Beyond model antigens: high-dimensional methods for the analysis of antigen-specific T cells

Evan W Newell1 & Mark M Davis2–4

Adaptive immune responses often begin with the formation of a molecular complex between a T-cell receptor (TCR) and a peptide antigen bound to a major histocompatibility complex (MHC) molecule. These complexes are highly variable, however, due to the polymorphism of MHC genes, the random, inexact recombination of TCR gene segments, and the vast array of possible self and pathogen peptide antigens. As a result, it has been very difficult to comprehensively study the TCR repertoire or identify and track more than a few antigen-specific T cells in mice or humans. For mouse studies, this had led to a reliance on model antigens and TCR transgenes. The study of limited human clinical samples, in contrast, requires techniques that can simultaneously survey TCR phenotype and function, and TCR reactivity to many T-cell epitopes. Thanks to recent advances in single-cell and cytometry methodologies, as well as high-throughput sequencing of the TCR repertoire, we now have or will soon have the tools needed to comprehensively analyze T-cell responses in health and disease.

From the advent of clonal selection theory1,2 to the present day, it has become increasingly clear that the adaptive immune response has, as its central unit, the expression of a single rearranged immunoglobulin or TCR on each B or T cell. And that, in general, single cells are the operational units or ‘quanta’ of immunity. With respect to T lymphocytes, this means that understanding their role in immune responses requires comprehensive methods of interrogating the phenotypic and functional characteristics of individual T cells. In this regard, the use of flow cytometry for high-throughput analysis of individual T cells has been the gold standard for many years3. Gradual improvements in flow cytometry, which allow simultaneous assessment of the expression of surface and intracellular markers4 and the precise temporal patterns of cytokine expression by T cells5–7, have enabled studies on the relationships between T-cell phenotype and/or function and clinical status in a range of diseases8–14. The study of antigen specificity, however, is complicated by enormous variability and unpredictability in terms of the epitopes targeted by T cells in any given T-cell response, especially considering the highly polymorphic nature of the MHC, and the fact that intact pathogens typically encode a wide variety of potential T-cell epitopes15. Furthermore, as the breadth or number of epitopes targeted by the T-cell response can be important, especially in rapidly evolving viral infections16–18, and the phenotypes of T cells targeting different epitopes from the same pathogen can vary widely19,20, it is important to be able to monitor recognition of numerous epitopes in the response to each pathogen. As a result, the number of parameters analyzed in any given experiment has grown beyond the number of colors (12–15) available for fluorescence-based flow cytometry, making the latter type of analysis more arduous and at times impossible. Recent developments in methods for analyzing antigen-specific T cells that extend these limits exploit multiplexing and single-cell mass spectrometry–based mass cytometry20–24. Other emerging technologies that promise to dramatically increase both the speed of obtaining information about T-cell responses and its depth include techniques enabling the analysis of single-cell mRNA transcripts25,26.

In addition, unlike most mouse models of immunological diseases, wherein the identity of the antigenic epitopes that drive disease initiation and/or progression are known, the instances of human immunological diseases wherein the precise specificities of T cells involved are known remain relatively rare. Therefore, until precise antigenic epitope specificities can be determined, study of these human T-cell responses requires alternative approaches; none appear to be more powerful than high-throughput sequencing of TCR repertoires. Data generated by this approach are providing insights into T-cell selection and the nature of repertoire diversity in various T-cell subsets in normal and pathological circumstances27,28. TCR sequencing approaches also allow the identification and tracking of TCR clonotypes or motifs involved in immune responses and various pathologies29–31. Moreover, high-throughput yeast-display approaches represent a way to identify peptide–MHC ligands that bind to these TCR clonotypes or motifs32,33. These approaches hold promise for identifying relevant antigens for immune responses for which the antigens are currently completely unknown. For instance, identification of antigens targeted by T cells in patients with auto-inflammatory diseases could facilitate the development of novel treatment options.

1Agency for Science, Technology and Research (A*STAR), Singapore Immunology Network (SIgN), Singapore. 2Department of Microbiology and Immunology, The Howard Hughes Medical Institute, Stanford, California, USA. 3Institute for Immunity, Transplantation and Infection, The Howard Hughes Medical Institute, Stanford, California, USA. 4The Howard Hughes Medical Institute, Stanford, California, USA. Correspondence should be addressed to E.W.N. (evan_newell@immunol.a-star.edu.sg) or M.M.D. (mmdavis@stanford.edu).

Received 2 October 2013; accepted 4 December 2013; published online 19 January 2014; doi:10.1038/nbt.2783
In this Review we discuss the advantages, disadvantages and complementarity of these high-dimensional approaches for the study of antigen-specific T cells. Common to each approach is the goal of understanding and/or exploiting the specificity of the T-cell–mediated immune response to manipulate or predict outcomes of immunological diseases or vaccine responses. These recent technological advances seem poised to finally make possible the comprehensive analyses of T-cell responses.

Analyzing T-cell phenotypic and functional diversity

Each individual αβ T cell expresses one of >1014 different TCR heterodimers34 and each of these TCRs is specific for a very small fraction of possible self or foreign antigens presented in the context of an individual’s MHC molecules (Fig. 1a). Thus, in terms of diversity of antigen specificity alone, T cells are one of the most diverse cell subsets in the body. Several approaches for analyzing this diversity exist, and each has advantages and disadvantages (Table 1).

Flow cytometry. Flow cytometry has been the leading method for measuring this diversity over the past three decades. The latest instruments are remarkably powerful, capable of analyzing and/or sorting cells based on up to ~18 cellular parameters at >10,000 cells per second4. Such high-dimensional and high-throughput analysis has revealed relationships between T-cell phenotype and/or function and the clinical status for a range of diseases. For instance, in several infectious disease settings, a correlation between pathogen control and the polyfunctionality (that is, the capacity to perform more than one effector function, such as producing more than one cytokine) of the corresponding antigen-specific T cells has been identified8–11. As less simplistic relationships between disease status and antigen-specific T-cell phenotypic and/or functional status must also exist, a great deal of effort is being dedicated to developing computer algorithms that can analyze and extract useful information from this otherwise impossible-to-understand, high-dimensional flow cytometry data15–38.

Human T cells are often segregated into four major categories based on surface markers that indicate their proliferative potential, cytotoxicity capacity and their ability to produce cytokines. They are the following: naive (CCR7+CD45RA+), central memory (CCR7+CD45RA−), effector memory (CCR7−CD45RA−) and short-lived effector cells (CCR7−KLRG1+Tbet+CD69+). Although simple, useful and well accepted, it is clear that much more heterogeneity exists than is captured by these subdivisions. For instance, expression of all possible combinations of CD27, CD28, CD62L and CCR7 are also observed, yet the significance of all of these distinctions is not clear40. Also, even cells fitting the most strict definition of naive (CD45RA+CD45RO−CD27−CCR7−CD62L−CD28−), yet expressing higher levels of CD95, interleukin 2 receptor subunit-β (IL-2Rβ), CXC-motif

Figure 1 Antigen recognition by the T-cell receptor and probing antigen specificity with peptide–MHC multimers. (a) Antigen-specific T-cell responses are initiated through the interaction of a TCR, expressed on T cells, and the corresponding peptide-MHC protein complex expressed by antigen-presenting cells. TCR engagement initiates a complex cell-signaling cascade that results in T-cell activation. (b) Binding affinity of TCR to its specific peptide–MHC ligand is very low (~1–100 μM) and has very fast dissociation kinetics (t1/2 usually much less than a minute). Thus, monomeric staining reagents are insufficiently stable for the detection of antigen-specific T cells. In contrast, by taking advantage of cooperative binding, multimeric complexes of peptide-MHC allow for remarkably sensitive and accurate detection of antigen-specific T cells37,58. (c) Mass cytometry and dimensionality-reduction methods allow integrated analysis of T-cell phenotype and function. In the example shown here, to visualize diversity of peripheral blood CD8+ T cells, 25 parameters were measured for each cell including 16 phenotypic markers and 9 functional markers. PCA was applied to generate three aggregate parameters describing ~60% of total variance. A representative donor’s 3D-PCA plot with cells annotated based on previously defined stringent criteria for naive, central-memory (Tcm), effector-memory (Tem) and short-lived effector cells (Tslec). Adapted with permission from ref. 22. (d) To illustrate the phenotypic and functional meaning of a non-naive cell progression along the PC2 axis of the PCA plot, average expression of phenotypic (left graph) and functional (right graph) parameters were normalized and plotted as a function of normalized PC2 values. This unsupervised analysis provides a hypothetical framework for graded T-cell differentiation involving progressive gains and/or losses of surface marker expression and functional capacities. Adapted with permission from ref. 22. (e) To illustrate the power of nonlinear dimensionality approaches, a linear PCA analysis of bone marrow–derived cells colored by a number of user-defined cell subsets is compared to two different nonlinear approaches, Isomap and visNE. Adapted with permission from ref. 38.
chemokine receptor 3 (CXCR3) and leukocyte function-associated antigen-1 (LFA-1), have been designated stem-cell–like memory cells with high proliferative capacity. T cells can also be classified by their functional capacities (e.g., to produce various cytokines and effector molecules), transcription factor expression profiles, as well as markers indicative of their tissue trafficking potential. Combined with antigen specificity, as discussed below, the number of parameters that can be clearly distinguished by even the latest fluorescent flow cytometers is insufficient for an integrated and comprehensive view of T-cell diversity.

**Mass cytometry.** The arrival of mass cytometry has allowed a quantum leap in phenotypic and functional characterization of single T cells. The mass spectrometry–based flow cytometry method (CyTOF) uses isotopically purified heavy metal atoms, instead of fluorophores, as tags. Whereas the number of parameters available to fluorescent flow cytometry is limited by broad spectral overlap leading to cross-talk between fluorescent channels, in CyTOF many more parameters (now ~40, and dozens more should be possible in the near future) can be distinguished. In addition, thanks to reduced cross-talk between channels due to accurate resolution of tags differing by as little as a single atomic mass unit, the analysis and interpretation of CyTOF data are greatly simplified. In terms of sample throughput, however, mass cytometry is substantially slower (capable of analyzing ~500 cells per second) than fluorescence flow cytometry (up to tens of thousands of cells per second). Although reduced sample throughput makes carrying out studies with large sample sizes or rare cells a challenge, using pre-enrichment strategies, such as T-cell purification methods or tetramer enrichment approaches as we have applied, makes it possible to analyze even extremely rare antigen-specific T cells. Mass-tag barcoding can also alleviate the throughput problem by allowing a large number (up to 96 so far) of uniquely tagged (barcoded) samples to be analyzed simultaneously. After data acquisition, simple and effective software is available for deconvolution of each barcoded sample.

Although mass cytometry is far from a comprehensive ‘proteomics’ method, extending cellular analysis into 40 dimensions means that each cell can be parsed into one of ~2^40 (~1 trillion) possible bins, allowing cells to be classified in unprecedented detail. It also means that a wide variety of T-cell markers can be assessed simultaneously providing a view of the overall diversity of a sample of cells. We have used mass cytometry to broadly probe the relationship between the CD8+ T-cell phenotype, function and antigen-specificity by simultaneously assessing several markers of each. For example, after using some of the parameters available at the time to isolate single CD8+ T cells, we probed six different antigen specificities using peptide–MHC tetramers (discussed below; [Fig. 1b]) and used the remaining 25 parameters to evaluate expression of 16 surface and 9 functional markers. In that study, we chose the parameters and the timing of assessment to maximize the amount of information obtained from each cell and to reveal variation. Nonetheless, a caveat of this approach is that only a snapshot of information from a single time-point taken from a single blood sample is available for each cell. Accurately tracking the kinetics of cellular responses requires an ability to continuously monitor single cells (as described below for the kinetics of cytokine secretion and in references 6-7).

As there is a great deal of interest in T-cell polyfunctionality as an index of the strength of an immune response, in the same study we took advantage of the large number of cellular parameters available to mass cytometry by assessing nine different functional capacities (eight cytokines and a marker of granular release, CD107), in populations of activated human CD8+ T cells. This gave us 512 (2^9) possible combinations, and we were able to observe at least 242 of these, indicating that a vast number of possible functional combinations can be expressed. Furthermore, by using MHC tetramers bearing influenza, cytomegalovirus (CMV) and Epstein-Barr virus (EBV) epitopes together with these nine functional markers, we showed that T cells specific for each of the different viruses expressed partially overlapping but distinct combinations of cytokines, ranging from 50 to 100 different combinations of cytokines observed. Although only a few antigen specificities have been analyzed in this manner so far, these results indicate that T-cell responses are far more complex than previously thought, at least for CD8+ T cells, and that they seem tailored to the particular pathogen.

To integrate both phenotypic and functional capacities of the cells, we also combined all the measured information about each cell and performed dimensionality reduction using principal component analysis (PCA). PCA uses weighted combinations of each of the measurements to create composite parameters that maximally represent the data with a minimal number of new parameters. This approach provides an unbiased composite representation of the cellular diversity that incorporates both phenotypic and functional information. It also provides insights about how these markers, and different T-cell subsets, are related to each other ([Fig. 1c,d])22. However, the represented pattern depends heavily on the cellular markers chosen for the analysis and on the composition of the cells being analyzed. We anticipate that this and other challenges of analyzing high-dimensional cellular data will improve as new computational methods are developed and their utilities compared for various applications. For example, more sophisticated analogs of PCA that combine the measured parameters in a nonlinear fashion are capable of explaining more variation in fewer dimensions and should enable better resolution of T-cell diversity ([Fig. 1e]).

The advantages of both fluorescence and mass cytometry include their ability to accurately and sensitively quantify proteins at the individual cell level in rapid succession (hundreds to tens of thousands per second). However, their dependence on the availability of reliable antibodies specific for each protein severely limits their utility for discovery of novel proteins of interest. Furthermore, though the number of channels continues to increase, even for mass cytometry the number of isotopes in the periodic table represents a hard cap on the number of probes possible by this approach.
Single T-cell phenotyping and temporal analysis using microwells. Another important new technique was developed by Love and colleagues, who spread T cells onto slides engineered to contain 100,000 microwells, such that many wells are occupied by only a single cell. These cells can then be stimulated and many potential cytokines analyzed using antibody capture. This system revealed something unique, which is that individual T cells typically make multiple cytokines, not necessarily at the same time but often sequentially over a span of several hours. This temporal complexity is not captured in other systems and takes advantage of the unique ability of this system to analyze the output of the same cells over time, and in a very high-throughput manner.

Single-cell transcription profiling. Many of the limitations of flow cytometry and mass cytometry will eventually be overcome through the use of single-cell gene transcription profiling. However, because copy numbers of mRNA can be as low as just a few copies per cell, single-cell gene expression analysis remains difficult and the data need to be interpreted carefully. Nonetheless, after careful validation, higher-throughput (up to 96 cells at a time), single-cell, quantitative RT-PCR systems are already allowing up to 96 mRNA transcripts to be analyzed in each cell in a quantitative fashion. (Fig. 2). Highlighting diversity in the abundance of cellular mRNA transcripts, this approach proved to be useful in distinguishing gene expression profiles of tetramer-stained and sorted antigen-specific T cells responding to the same antigen in the context of different gene-based vaccination. Though it is a dramatic improvement over bulk analysis without single-cell resolution (Fig. 2), the analysis of only 96 cells at a time severely limits the current applicability of this approach. We anticipate that implementation of barcoding-based technologies (as discussed below) will gradually increase throughput and broaden the applicability of single-cell transcriptional profiling.

Also exciting is the emergence of single-cell RNA sequencing methods that allow genome-wide mRNA transcription quantification. The use of barcoded primers used to multiplex the analysis of a large number of cells in a single DNA sequencing reaction will surely broaden the utility of the single-cell transcriptomics approach. Although it will always be important to evaluate cellular protein levels instead of relying solely on mRNA levels, accurate genome-wide transcriptomic information from single cells, if of sufficient throughput and sensitivity, will add insight to that gained using proteomic methods; transcriptomic information might also lead to markers that can then be analyzed by proteomic methods. That said, cytometry methods that directly evaluate T-cell antigen specificity by using peptide–MHC multimers will remain essential.

Characterizing T-cell antigen specificity. Studies of human immunology are limited by the availability of cellular material, which is usually restricted to small volumes of blood. Furthermore, to varying extents, depending on the immune response of interest, responsive T cells often represent a very small minority of total blood lymphocytes. Thus, to achieve sufficient signal above the inherent noise of complex blood samples, one must restrict analysis to the relevant cells—ideally through identification of T cells specific for the antigen of interest. However, the precise antigen specificities of T cells are known for only a handful of immune responses. Without knowledge of precise antigenic epitopes, only indirect methods that rely on a cellular response to stimulation with whole antigen (e.g., proliferation, cytokine production, surface marker expression changes) can be used. Furthermore, in these approaches, the analysis is ‘single-plex’ in that each sample of cells can be interrogated with only a single antigen (or single mixture of antigens). Thus, epitope mapping is either very low resolution or requires large volumes of sample. Detecting rare, antigen-specific T cells is also problematic especially if any other cells in the sample are activated and contribute to background signal. Lastly, depending on the method of detection (unless high-dimensional approaches are used), the phenotypic and/or functional profile of the antigen-specific T cells can be markedly limited in depth and influenced in unpredictable ways by the stimulation used to identify them.

Improvements to methods enabling direct detection of antigen-specific cells are overcoming each of these limitations. As a result of their cooperative binding and ability to specifically label antigen-specific T cells (Fig. 1), peptide–MHC tetramers (or multimers) are now widely used for directly identifying and characterizing antigen-specific T cells. Several recent technical advances have made this approach especially practical, including (i) higher-throughput production of hundreds or thousands of tetramers from one batch.
of prepared MHC class I protein–loaded ultraviolet (UV)–cleavable and exchangeable peptides\textsuperscript{59,60} or cleavable and exchangeable class II–associated li peptide (CLIP)-loaded MHC class II proteins\textsuperscript{61}; (ii) detection of rare antigen–specific T-cell populations using magnetic bead–based enrichment of antigen–specific cells labeled with the peptide–MHC tetramer, allowing the characterization of very rare T cells, even those in the naïve T-cell repertoire\textsuperscript{61,62}; (iii) probing of larger numbers of T-cell antigen specificities in a single sample using combinatorial staining approaches\textsuperscript{23,24,63}.

Most recently, we demonstrated an approach that incorporates each of these improvements and exploits mass cytometry\textsuperscript{20}. In this approach >100 antigen specificities and >20 phenotypic or functional markers can be probed in a single blood or tissue sample (Fig. 3a–c). We used this approach to simultaneously screen for, identify and phenotype T cells that target different epitopes derived from the same virus\textsuperscript{20} (Fig. 3d,e). Assuming that it will be possible to increase the number of T-cell specificities that can be probed simultaneously using this approach (reconstruction experiments show that even a thousand different specificities may be possible\textsuperscript{20} and to apply it to study CD4\textsuperscript{+} T cells and MHC class II–restricted antigens (H. McGuire & M.M.D., unpublished data), we anticipate that this method will enable the rapid identification of antigens related to a wide range of pathologies currently characterized by poorly defined T-cell responses.

It is currently impossible to predict a priori which of the multitude of possible T-cell epitopes will be targeted by a T cell–mediated immune response\textsuperscript{15,64}. An ability to probe >100 or possibly >1,000 candidate epitopes is helpful but alone will not solve this problem, especially for diseases where the range of possible antigens is too large. As we and others have demonstrated for rotavirus and several other viral pathogens (e.g., influenza virus, Dengue virus and HIV\textsuperscript{20,65–68} with relatively small genomes, epitope– or MHC binding–prediction algorithms\textsuperscript{69,70} provide a means of narrowing the possibilities and have been applied with some success. However, even for these small-genome pathogens, the number of candidate epitopes can be quite large when prediction stringencies are relaxed in an effort to avoid missing epitopes or when viruses with high mutation rates and significant epitope variation are considered. Furthermore, binding prediction algorithms perform especially poorly for MHC alleles for which there is much less peptide binding data with which to train the algorithms\textsuperscript{71}. Sometimes it can help to broadly map epitopes through cellular stimulation–based approaches or, in the case of MHC class II–restricted cells, tetramer-guided epitope mapping\textsuperscript{72} before narrowing it down to the precise epitopes using peptide–MHC multimers. Another solution is to use careful transcript sequence analysis, as has been done for the identification of tumor–specific mutations encoding candidate tumor antigens, based on whole-exome sequencing\textsuperscript{73}. For example, in one recent study, >400 candidate antigens were screened using combinatorial tetramer staining of tumor–infiltrating lymphocytes in a lymph-node metastases to identify a single dominant epitope from a melanoma patient\textsuperscript{74}. Lastly, direct identification of MHC-bound peptides through peptide elution\textsuperscript{75,76} and ever-improving mass spectrometry/proteomics-based peptide sequencing should provide more accurate means of identifying candidate epitopes for
Although aided by the development of more sensitive and sophisticated peptide-identifying mass spectrometers, this approach is still currently limited by the large number of antigen-presenting cells required and the associated bias against low-abundance and low-affinity peptide ligands. Thus, despite the ongoing improvements being made to T-cell-epitope discovery methods, identification of antigens for a number of diseases, such as autoimmune and auto-inflammatory diseases, remain elusive. In these cases, TCR sequencing-based approaches hold promise as a reliable means of tracking and understanding antigen-specific T-cell responses, even when the identities of the antigens are unknown.

**Sequencing the TCR repertoire**

The advent of inexpensive and massively parallel DNA sequencing has begun to affect T-cell analysis, starting with the pioneering work of Robins and colleagues to understand the scope of the human TCR repertoire. From these studies, it was found that natural biases in VDJ recombination lead to an overrepresentation of common TCR sequences from individual cells much more efficiently than current practice.

**Figure 4** Strategies for high-throughput single-cell analysis of TCR sequences and identification of TCR ligands. (a) As demonstrated for B-cell immunoglobulin genes, sequencing of V regions of endogenously paired TCRα and TCRβ genes can be performed in high-throughput by mRNA capture and emulsion linkage RT-PCR, or by direct cellular-emulsion-linkage PCR. Cell-specific barcoded tags can also be introduced at the single-cell stage as a general means of increasing the throughput of this type of approach. Adapted with permission from refs. 48. (b) Libraries of yeast clones displaying 10^5–10^6 unique peptides tethered to MHC molecules enable high-throughput screening for peptides capable of binding MHC (because these peptides result in proper folding and surface expression of peptide-MHC complexes) and for peptides capable of binding to a TCR of interest (here the TCR is used as a tetrameric staining reagent). After multiple rounds of selection, hits are sorted, cloned and sequenced; the peptide sequences can then be analyzed. Adapted with permission from ref. 33.

Extending this approach, several higher-throughput means of obtaining endogenous pairs of TCRα and TCRβ sequences from individual cells are emerging (Fig. 4a). These include single-cell mRNA capture and emulsion linkage RT-PCR, as recently described for simultaneous sequencing of heavy and light chains of immunoglobulin in single B cells; single-cell barcoding, as has been used for single-cell mRNA sequencing, and direct cellular-emulsion-linkage PCR. Each of these methods allows the sequencing of α and β or γ and δ TCR sequences from individual cells much more efficiently than current practice.

In terms of applications, an especially promising approach would be to use single-cell methodologies and barcoding to analyze antigen-specific T-cell responses in the many cases where the major epitopes are unknown. A glimpse of the potential insights that can be gained with this approach is found in our recently published study of T-cell responses in patients with celiac disease. In this disease, patients have a strong autoimmune response to gluten, which inflames the gut and causes serious discomfort and damage. In most cases, eating a gluten-free diet eliminates the symptoms. We monitored the

---

*© 2014 Nature America, Inc. All rights reserved.*
peripheral blood lymphocytes of celiac patients who had ingested gluten for any unusual responses. What was expected based on the work of Sollid and colleagues\(^{80}\) was that gliadin (a subset of gluten proteins)-specific CD4\(^+\) T cells would be activated and then enter the circulatory system approximately six days after the first gluten exposure. However, we found that populations of CD8\(^+\) T cells and γδ T cells appeared (and disappeared) with the same kinetics, all expressing gut-homing receptors. We then employed a single-cell TCR amplification method using a large panel of primers to show that both the CD8\(^+\) and γδ T-cell populations were highly enriched for particular sequences and that complementarity-determining region 3 (CDR3) encoded by these sequences was shared between many of the patients. This implies that all three of these different types of T cells are recognizing specific ligands in this disease and somehow coordinating their responses. It also shows that it is possible to analyze T-cell responses directly after particular immunological events (albeit with a time lag) from a blood sample and assess the clonality of the response. In this study and in an analogous study that identified oligoclonal, skin-resident, T-cell populations responding to herpes virus infection\(^{80}\), one chain from each T-cell population (TCRs, \(\alpha, \beta, \gamma\) or \(\delta\)) was analyzed. However, it is also possible to amplify both chains from single T cells as described above (which we have done; unpublished data).

Adapting this methodology to a high-throughput system of single-cell analysis, as has been done by Georgiou\(^{88}\) and W. Robinson (Stanford University, Department of Medicine; personal communication) for plasmablasts, one could envisage a very powerful way to analyze T-cell responses to disease or vaccination, which is entirely independent of other methods and requires no knowledge of the major antigens, but can lead to them.

**Obtaining antigen specificity from TCR sequences**

It would be ideal if high-throughput sequencing of enough TCR sequences could be used to computationally predict T-cell specificity from a given TCR heterodimer sequence. This would solve many of the problems and challenges discussed in this Review. However, unfortunately, we do not anticipate this to be possible anytime soon. There are several reasons for this but all stem from the highly variable nature of each of the components of the TCR-peptide-MHC complex. Unlike the binding of short, linear peptides to MHC, which is mediated by relatively simple motifs and mostly predictable anchor residues\(^{81}\), the energetics of TCR binding to peptide–MHC rely on numerous, highly unpredictable contacts between the TCR and both the MHC and the antigenic peptide\(^{82}\), although peptide residues generally contribute more than the MHC residues\(^{83}\). Solving such a problem depends on overcoming the difficulties in the field of predicting protein–protein interactions in general, a challenge even when there are high-resolution structures for each\(^{84,85}\). Furthermore, it is clear that such a problem has far from a unique solution for a given TCR. This is because TCRs generally have a great deal of flexibility in one or both of the CDR3 loops, which are the principal determinants of peptide specificity\(^{86}\); this flexibility allows them to recognize many different peptides that may not share any sequence homology. For example, Garcia and colleagues solved the structures of one TCR binding to four structurally distinct peptide–MHC complexes; in each case, the CDR3s of the TCR adopted a different conformation\(^{87,82}\).

Considering these observations together with the relatively ubiquitous cross-reactivity of TCRs to homologous peptides\(^{87,88}\), one can conclude that most TCRs can likely recognize many different peptides, and with very low affinities (~10,000 times lower than most affinity-matured antibodies, for example). This means the ‘fit’ between a TCR and a ligand is much less stringent and therefore substantially harder to model than the fit between an antibody and its cognate antigen. Thus, we are not optimistic about the possibility of being able to computationally predict TCR antigen specificity solely from sequence information, at least by some form of structural prediction. An alternative may be to identify specific sequence motifs that can be associated with particular antigen responses (as found for antibody sequences in Dengue fever virus infection\(^{89}\)). But for now we will have to rely on experimental approaches for identifying T-cell antigens based on cloned TCRs. Fortunately, some such approaches are already proving to be useful.

One general strategy for identifying antigens bound by a particular cloned TCR is to create libraries of cells (insect/baculovirus or yeast), each expressing a single peptide–MHC antigen that can be selected by staining with tetramerized soluble TCR proteins. For this approach, one needs to know the restricting element (e.g., the appropriate MHC allele) before addressing the even larger challenge of identifying the corresponding peptide epitope. The power of this approach can be illustrated by the identification of the peptide–MHC complex recognized by the TCR expressed on the murine diabetogenic T-cell clone, BDC-2.5. Insect cell display was used to identify several mimotopes (presumably irrelevant peptides also capable of stimulating these cells\(^{100,101}\) that eventually aided in the identification of the true endogenous antigen. This case was made especially difficult by the abnormal MHC binding register of the pathogenic epitope, which was not found in any of the insect cell libraries\(^{102}\).

A possibly faster approach is to generate yeast display libraries restricted to sequences derived from the pathogen of interest, as was demonstrated for influenza virus–specific human leukocyte antigen DR (HLA-DR)–restricted CD4\(^+\) T cells\(^{102}\). Extending this to large random libraries\(^{88,90}\), one can construct yeast display libraries in which a large number (10\(^5\)–10\(^7\)) of different peptides bound to a given MHC molecule can be expressed, and these libraries can then be screened with a particular TCR to identify ligands\(^{32,33}\) (Fig. 4b).

This system enabled identification of a number of interesting H-2L\(^d\), restricted ligands of well-described mouse T-cell receptor clones. These included an epitope with a sequence totally unrelated to the sequence of the original antigen; although it had a high affinity for the TCR, this epitope failed to trigger T-cell activation, apparently owing to an abnormal docking orientation\(^{32}\). Hits like this and other mimotopes may be difficult to distinguish from immunologically relevant epitopes by this approach. However, as demonstrated\(^{101}\), identification of a consensus sequence from the hits searched against sequence databases containing candidate antigens should help tremendously to narrow down the possibilities especially in cases where the source of the antigen is totally unknown. From this, shortlisted candidates can be used as input for a tetramer panel or stimulation-based studies of polyclonal T-cell populations from which the TCRs of interest have been identified. Thus, as large-scale, yeast-display screening approaches continue to improve in throughput and depth, we are optimistic that many more elusive, disease-associated, T-cell antigens will be identified, possibly leading to novel therapeutics and/or diagnostics.

**Conclusions**

The complexity of T-cell responses to pathogens in outbred populations has, in most cases, severely limited their assessment. This has forced a reliance on highly constrained animal models, which cannot give us a complete picture of how the different types of T cells function during an actual disease or vaccine challenge. It has also greatly inhibited attempts to find T-cell correlates of protection in vaccine or epidemiological work. Fortunately, the technologies discussed here...
represent a tremendous advance in our ability to capture that complexity, and thus we may finally have the tools we need to broadly assess T-cell responses in most situations and to understand their contributions to immunity in much more depth. We anticipate that these approaches will give us much better ways of evaluating new vaccines and immunotherapies in the very near future.

COMPEETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.
T-cell epitopes

Nature America, Inc. All rights reserved.

81. Robins, H.S.
80. Baker, E.S.
79. Fortier, M.H.
78. Marrack, P., Ignatowicz, L., Kappler, J.W., Boymel, J. & Freed, J.H. Comparison of peptides bound to spleen and thymus class II.
77. Rotzschke, O.
76. Yang, J.
75. Rivino, L.
74. Weiskopf, D.
73. Assarsson, E. et al. Immunoinformatics: how in silico methods are re-shaping the investigation of peptide immune specificity.
72. Yang, J.
71. Wee, L.J., Lim, S.J., Ng, L.F. & Tong, J.C. Immunoinformatics: how in silico methods are re-shaping the investigation of peptide immune specificity.
67. Yang, J.
66. Chang, C.X.
65. Assarsson, E.
64. Chang, C.X.
63. Andersen, R.S. et al. Parallel detection of antigen-specific T cell responses by combinatorial encoding of MHC multimers.
62. Moon, J.J. et al. Naive CD4+ T cell frequency varies for different epitopes and predicts repertoire diversity and response magnitude.
61. Lundegaard, C. et al. NetMHC-3.0: accurate web accessible predictions of human, mouse and monkey MHC class I affinities for peptides of length 8–11.
59. Wee, L.J., Lim, S.J., Ng, L.F. & Tong, J.C. Immunoinformatics: how in silico methods are re-shaping the investigation of peptide immune specificity.
58. Rivino, L. et al. Defining CD8+ T cell determinants during human viral infection.
57. Hoof, I. et al. Interdisciplinary analysis of HIV-specific CD8+ T cell responses against variant epitopes reveals restricted TCR promiscuity.
56. Weiskopf, D. et al. Multiplex mapping of CD4+ T cell epitopes using class II tetramers.
55. Liu, C.J. et al. Pairing of T-cell receptor chains via emulsion PCR.
54. Garcia, K.C., Teyton, L. & Wilson, I.A. Structural basis of T cell recognition.
53. Wu, L.C., Tuot, D.S., Lyons, D.S., Garcia, K.C. & Davis, M.M. Two-step binding mechanism for T-cell receptor recognition of peptide MHC.
52. Garcia, K.C., Teyton, L. & Wilson, I.A. Structural basis of T cell recognition.
51. Reiser, J.B.
50. Garcia, K.C., Teyton, L. & Wilson, I.A. Structural basis of T cell recognition.
49. Reiser, J.B.
48. Garcia, K.C., Teyton, L. & Wilson, I.A. Structural basis of T cell recognition.
47. Wu, L.C., Tuot, D.S., Lyons, D.S., Garcia, K.C. & Davis, M.M. Two-step binding mechanism for T-cell receptor recognition of peptide MHC.
46. Garcia, K.C., Teyton, L. & Wilson, I.A. Structural basis of T cell recognition.
45. J. Exp. Med.
44. J. Immunol.
43. Proc. Natl. Acad. Sci.
42. J. Exp. Med.
41. Mol. Biol. 1023, 203–218 (1993).
40. Sollid, L.M. & Jabri, B. Triggers and drivers of autoimmunity: lessons from coeliac disease.
39. Nature 418, 552–556 (2002).
38. Garcia, K.C., Teyton, L. & Wilson, I.A. Structural basis of T cell recognition.
37. J. Immunol.
36. J. Virol.
35. Proc. Natl. Acad. Sci.
34. J. Immunol.
33. J. Exp. Med.
32. Cell Host Microbe
31. Nature Biotechnology
30. J. Immunol.
29. Nat. Rev. Immunol.
28. Mol. Biosyst.
27. Cell Host Microbe
26. Science 319, 103–117 (2008).
25. J. Exp. Med.
24. Mol. Biosyst.
23. Cell Host Microbe
22. Cell Host Microbe
21. Nature 418, 552–556 (2002).
20. J. Immunol.
19. J. Exp. Med.
18. Mol. Immunol.
17. J. Immunol.
16. J. Exp. Med.
15. Mol. Immunol.
14. J. Exp. Med.
13. Mol. Immunol.
12. J. Exp. Med.
11. Mol. Immunol.
10. J. Exp. Med.
9. Mol. Immunol.
8. J. Exp. Med.
7. Mol. Immunol.
6. J. Exp. Med.
5. Mol. Immunol.
4. J. Exp. Med.
3. Mol. Immunol.
2. J. Exp. Med.
1. Mol. Immunol.

© 2014 Nature America, Inc. All rights reserved.