fraction of HLA class I ligands are proteasome-generated spliced peptides. *Science* **354**, 354–358 (2016).
3. O. V. Krokhin, V. Spicer, Predicting peptide retention times for proteomics. *Curr. Protoc. Bioinformatics* **31**, 13.14.1–13.1415 (2010).
4. P. Faridi, A. W. Purcell, N. P. Croft, In immunopeptidomics we need a sniper instead of a shotgun. *Proteomics* **18**, e1700464 (2018).
5. J. E. Elias, S. P. Gygi, Target-decoy search strategy for mass spectrometry-based proteomics. *Methods Mol. Biol.* **604**, 55–71 (2010).
6. J. Liepe, J. Sidney, F. K. M. Lorenz, A. Sette, M. Mishto, Mapping the MHC class I spliced immunopeptidome of cancer cells. *Cancer Immunol. Res.* **7**, 62–76 (2019).
7. T. Aaron Wiles, R. Powell, C. R. Michel, K. S. Beard, A. Hohenstein, B. Bradley, N. Reisdorff, K. Haskins, T. DeLong, Identification of hybrid insulin peptides (HIPs) in mouse and human islets by mass spectrometry. *J. Proteome Res.* **18**, 814–825 (2018).
8. R. Mylonas, I. Beer, C. Iseli, C. Chong, H.-S. Pak, D. Gfeller, G. Coukos, I. Xenarios, M. Müller, M. Bassani-Sternberg, Estimating the contribution of proteasomal spliced peptides to the HLA-I ligandome. *Mol. Cell. Proteomics* **17**, 2347–2357 (2018).
9. Z. Rolfs, S. K. Solntsev, M. R. Shortreed, B. L. Frey, L. M. Smith, Global identification of post-translationally spliced peptides with neo-fusion. *J. Proteome Res.* **18**, 349–358 (2018).
10. K. Woods, P. Faridi, S. Ostrouska, C. Deceneux, S. Q. Wong, W. Chen, R. Aranha, N. P. Croft, D. Duscharla, C. Li, R. Ayala, J. Cebon, A. W. Purcell, R. B. Schittenhelm, A. Behren, The diversity of the immunogenic components of the melanoma immunopeptidome. *bioRxiv* 623223 [Preprint]. 2 May 2019. <https://doi.org/10.1101/623223>.

Submitted 30 January 2019

Accepted 15 July 2019

Published 16 August 2019

10.1126/sciimmunol.aaw8457

Citation: P. Faridi, C. Li, S. H. Ramarathinam, P. T. Illing, N. A. Mifsud, R. Ayala, J. Song, L. J. Gearing, N. P. Croft, A. W. Purcell, Response to Comment on “A subset of HLA-I peptides are not genomically templated: Evidence for cis- and trans-spliced peptide ligands.” *Sci. Immunol.* **4**, eaaw8457 (2019).

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

Sci. Immunol. 4, eaaw8457.
DOI: 10.1126/sciimmunol.aaw8457

ARTICLE TOOLS

<http://immunology.sciencemag.org/content/4/38/eaaw8457>

REFERENCES

This article cites 9 articles, 3 of which you can access for free
<http://immunology.sciencemag.org/content/4/38/eaaw8457#BIBL>

PERMISSIONS

<http://www.sciencemag.org/help/reprints-and-permissions>

Use of this article is subject to the [Terms of Service](#)

Science Immunology (ISSN 2470-9468) is published by the American Association for the Advancement of Science, 1200 New York Avenue NW, Washington, DC 20005. 2017 © The Authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original U.S. Government Works. The title *Science Immunology* is a registered trademark of AAAS.