Most viral peptides displayed by class I MHC on infected cells are immunogenic

Nathan P. Croft[a,b,1], Stewart A. Smith[c], Jana Pickering[c], John Sidney[d], Bjoern Peters[d,e], Pouya Faridi[a,b], Matthew J. Witney[c], Prince Sebastian[c], Inge E. A. Flesch[c], Sally L. Heading[c], Alessandro Sette[d,e], Nicole L. La Gruta[a,b,f], Anthony W. Purcell[a,b,1,2], and David C. Tscharke[c,1,2]

[a]Infection and Immunity Program, Biomedicine Discovery Institute, Monash University, Clayton, VIC 3800, Australia; [b]Department of Biochemistry and Molecular Biology, Monash University, Clayton, VIC 3800, Australia; [c]John Curtin School of Medical Research, The Australian National University, Canberra, ACT 2601, Australia; [d]Division of Vaccine Discovery, La Jolla Institute for Allergy and Immunology, La Jolla, CA 92037; [e]Department of Medicine, University of California, San Diego, La Jolla, CA 92093; and [f]Department of Microbiology and Immunology, University of Melbourne, The Peter Doherty Institute for Infection and Immunity, Melbourne, VIC 3000, Australia

Edited by Emil R. Unanue, Washington University, St. Louis, MO, and approved January 4, 2019 (received for review September 4, 2018)

CD8$^+$ T cells are essential effectors in antiviral immunity, recognizing short virus-derived peptides presented by MHC class I (pMHCI) on the surface of infected cells. However, the fraction of viral pMHCI on infected cells that are immunogenic has not been shown for any virus. To approach this fundamental question, we used peptide sequencing by high-resolution mass spectrometry to identify more than 170 vaccinia virus pMHCI presented on infected mouse cells. Next, we screened each peptide for immunogenicity in multiple virus-infected mice, revealing a wide range of immunogenicities. A surprisingly high fraction (>80%) of pMHCI were immunogenic in at least one infected mouse, and nearly 40% were immunogenic across more than half of the mice screened. The high number of peptides found to be immunogenic and the distribution of responses across mice give us insight into the specificity of antiviral CD8$^+$ T cell responses.

antigen presentation | virus | CD8$^+$ T cells | MHC class I | vaccinia virus

The recognition of peptide-MHC class I (pMHCI) complexes by CD8$^+$ T cells plays an important role in mediating antiviral immunity. A comprehensive view of the specificity of these antiviral responses requires an understanding of both the pMHCI presented and the corresponding CD8$^+$ T cell responses. This requires systems where pMHCI can be identified and any CD8$^+$ T cells elicited detected and quantified. The main approach in mouse models, where quantification of responses is not confounded by the numerous sources of diversity in human populations, has focused around identification of immunogenic pMHCI (referred to here as epitopes). For small viruses it is possible to generate overlapping libraries of synthetic peptides that can be screened to find many epitopes (1, 2). This approach is not feasible for larger viruses, so epitope discovery for these has been done largely by antigen expression screening and/or screening of synthetic peptides predicted to bind MHCI (3–9). For vaccinia virus (VACV), despite intensive efforts, the specificity of only around half of the total antiviral CD8$^+$ T cell response is known (10). Further, without knowledge of the pMHCI presented during infection it is unclear how these T cell specificities relate to the targets available on infected cells. Indeed more generally, while estimates have been made, it remains unknown for any virus the fraction of viral pMHC specificities presented on infected cells that are immunogenic (11).

Here we answer this question by bringing together high-resolution mass spectrometry (MS) to profile a viral immunopeptidome with a mouse model where rigorous assessment of immunogenicity is possible.

Results and Discussion

Comprehensive Identification of VACV-Derived pMHC. To generate a comprehensive list of the VACV pMHCIs that might be presented during infection of C57BL/6 mice, we used DC2.4 cells as the substrate because they have high expression levels of H-2^b-haplotype MHCI and high infection efficiency with VACV (12). The cells were infected with VACV for 6 h because we have demonstrated previously that presentation of epitopes from all kinetic classes can be detected at this time (12, 13). H-2K^b and D^b molecules were individually isolated by immunoaffinity purification, and acid-eluted peptides were then analyzed by high-resolution liquid chromatography–tandem mass spectrometry (LC-MS/MS). Spectra were searched against the VACV proteome. The distribution of the candidate VACV peptides according to their statistical confidence shows a clustering of sequences at the top, with around half having a confidence of >95% (Fig. 1*A*). We applied an inclusive threshold of 50% based on the appearance of two previously identified epitopes with confidence levels as low as 70%. Subsequent screening of the same data against combined mouse and VACV proteomes flagged several ambiguous spectra that could be assigned as either a VACV- or mouse-derived sequence, resulting in a list of 191 peptides (Dataset S1). Next, we validated this list in two ways: First, the integrity of the MS data were tested by either assessing the degree of similarity between spectra from peptides eluted from infected cells and that of a corresponding synthetic peptide (139 peptides) or through the sensitive detection method of multiple reaction monitoring (MRM) of mock versus infected

Significance

CD8$^+$ T cells are key to the defense of animals against virus infection. These immune cells recognize peptides derived from viral proteins that are displayed on the surface of infected cells in a complex with host proteins known as MHC I. Many viral peptides are displayed by MHC I on infected cells, but it has never been shown what fraction of these can induce an immune response. We answered this long-standing question, finding that more than 80% of vaccinia virus peptides presented by MHC I on infected mouse cells were immunogenic across a population of mice.

Author contributions: N.P.C., N.L.L.G., A.W.P., and D.C.T. designed research; N.P.C., S.A.S., J.P., J.S., M.J.W., P.S., I.E.A.F., and S.L.H. performed research; N.P.C., J.S., B.P., P.F., and D.C.T. analyzed data; and N.P.C., J.S., B.P., P.F., A.S., N.L.L.G., A.W.P., and D.C.T. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Data deposition: LC-MS/MS data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD010811.

[1]To whom correspondence may be addressed. Email: nathan.croft@monash.edu, anthony.purcell@monash.edu, or david.tscharke@anu.edu.au.

[2]A.W.P. and D.C.T. contributed equally to this work.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1815239116/-/DCSupplemental.

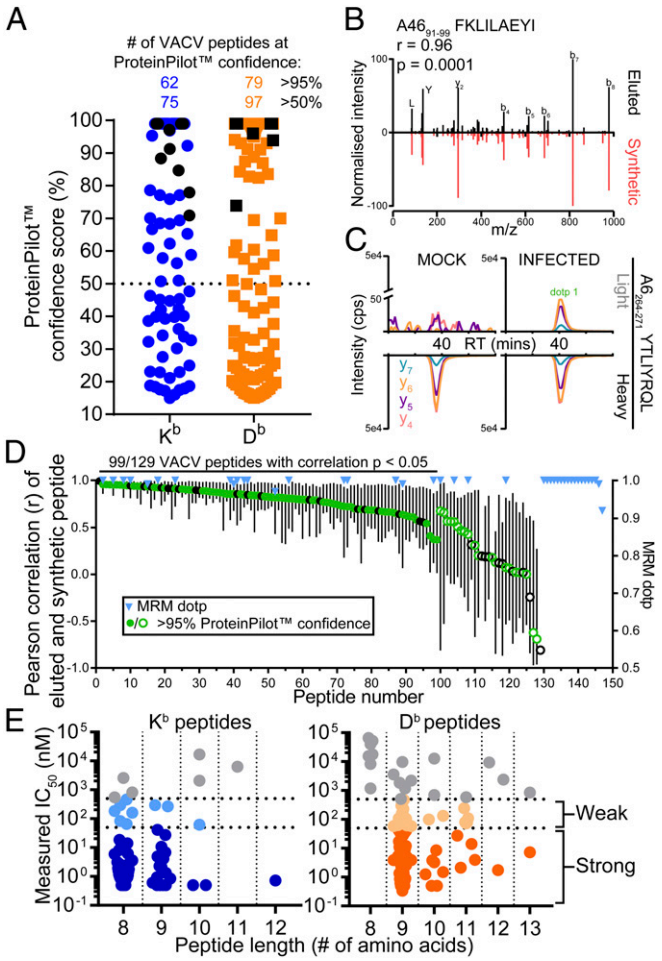

Fig. 1. Identification of a viral immunopeptidome. Cells were infected with VACV for 6 h, and MHCI-bound peptides were identified by LC-MS/MS. (*A*) ProteinPilot confidence score for VACV peptide sequences. Known immunogenic peptides are in black. Number of peptides above 50 and 95% confidence levels are shown at the top. (*B*) Example of a comparison between synthetic (red, *Lower*) and eluted (black, *Upper*) peptide spectra. Dominant y and b ions are indicated along with Pearson *r* correlation between spectra. (*C*) Example of MRM detection of a peptide, showing the native (*Upper*) sequence only in infected cells and spiked heavy peptide (*Lower*) in mock and infected samples. Dot-product (dotp) of eluted detection is indicated. (*D*) Pearson *r* correlation values (left axis) between synthetic and eluted spectra across 129 peptides. Filled circles are peptides with a correlation $P < 0.05$; circles in green are eluted peptides with ProteinPilot confidence >95%. The right axis shows MRM dotp values for each peptide validated by MRM. (*E*) Distribution of binding affinities of VACV peptides for MHCI.

cells in combination with isotopically labeled heavy peptides (45 peptides) (14, 15). These paired spectra and MRMs were analyzed both by eye (e.g., Fig. 1 *B* and *C*) and more objectively by calculating Pearson correlations between the sets of log-intensities of the observed b and y ions (16) or computed dot-products (dotp) for MRM (*SI Appendix*, Fig. S1*A*). Nearly 80% of tested peptides had a statistically significant ($P < 0.05$) similarity to the eluted peptide (Fig. 1*D*); of those that failed, some were due to low signal intensity of the experimentally acquired spectra. Indeed, four peptides with failed Pearson correlations were detected with a rigorous dotp of 1 by the more sensitive MRM assay. This process eliminated 19 potential false-positive peptides, most likely of mouse origin, leaving 172 LC-MS/MS–identified VACV-derived pMHCI. Second, we checked the distribution of lengths and the amino acids at predicted MHCI anchor positions, finding the sets of peptides typical of ligands for H-2K^b and D^b (*SI Appendix*, Fig. S1 *B* and *C*). Taking this a step further we measured the binding of these peptides to MHCI, which we note correlated very well with predictions made with NetMHC 4 (*SI Appendix*, Fig. S1*D*), and demonstrated that 84% of the VACV-derived peptides bound MHCI with <500 nM affinity (Fig. 1*E*). Finally, we note that our rate of discovery (73 and 97 for H-2K^b and D^b, respectively) was similar to that published for VACV in the context of two human MHC allomorphs (110 and 64 peptides for HLA-A2 and B7, respectively) (17).

Characteristics of VACV-Derived, MHCI-Presented Peptides. The 172 peptides identified by LC-MS/MS were spread across the VACV proteome, being encoded by the four kinetic classes of viral genes (18–20). Similar to other recent studies (4, 6, 17, 21), we found at least one epitope in 40% of all VACV proteins (92 of 230 proteins), but some of these proteins were richer sources of presented peptides (*SI Appendix*, Fig. S2*A*). In some proteins, the multiple pMHC detected comprised sets of peptides that were largely overlapping but with extensions at either the amino- or carboxy-terminus (for example, proteins J4 and F5, respectively; Dataset S1). In other cases, there were multiple non-overlapping peptides from the same protein. Most of the proteins here that are sources of multiple, nonoverlapping pMHC are immunoprevalent, having multiple epitopes presented in the context of several MHC allomorphs, e.g., A10, A3, A47, B8, and J6 as defined by others (6, 17). Others are unique to this study, e.g., A18, with five peptides and A23, A24, and A8, with three peptides each. There is no obvious characteristic that links these proteins, but understanding why they are so frequently presented should help open one of the remaining black boxes in antigen presentation: The selection of antigens and peptide precursors that occurs before loading onto MHCI. To understand the frequency of presentation across the proteome in a way that takes into account the different sizes of viral proteins, we calculated the number of presented peptides per 100 aa for each protein. These proteins were then grouped according to kinetic class, allowing an analysis of the role of expression time on pMHCI display that was not confounded by the differing numbers and sizes of genes in each class (Fig. 2*B*). This showed that there was no significant difference in the frequency of epitope occurrence by any kinetic class, suggesting that all are equally available to the antigen presentation pathway (Fig. 2*B*). To extend this analysis, we analyzed the abundance of pMHCI as inferred by the precursor ion intensity and the number of times a given peptide was identified by the mass spectrometer (i.e., spectral counts). These methods are approximations, but we note that they broadly correlate with each other across our data (*SI Appendix*, Fig. S2*B*). Intriguingly, the earliest class of genes was found to be the source of significantly more pMHCI than the later classes when normalized to their coding capacity (Fig. 2 *C* and *D*), but this was not related to the affinity of these peptides for MHC (*SI Appendix*, Fig. S2*C*). Looking at the individual protein level, the distribution of presented peptides within a viral protein shows that MHCI ligands are equally likely to be processed from any part of an antigen (*SI Appendix*, Fig. S2*D*). The start positions of the peptides were distributed along the entire length of the protein, with the average sitting at almost the exact midpoint of the protein, both for H-2K^b- and for D^b-binding peptides. Finally, VACV has been the subject of several studies to identify MHCI-restricted epitopes and is already the best characterized large virus in this regard, with >70 VACV-derived, H-2^b–restricted peptides or epitopes identified (Immune Epitope Database, IEDB; ref. 22). Most of these peptides were found first or confirmed in a single, landmark, proteome-wide study that identified 49 epitopes from a library of ~2,500 predicted pMHCI (4). This study set the benchmark for identification of CD8$^+$ T cell epitopes by prediction and screening and provides a good dataset for comparison with our LC-MS/MS identified peptides. The previously identified epitopes only partially overlapped with the MHCI ligands identified here (Fig. 2*E*). This suggests limitations in all approaches. For our LC-MS/MS study,

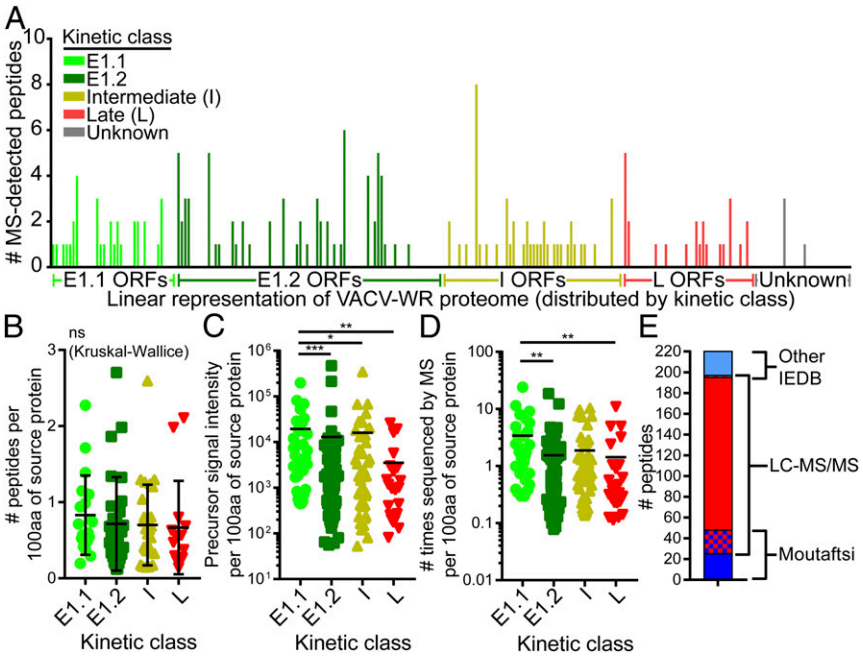

Fig. 2. The protein sources of a viral immunopeptidome. (*A*) The number of H-2^b–presented peptides from each VACV protein. The viral proteome is ordered by kinetic class and then position in genome, from left to right. Gaps are proteins from which no peptide was detected. (*B*) Peptide distribution according to the kinetic class of source proteins, normalized to protein length. (*C* and *D*) Abundance of pMHC from MS precursor signal intensity (*C*) and number of times each peptide was sequenced by MS (*D*), according to the kinetic class of source proteins, normalized to protein length. (*E*) Comparison of number of peptides identified here versus those identified previously. Epitopes identified in Moutaftsi et al. (4) and all others from the IEDB are shown separately. Kruskal–Wallis multiple comparison tests were used to determine significance between all possible pairs in *B–D*; ns, not significant, $*P < 0.05$, $**P < 0.01$, $***P < 0.001$.

pMHCI found by prediction will be missed if they are not presented on DC2.4, which are only a model of the actual antigen-presenting cell in vivo, or if they are only produced by cross-presentation. Also there are limits to the sensitivity of LC-MS/MS that mean that we will have missed some low-abundance pMHCI. Overall, however, the striking result is that our LC-MS/MS approach has more than tripled the number of pMHCI now associated with VACV infection of C57BL/6 mice.

A High Fraction of VACV-Derived pMHC Are Immunogenic. Next we screened the 172 candidate epitopes identified by LC-MS/MS for immunogenicity. Synthetic peptides were used to stimulate splenocytes from VACV-infected mice, with a readout of intracellular IFN-γ production, which we have validated for quantification of anti-VACV CD8$^+$ T cells (23). Reactivity to each peptide was tested eight times, and they were marked as immunogenic in a test if the response exceeded the mean plus three SDs of a set of negative controls, which received no peptide (Fig. 3 *A* and *B*). The number of peptides that were found to be immunogenic at least one time out of the eight tests was surprisingly high (142/172; 83%). From this we conclude that the vast majority of all viral pMHC have the potential to be immunogenic across a population of infected hosts. Even if we were more conservative and required a peptide to be positive in at least two tests, the fraction of pMHC deemed immunogenic would be almost 70%, which is also higher than previous estimates (11). These data suggest that immunogenicity was not binary, with peptides being found to be immunogenic in every number of tests, from nil to eight. However, an examination of the distribution of responses across the eight tests suggests a more nuanced picture. In general, individual responses are distributed evenly around a mean that rises for peptides that are found to be immunogenic in a greater number of mice. This can be seen for selected examples (Fig. 3*A*) and also when the data are plotted in aggregate for all mice and peptides (*SI Appendix*, Fig. S3*A*). Indeed, the increase in population mean as pMHC are positive in more mice can still be seen when values above the threshold are removed (*SI Appendix*, Fig. S3*B*). This suggests that our threshold limits the number of mice positive for pMHC where responses are close to the limit of detection. It is possible that adding further mice will find more peptides to be positive, but we note that for the pMHC we have designated to be non-immunogenic, the mean of all responses overlaps zero at the 95% confidence level. Having noted this general pattern, there are examples of peptides (e.g., E10$_{15–23}$; Fig. 3*A*) where the responses appear to be bimodally distributed, with the peptide seemingly immunogenic in some mice, but not others. Perhaps this occurs because particular T cell receptors are not available in all mice. Further, we note all assays for immunogenicity are subject to a threshold of detection. Altogether, we conclude that from an experimental point of view, immunogenicity will typically be found to be broader across a population than within a single individual, but in most cases this will reflect limits of detection and not absolute immunogenicity. As a final note, we tested mice by i.p. infection here, but immunogenicity of VACV epitopes can be influenced by route of infection, dose, or virulence of the strain used (24, 25).

The use of thresholds in individual mice also enables pMHC to be divided into immunogenicity groups for the purpose of further analyses. We have divided the pMHCI into three groups: Nonimmunogenic, which were never positive; major epitopes, which were positive in more than half of the mice (five to eight of eight mice); and minor epitopes, which are the remaining pMHC that were less frequently positive (one to four of eight mice). Using this definition, we found 39 and 44% of pMHC detected on VACV-infected cells to be major and minor epitopes, respectively (Fig. 3*B*).

A Comprehensive Set of VACV CD8$^+$ T Cell Epitopes. The peptides detected by LC-MS/MS only partially overlapped the known H-2^b-restricted epitopes for VACV. So we extended the immunogenicity screen to the 46 VACV-derived H-2^b–restricted peptides/epitopes from the IEDB that were not found by LC-MS/MS, in addition to one entirely unpublished and one more finely mapped version of a published (26) epitope from other work in our laboratories. Combining these with the set found by LC-MS/MS gave 220 VACV peptides tested eight times, giving a number of times positive and the average size of response for each peptide. As expected from the above, the average size of the CD8$^+$ T cell response across this complete set of peptides correlated with the frequency with which a response was detected (*SI Appendix*, Fig. S4*A*). This correlation was significant ($r = 0.8219$, $P < 0.0001$) by a Spearman rank test. Across the 220 pMHCI in this extended set, 84 were defined as major epitopes, and most of these (67 peptides) were found by our LC-MS/MS experiment (*SI Appendix*, Fig. S4*B*). By contrast, the large

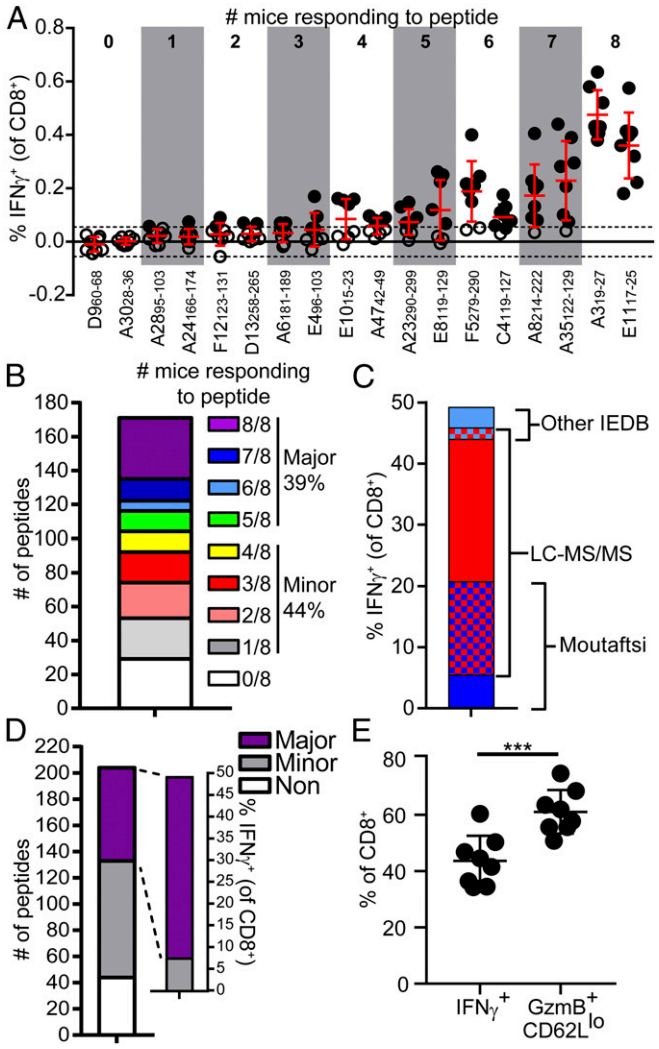

Fig. 3. The majority of members in a viral immunopeptidome are immunogenic. Peptides were tested for reactivity against $CD8^+$ T cells in spleens from mice infected with VACV. Each peptide was tested in eight mice. (*A*) The distribution of data from individual mice for a selection of 18 pMHC, two to represent each possible number of times positive (as noted on top). The dashed line represents the average of three times the SD of negative controls across the eight mice, and closed circles are values that were above the threshold for immunogenicity. (*B*) The distribution of pMHCI according to the number of times that each was immunogenic out of eight tests. (*C*) Comparison of the cumulative size of the response for VACV peptides identified by LC-MS/MS here versus those identified previously. (*D*) Summed size of the $CD8^+$ T cell response of the major and minor peptides. (*E*) Comparison of the cumulative response to the top 90 most immunogenic VACV epitopes against the total anti-VACV $CD8^+$ T cell response ($GzmB^+$, $CD62L^-$) 7 d after infection. ***$P = 0.0003$ by paired *t* test.

predictive study by Moutaftsi et al. (4) identified less than half of these major epitopes (35 peptides). The LC-MS/MS approach also found most of the minor epitopes. We note that we were not able to confirm the immunogenicity of all VACV peptides in the IEDB, so some revision of the status of these peptides is required (Dataset S1). Summed responses were used to show the fraction of the total response contributed by specificities found by LC-MS/MS compared with those already known (Fig. 3*C*). This included some sets of peptides with largely overlapping sequences, so to be conservative in our estimates of summed responses, we assumed that these pMHC sets elicit responses from largely overlapping sets of T cells. For this reason, only data from the peptide in each set that gave the largest response were included in summed responses. This revealed that our LC-MS/MS approach and the large predictive study of Moutaftsi et al. (4) identified the epitope targets for ∼90 and 50% of the total anti-VACV T cell specificities known thus far, respectively (Fig. 3*C*). The limitations of our LC-MS/MS approach are noted above, but we examined further why the predictive study failed to find so many of the pMHC presented here. Three reasons were identified: (*i*) Length: The original predictions only considered 8- and 9-mer and 9- and 10-mer peptides for $H\text{-}2K^b$ and D^b, respectively; (*ii*) Screening: Some peptides were predicted, but not found to be immunogenic, in the earlier study; (*iii*) Prediction: The original algorithms failed to rank some peptides highly enough to include in testing. Each of these reasons account for roughly a third of the peptides missed in the original study and found by LC-MS/MS here (*SI Appendix*, Fig. S4 *C* and *D*). We note that the original predictive study was done over a decade ago, and significant advances in MHC-binding predictions have occurred since then, which we expect will improve their accuracy. Our data will be useful to objectively evaluate the performance of newer algorithms and to identify weaknesses and refine MHCI-binding predictions (e.g., as in *SI Appendix*, Fig. S1*D*).

Summing the responses to all nonoverlapping major and minor epitopes shows that the vast majority of anti-VACV $CD8^+$ T cells are elicited by the major epitopes (Fig. 3*D*). Next, to estimate the fraction of the total response to VACV that is contributed by the newly expanded set of pMHCI, we compared the fraction of $CD8^+$ T cells from spleens of acutely infected mice that were $CD62L^-$ and Granzyme B^+ with the summed response to the 90 most immunogenic peptides (excluding overlapping peptides as noted above) in these same mice (Fig. 3*E*). It is not possible to test all of the peptides in a single mouse, but the 90 most immunogenic peptides account for 95% of the response found by all of the immunogenic peptides, so is a close estimate of the total now mapped. Using CD62L and Granzyme B is a more accurate way to determine the total anti-VACV $CD8^+$ response than using virus-infected cells as antigen-presenting cells and intracellular IFNγ as a readout, this latter method underestimating the response by around 50% (10). The comparison of these measures finds that these 90 pMHCI account for around 70% of the total anti-VACV $CD8^+$ T cell response (Fig. 3*E*). This suggests that despite the depth of all of the studies to date, further specificities remain to be defined. In addition to the limitations of using DC2.4 as the substrate for LC-MS/MS, there may also be immunogenic peptides that have post-translational modifications that we did not test, or come from outside conventional ORFs, or that are spliced (27–31). Finally, the 25 most immunogenic peptides are shown in a hierarchy (*SI Appendix*, Fig. S4*E*), excluding any less immunogenic variants from overlapping sets. This hierarchy includes four epitopes that we have refined by identifying more immunogenic variants here (e.g., $A8_{189–198}$ ITYRFYLINL, which is almost twice as immunogenic as the previously defined $A8_{189–196}$ ITYRFYLI), and eight peptides for which no variants exist in the IEDB.

Factors Contributing to Immunogenicity. Noting the importance of the major $CD8^+$ T cell epitopes in the total response, we then asked what factors were associated with these larger and more prevalent responses, examining time of expression, amount presented, and affinity of peptides for MHC.

First, the distribution of anti-VACV responses according to antigen and arranged by kinetic class visually suggests that late viral proteins were recognized by fewer $CD8^+$ T cells (Fig. 4*A*) as previously suggested (32). We examined this more quantitatively in the same way that we analyzed presentation on MHCI, by calculating the occurrence of major epitopes (Fig. 4*B*) and the total size of the response (Fig. 4*C*) per 100 aa for each viral protein. This allowed an unbiased comparison of immunogenicity across the kinetic classes. Immunogenicity was not significantly different across the kinetic classes of genes as a whole, and the distribution of major epitopes was not significantly different across the kinetic classes ($P = 0.2353$, Fisher's exact test). However, the distribution of highly immunogenic proteins, defined as the top 10% by total size of

response, was significantly different ($P = 0.04033$, Fisher's exact test), with no late proteins in this category. Overall then, our data support the notion that while late genes are equally likely to be a source of pMHC on infected cells (Fig. 2), their epitopes are less likely to be highly immunogenic.

Second, we investigated epitope abundance as inferred from precursor ion intensities (Fig. 4*D*) and spectral counts (Fig. 4*E*). On average, major pMHCI were identified by the mass spectrometer significantly more times than minor, or nonimmunogenic peptides. A similar trend was seen with precursor ion abundance, but pair-wise comparisons between groups were not statistically significant. Together these analyses suggest that abundance plays some role in determining immunogenicity, but the evidence was not especially strong. This conclusion is supported by the finding that the E1.1 class of VACV genes was more abundantly presented (Fig. 2 *C* and *D*), but not more immunogenic (Fig. 4 *B* and *C*) than other classes. Further, we note that a previous and more rigorously quantitative study with a subset of these epitopes found no correlation between dominance and immunogenicity (12).

Finally, we compared the binding affinity for MHCI of the major and minor epitopes, focusing only on those peptides that we identified by LC-MS/MS to ensure all were naturally presented (Fig. 4*F*). This found a very clear preference for high-affinity MHCI binding among the major epitopes, with these peptides having a significantly lower IC_{50} than minor epitopes and nonimmunogenic pMHC ($P < 0.0001$ and $P = 0.0005$, respectively, Kruskal–Wallis). While at face value this seems to be an obvious finding, we emphasize that this analysis only included peptides that were eluted from MHCI, and so all are of sufficient affinity to be presented. Further, we ruled out a role for confidence in the assignment of peptide identity in the original LC-MS/MS data as a factor that may contribute to the observed correlates with immunogenicity (*SI Appendix*, Fig. S5 *A* and *B*).

Across these three factors, affinity for MHC was the most rigorously associated with dominance, even in the set of peptides we identified as being presented on infected cells. Therefore, we speculate that the affinity of a peptide for MHC may have a significance for $CD8^+$ T cell immunity beyond making a pMHC available at the cell surface.

Summary. To conclude, this study provides the largest resource of LC-MS/MS–verified MHCI ligands along with their immunogenicity for any viral infection model. Using this information it was found that >80% of viral pMHC can be immunogenic across a set of individuals, a surprisingly high fraction. Immunogenicity was found to be higher on a population basis than in individuals due to weak epitopes not stimulating responses above the limit of detection in all individuals. Finally, we found that while roles can be shown for abundance and time of gene expression, major $CD8^+$ T cell responses are most strongly associated with high affinity for MHCI, even among peptides that are known to be presented.

Materials and Methods

Detailed methods are provided in *SI Appendix*.

Viruses and Cell Lines. Sucrose cushion purified stocks VACV strain Western Reserve (ATCC #VR1354) and DC2.4 cells (33) were used for LC-MS/MS experiments.

Infection of Cells for LC-MS/MS. Many 1×10^8 DC2.4 cells (or 5×10^7 cells for MRM validations) were infected at 10 plaque-forming units (PFU) per cell for a total of 6 h at 37 °C, after which cells were centrifuged, media removed, and the pellet snap frozen.

Mice and Infections. Specific pathogen-free female C57BL/6 mice 8–16 wk of age were obtained from The Australian National University (ANU) Phenomics Facility (Canberra, Australia). Mice were housed, and experiments were done according to the relevant ethical requirements and under an approval from the ANU animal ethics and experimentation committee (approvals A2011.01, A2013.37, and A2016.45). Mice were infected by the i.p. route with 1×10^6 PFU of VACV in 200 μL of PBS. To test all peptides eight times, they were screened in two main batches, so 16 mice were required for the main screen. Additional mice were then used to fill data for some failed assays in the main screen.

Identification of pMHC by LC-MS/MS. VACV-infected cells were lysed, and pMHCI complexes were purified (12). Briefly, cell lysates were cleared by centrifugation, and pMHC immunoaffinity was purified on Sepharose beads with anti-K^b or D^b before elution with 10% acetic acid. The mixture of peptides and MHC protein chains was fractionated on a reversed-phase C_{18} HPLC column. Peptide-containing fractions were analyzed on a 5600+ TripleTOF (SCIEX) mass spectrometer, operating in information-dependent acquisition mode via an online Eksigent Tempo nanoLC autosampler (SCIEX) and Eksigent cHiPLC nanoflex (SCIEX) system. For peptide identification, data were searched by the Paragon algorithm of ProteinPilot (v4.5; SCIEX) against the VACV strain WR proteome (2012–01; Uniprot). We initially searched against the VACV proteome alone, making the following choices when analyzing the

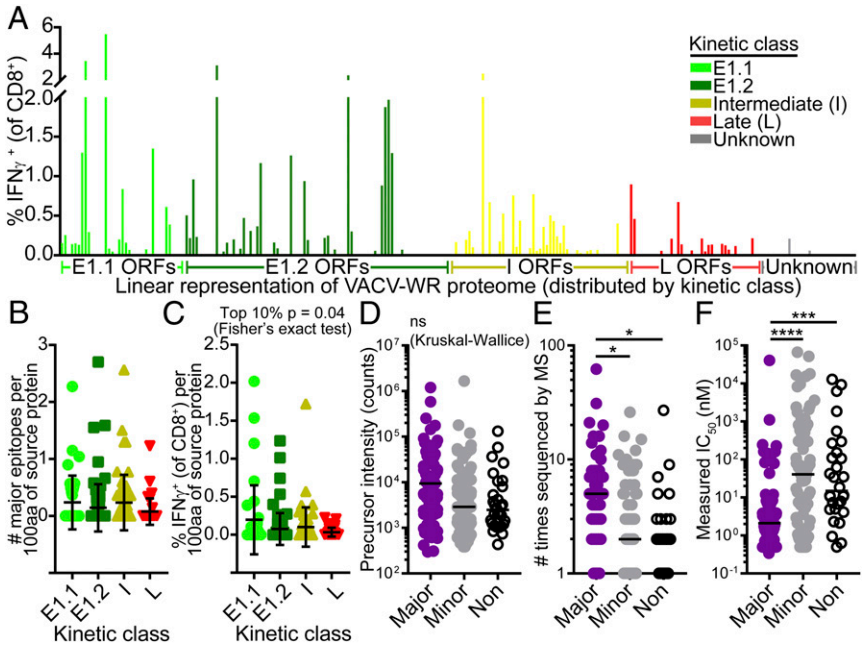

Fig. 4. Factors associated with immunogenicity for viral pMHC. (*A*) The cumulative size of CD8+ T cell responses to each viral protein when ordered as in Fig. 2*A*. Gaps are nonimmunogenic proteins. (*B*) Distribution of major epitopes and (*C*) immunogenicity of viral proteins across the kinetic classes, top 10% of immunogenic proteins are significantly reduced for late proteins (Fisher's exact test, $P = 0.04033$). (*D* and *E*) Abundance of pMHC estimated by precursor ion intensity (*D*) and number of times sequenced by MS (*E*) across major, minor, and non-immunogenic (Non) peptides. (*F*) Comparison of IC_{50} for major, minor, and nonimmunogenic peptides. Kruskal–Wallis multiple comparison tests were used to determine significance between all possible pairs in *D–F*; $*P < 0.05$, $***P < 0.001$, $****P < 0.0001$.

sequenced spectra. (*i*) Peptide FDR was ignored, instead allowing for a ProteinPilot confidence score of as low as 50% to be acceptable for subsequent testing (34); (*ii*) peptides that were detected as modified were included, but considered as unmodified. To account for potential false-positives deriving from mouse-derived pMHCI sequences, all spectra were sequenced against the *Mus musculus* proteome (2018–08; Uniprot) appended to the above VACV proteome. Spectra matching to mouse sequences in the latter search and VACV sequences in the former search were excluded, subject to further validation.

Correlation of Synthetic Spectra with Observed Eluted Spectra. Synthetic peptides were analyzed under LC-MS/MS conditions as above. MS2 spectra from each synthetic peptide were then compared with that of the original identified peptide from VACV-infected DC2.4 cells. Similarity between the spectra was assessed manually and by taking the $\log_{10}$-intensities of identified b and y ions and calculating their Pearson correlation coefficient and corresponding *P* value.

MRM. A mixture of each synthesized isotopic heavy peptide was spiked into the acid eluate of immunoprecipitated pMHCI complexes from VACV- or mock-infected cells before fractionation by RP-HPLC. Samples were assessed for the presence of peptide by MRM, acquired on a QTRAP 5500 (SCIEX) mass spectrometer equipped with an Eksigent Tempo nanoLC (SCIEX) autosampler and an Eksigent cHiPLC nanoflex (SCIEX) and utilizing Analyst 1.6 (SCIEX) software. MRM transitions are listed in Dataset S1. Data were analyzed in Skyline 64-bit [v4.1.0.18169; MacCoss Laboratory (35)], with dot-product values reported in Dataset S1.

Measurement of H-2^b–Binding Affinity. Measurements of peptide affinity for mouse class I H-2 K^b and D^b molecules were based on the inhibition of binding of high-affinity radiolabeled peptides to purified MHC molecules and performed as detailed elsewhere (36). Each competitor peptide was tested at six concentrations over a 10^5-fold range and in three or more independent experiments. Measured IC_{50} values using this method are reasonable approximations of true K_d (37, 38).

Quantification of Peptide-Specific CD8$^+$ T Cells. Splenocytes from mice infected with VACV 7 d previously were used to determine $CD8^+$ T cell responses to peptides (23–25). Briefly, 1×10^6 splenocytes were stimulated with peptides at 10^{-6} M for a total of 4 h (with 5 μg/mL Brefeldin A for the last 3 h), before being stained with monoclonal antibodies for surface CD8 and intracellular IFNγ. Negative controls had no peptide added. The average plus three SDs (of IFN-γ^+ as a percent of $CD8^+$) of negative controls was the threshold to determine whether a particular peptide was immunogenic in a test. To estimate the size of the response to each peptide, the average from the background wells was subtracted from values for each well. To determine the total size of anti-VACV $CD8^+$ T cell responses, splenocytes were stained directly ex vivo for surface CD8 and CD62L and intracellular granzyme B.

Data Analysis and Statistics. As described above, mass spectra were analyzed using ProteinPilot v4.5 (SCIEX). Raw mass spectra were analyzed using Peakview v2.2 (SCIEX). MHC-binding predictions (IC_{50}) were done using IEDB analysis resources NetMHCcons (ver. 1.1) and NetMHC4.0 (39). Statistical analyses of data were carried out in GraphPad Prism v7.01 or R studio using the tests noted in the text and/or figure legends. For statistical tests, $P < 0.05$ was considered to be significant; unless noted otherwise, error bars or graphs denote SD.

ACKNOWLEDGMENTS. We thank the staff of the ANU Australian Phenomics Facility for animal husbandry. This work was supported by a Project Grant from the National Health and Medical Research Council Australia (NHMRC) (APP1084283) (to D.C.T., A.W.P., and N.P.C.); an NHMRC Senior Research Fellowship (APP1104329) (to D.C.T.); an NHMRC Principal Research Fellowship (APP1137739) (to A.W.P.); and a Viertel Fellowship, ARC Future Fellowship, and NHMRC Program Grant (APP1071916) (to N.L.L.G.).

1. Precopio ML, et al. (2008) Optimizing peptide matrices for identifying T-cell antigens. *Cytometry A* 73:1071–1078.
2. Roederer M, Koup RA (2003) Optimized determination of T cell epitope responses. *J Immunol Methods* 274:221–228.
3. Tscharke DC, et al. (2005) Identification of poxvirus $CD8^+$ T cell determinants to enable rational design and characterization of smallpox vaccines. *J Exp Med* 201:95–104.
4. Moutaftsi M, et al. (2006) A consensus epitope prediction approach identifies the breadth of murine $T_{(CD8+)}$-cell responses to vaccinia virus. *Nat Biotechnol* 24:817–819.
5. Tscharke DC, et al. (2006) Poxvirus $CD8^+$ T-cell determinants and cross-reactivity in BALB/c mice. *J Virol* 80:6318–6323.
6. Oseroff C, et al. (2008) Dissociation between epitope hierarchy and immunoprevalence in CD8 responses to vaccinia virus western reserve. *J Immunol* 180:7193–7202.
7. St Leger AJ, Peters B, Sidney J, Sette A, Hendricks RL (2011) Defining the herpes simplex virus-specific CD8+ T cell repertoire in C57BL/6 mice. *J Immunol* 186:3927–3933.
8. Munks MW, et al. (2006) Genome-wide analysis reveals a highly diverse CD8 T cell response to murine cytomegalovirus. *J Immunol* 176:3760–3766.
9. Gredmark-Russ S, Cheung EJ, Isaacson MK, Ploegh HL, Grotenbreg GM (2008) The CD8 T-cell response against murine gammaherpesvirus 68 is directed toward a broad repertoire of epitopes from both early and late antigens. *J Virol* 82:12205–12212.
10. Yuen TJ, et al. (2010) Analysis of A47, an immunoprevalent protein of vaccinia virus, leads to a reevaluation of the total antiviral $CD8^+$ T cell response. *J Virol* 84: 10220–10229.
11. Assarsson E, et al. (2007) A quantitative analysis of the variables affecting the repertoire of T cell specificities recognized after vaccinia virus infection. *J Immunol* 178: 7890–7901.
12. Croft NP, et al. (2013) Kinetics of antigen expression and epitope presentation during virus infection. *PLoS Pathog* 9:e1003129.
13. Croft NP, et al. (2015) Simultaneous quantification of viral antigen expression kinetics using data-independent (DIA) mass spectrometry. *Mol Cell Proteomics* 14:1361–1372.
14. Croft NP, Purcell AW, Tscharke DC (2015) Quantifying epitope presentation using mass spectrometry. *Mol Immunol* 68:77–80.
15. Purcell AW, Croft NP, Tscharke DC (2016) Immunology by numbers: Quantitation of antigen presentation completes the quantitative milieu of systems immunology! *Curr Opin Immunol* 40:88–95.
16. Fälth M, et al. (2008) Validation of endogenous peptide identifications using a database of tandem mass spectra. *J Proteome Res* 7:3049–3053.
17. Gilchuk P, et al. (2013) Discovering naturally processed antigenic determinants that confer protective T cell immunity. *J Clin Invest* 123:1976–1987.
18. Assarsson E, et al. (2008) Kinetic analysis of a complete poxvirus transcriptome reveals an immediate-early class of genes. *Proc Natl Acad Sci USA* 105:2140–2145.
19. Yang Z, Bruno DP, Martens CA, Porcella SF, Moss B (2010) Simultaneous high-resolution analysis of vaccinia virus and host cell transcriptomes by deep RNA sequencing. *Proc Natl Acad Sci USA* 107:11513–11518.
20. Yang Z, et al. (2011) Expression profiling of the intermediate and late stages of poxvirus replication. *J Virol* 85:9899–9908.
21. Moutaftsi M, et al. (2010) Uncovering the interplay between CD8, CD4 and antibody responses to complex pathogens. *Future Microbiol* 5:221–239.
22. Vita R, et al. (2015) The Immune Epitope Database (IEDB) 3.0. *Nucleic Acids Res* 43: D405–D412.
23. Flesch IEA, Hollett NA, Wong YC, Tscharke DC (2012) Linear fidelity in quantification of anti-viral $CD8^+$ T cells. *PLoS One* 7:e39533.
24. Lin LCW, Flesch IEA, Tscharke DC (2013) Immunodomination during peripheral vaccinia virus infection. *PLoS Pathog* 9:e1003329.
25. Flesch IEA, et al. (2015) Extent of systemic spread determines $CD8^+$ T cell immunodominance for laboratory strains, smallpox vaccines, and zoonotic isolates of vaccinia virus. *J Immunol* 195:2263–2272.
26. Hersperger AR, Siciliano NA, Eisenlohr LC (2012) Comparable polyfunctionality of ectromelia virus- and vaccinia virus-specific murine T cells despite markedly different in vivo replication and pathogenicity. *J Virol* 86:7298–7309.
27. Hanada K, Yewdell JW, Yang JC (2004) Immune recognition of a human renal cancer antigen through post-translational protein splicing. *Nature* 427:252–256.
28. Warren EH, et al. (2006) An antigen produced by splicing of noncontiguous peptides in the reverse order. *Science* 313:1444–1447.
29. Ebstein F, et al. (2016) Proteasomes generate spliced epitopes by two different mechanisms and as efficiently as non-spliced epitopes. *Sci Rep* 6:24032.
30. Liepe J, et al. (2016) A large fraction of HLA class I ligands are proteasome-generated spliced peptides. *Science* 354:354–358.
31. Faridi P, et al. (2018) A subset of HLA-I peptides are not genomically templated: Evidence for cis- and trans-spliced peptide ligands. *Sci Immunol* 3:eaar3947.
32. Sette A, et al. (2009) Definition of epitopes and antigens recognized by vaccinia specific immune responses: Their conservation in variola virus sequences, and use as a model system to study complex pathogens. *Vaccine* 27:G21–G26.
33. Shen Z, Reznikoff G, Dranoff G, Rock KL (1997) Cloned dendritic cells can present exogenous antigens on both MHC class I and class II molecules. *J Immunol* 158: 2723–2730.
34. Faridi P, Purcell AW, Croft NP (2018) In immunopeptidomics we need a sniper instead of a shotgun. *Proteomics* 18:e1700464.
35. MacLean B, et al. (2010) Skyline: An open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics* 26:966–968.
36. Sidney J, et al. (1998) Measurement of MHC/peptide interactions by gel filtration. *Current Protocols in Immunology* (John Wiley & Sons, Inc., Hoboken, NJ), pp 18.13.11–18.13.19.
37. Cheng Y, Prusoff WH (1973) Relationship between the inhibition constant (K1) and the concentration of inhibitor which causes 50 per cent inhibition (I50) of an enzymatic reaction. *Biochem Pharmacol* 22:3099–3108.
38. Gulukota K, Sidney J, Sette A, DeLisi C (1997) Two complementary methods for predicting peptides binding major histocompatibility complex molecules. *J Mol Biol* 267: 1258–1267.
39. Karosiene E, Lundegaard C, Lund O, Nielsen M (2012) NetMHCcons: A consensus method for the major histocompatibility complex class I predictions. *Immunogenetics* 64:177–186.