Journal Pre-proofs

Review

T cell receptor sequence clustering and antigen specificity

Milena Vujovic, Kristine Fredlund Degn, Frederikke Isa Marin, Anna-Lisa Schaap-Johansen, Benny Chain, Thomas Lars Andresen, Joseph Kaplinsky, Paolo Marcatili

PII: S2001-0370(20)30330-5
DOI: https://doi.org/10.1016/j.csbj.2020.06.041
Reference: CSBJ 579

To appear in: *Computational and Structural Biotechnology Journal*

Received Date: 23 March 2020
Accepted Date: 27 June 2020

Please cite this article as: M. Vujovic, K.F. Degn, F.I. Marin, A-L. Schaap-Johansen, B. Chain, T.L. Andresen, J. Kaplinsky, P. Marcatili, T cell receptor sequence clustering and antigen specificity, *Computational and Structural Biotechnology Journal* (2020), doi: https://doi.org/10.1016/j.csbj.2020.06.041

T cell receptor sequence clustering and antigen specificity

Abstract

There has been increasing interest in the role of T cells and their involvement in cancer, autoimmune and infectious diseases. However, the nature of T cell receptor (TCR) epitope recognition at a repertoire level is not yet fully understood. Due to technological advances a plethora of TCR sequences from a variety of disease and treatment settings has become readily available. Current efforts in TCR specificity analysis focus on identifying characteristics in immune repertoires which can explain or predict disease outcome or progression, or can be used to monitor the efficacy of disease therapy. In this context, clustering of TCRs by sequence to reflect biological similarity, and especially to reflect antigen specificity have become of paramount importance. We review the main TCR sequence clustering methods and the different similarity measures they use, and discuss their performance and possible improvement. We aim to provide guidance for non-specialists who wish to use TCR repertoire sequencing for disease tracking, patient stratification or therapy prediction, and to provide a starting point for those aiming to develop novel techniques for TCR annotation through clustering.

Keywords: T cell receptor (TCR), clustering, epitope specificity, T cell receptor distance, T cell receptor similarity, T cell repertoire

*Declarations of interest: none

**Funding: MV and TLA acknowledge funding from the Interdisciplinary Synergy Programme Grant from the Novo Nordisk Foundation

[1]DTU HealthTech, Department of Health Technology, Technical University of Denmark, Ørsteds Plads, Building 345C, DK-2800 Kgs. Lyngby, Denmark

[2]UCL Division of Infection and Immunity, University College London, Wing 3.2, Cruciform Building, Gower Street, London, WC1E 6BT, United Kingdom

[3]Ludwig Institute for Cancer Research Ltd, University of Oxford, Nuffield Department of Medicine, Old Road Campus Research Building, Roosevelt Drive, Oxford, OX3 7DQ, United Kingdom

1. Introduction

Understanding T cell biology has long been essential to the study of infectious and autoimmune diseases. More recently, as immunotherapy has joined the traditional pillars of surgery, chemotherapy and radiation, it has also become more and more central to cancer biology.

The advent of high throughput sequencing has opened a new window on to the T cell receptor (TCR) repertoire. While there is much scope for improvement in TCR repertoire sequencing, these experiments are becoming increasingly routine. Two technological developments can be highlighted. First, the commercial availability of repertoire sequencing as a service and in the form of kits. Second, the availability of single cell sequencing. This allows the linking of the α and β (or γ and δ) chains of the TCR, while linking this TCR sequence to a phenotype such as memory or regulatory cell through single cell RNA-seq. Finally, the development of unique molecular identifiers allows for quantitation from sequencing data unbiased by PCR amplifications steps [1]. Because these technologies are now well established, T cells have been sequenced in a plethora of therapeutic and disease settings, as well as healthy control groups, and the data has been deposited on online databases such as the Sequence Read Archive (SRA)[2], VDJdb[3], TCR3d [4] and ImmuneACCESS database [5]. Most sequencing data available are still bulk unpaired α and β TCR sequences, due to the lower throughput and much higher cost of single-cell sequencing platforms.

However, the outstanding question in TCR repertoire analysis remains understanding the relationship between TCR sequence and TCR binding specificity. Sequence data itself contains no direct information on epitope specificity involved. While this may contribute towards models of sequence-binding specificity it will require more focused data sets to make substantial progress. *In silico* annotation of TCR specificity, would, for example, allow tracking of the number and expansion of clones that respond during the natural history of a disease, after vaccination, or during therapy. An example of this application would be to track 'epitope spreading' in response to cancer

immunotherapy[6, 7].

Antigens are presented to T cells in the form of short peptides via the major histocompatibility complex (MHC). There are two classes of MHC, class I, recognised by $CD8^+$ T cells, and class II, recognised by $CD4^+$ T cells. Antigen presentation via MHC I and MHC II differs, as shown in Figure 1. While there is high overall homology between MHC I and MHC II, differences in structure and antigen processing results in shorter peptides (typically 8-13 amino acids) with buried ends presented on MHC I, than on MHC II (usually 10-22 amino acids) [8, 9, 10].

To achieve high specificity and diversity of TCRs that allow for a directed response against a vast number of epitopes. TCRs undergo a stochastic process of V(D)J recombination in the thymus, through which they form three complementarity determining regions (CDRs) on each of the α and β chains [8]. TCR-pMHC complexes adopt diverse conformations, but in the majority of cases it is the loops formed by the CDRs which come into most direct contact with the peptide-MHC complex (pMHC), as shown in Figure 2. In particular the $CDR_\beta 3$ loop, which is also the most diverse in sequence in the TCR, usually accounts for the largest part of contacts with the epitope.

The process of V(D)J recombination has the potential to generate an indefinite number of distinct TCRs. It is estimated that up to 10^{20} distinct TCRs can be generated with biologically significant probability [11, 12]. The human body contains on the order of 10^{11} T cells [13], and little overlap is generally observed between the repertoires in different individuals. It is therefore likely that each individual will respond with a unique set of TCRs to each epitope. A second important consequence of this extraordinary amount of sequence diversity is that many different sequences must code for TCRs which recognise the same epitope. Otherwise, many individuals would end up with no TCR for many antigens. In fact experimental measurements suggest that hundreds, or thousands of TCRs in each individual react with each peptide MHC complex [3].

On the other hand, there are several orders of magnitude more possible epitopes than T cells in an adult human [14]. Consequently, to provide protection against a broad spectrum of pathogens, the limited number of T cells within an individual must

react with broad specificity towards foreign antigens, ignoring self, but simultaneously exhibiting cross-reactivity. In other words, many different TCRs must recognise the same peptide, but each TCR must recognise many peptides. This biological balancing act has made it difficult to understand which TCRs are responsible for an antigen response.

The most direct and detailed method for studying TCR-pMHC binding is X-ray crystallography. The progress in the field has provided very precise knowledge of some TCR-pMHC binding sites. The number of TCR-pMHC structures which have been solved is still limited (less than 100 unique currently available) [15, 16]. One approach to extend this data set is to use structural predictions, based on sequence. Despite the difficulties of modelling flexible loops, such as the CDR regions of the TCR, several tools have been explored, and the field is an active area of research. Models predicting TCR-pMHC binding based on their structure have already been investigated [15, 17, 18, 19, 20].

A number of other techniques probe the nature and quality of the T cell receptor interaction with pMHC. The ELISPOT assay [21] is one of the simplest methods for such an analysis, and has been widely used in assessing the quality of T cell responses. The surface of wells in a well-plate is coated with antibodies designed to capture cytokines secreted upon T cell activation. T cells are added to each of the wells, and upon addition of the antigen the number of activated T cells in each well and the magnitude of their response can be measured by the amount of bound cytokines surrounding each cell [22]. This analysis provides information on both the clonal size and the effector function of activated T cells. Despite the simplicity of the method, its major drawback is that no information is obtained on the TCR sequences of the T cells involved. Furthermore, the number of antigens tested in a single experiment is limited.

The key invention for sequencing of antigen-specific TCR subsets is labelled multimer technologies [23, 24, 25]. These allow for *in vitro* specificity testing and sorting of antigen specific T cells by binding to synthetic conjugates of peptide MHC (pMHC) molecules. The same restriction applies as with ELISPOT, in that there is a limited

number of peptides that can be tested in this manner. However, unlike ELISPOT these T cells can be separated subsequently and sequenced to reveal information on nature of TCRs involved in a response to a single epitope. As the method is fully compatible with sequencing, it provides an unprecedented view into TCR-antigen specificity, by allowing simultaneous collection of information about both the epitope and TCR.

These experimental techniques provide abundant complementary data on TCR-epitope binding. Ideally, to make sense of this plenitude of sequence data one would like to be able to read out which epitope specificities are present in a sample, or in a more restricted way to test for reaction against specific epitopes, using sequence data alone. However, inferring this from primary sequence information is a challenging task as it involves prediction of protein-protein binding without knowledge of exact structures of proteins involved. Still, both TCRs and pMHCs have some defined structure with known variable regions and restricted number of binding conformations. As tertiary and quaternary structure of functional proteins is dependent on their primary sequence, it is reasonable to believe that protein-protein interactions could be inferred from the sequence information alone. Structure prediction of pMHC is relatively straight-forward, unlike the prediction of TCR structure which becomes quite the ordeal due to the high variability of the CDRs. Current TCR structure prediction tools such as LYRA [26] and TCRmodel[4] are able to predict TCR structure with a striking reported accuracy for a protein with such a high degree of variability, with benchmarked average RMSD accuracy of 1.48 Å reported for LYRA. Even though these predictions have not yet reached the accuracy of the related protein family of antibodies, the models are quite useful as they convey information about the true protein structure.

The main challenge is constructing a TCR comparison strategy that will somehow reflect the epitope specificites of TCRs involved, as illustrated in Figure 3. Understanding the complex mechanisms of TCR antigen reactivity and expansion, could lead to correct patient stratification, track response to disease, help guide immunotherapy and further the development of precision medicine. Further, understanding the binding determinants might allow design of TCRs (or vaccines). Currently there are a number of

approaches that aim to cluster TCRs by extrapolating information from their primary sequences to study their specificites.

In the remaining part of this review, we discuss the latest discoveries in the field of TCR specificity and repertoire analysis. We aim to provide a complete overview of all TCR clustering methods and repertoire analysis, their advantages and pitfalls,in hopes of facilitating the choice of data analysis choice for experimentalists and bioinformaticians alike. We aim to showcase all current TCR grouping strategies and their ability to translate into biological similarity or classification of repertoires. It is also our hope that outlining current state-of-the art will facilitate further development of improved TCR clustering techniques.

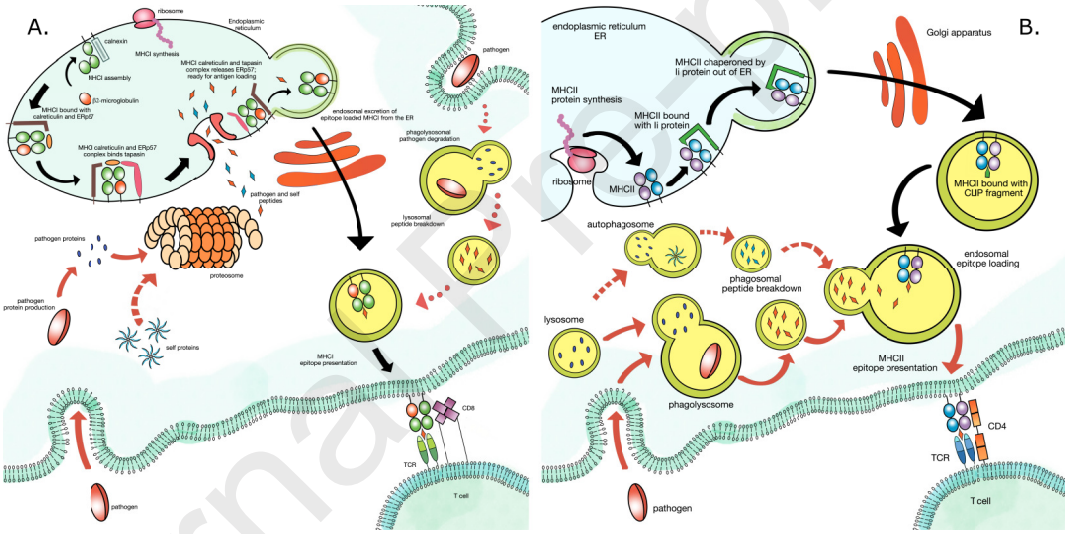

Fig. 1. Schematic representation of MHC antigen processing and presentation adapted from cellular and immunobiological textbooks by Janeway [8] and Abbas et al [27]. The MHC class I or II antigen presenting molecule comes into contact with $CD8^+$ or $CD4^+$ T cells, respectively. The binding to the T cell receptor (TCR), which induces T cell activation, is aided by the CD8 or CD4 protein, for MHC class I or II binding mechanisms, respectively. In both figures black arrows follow MHC synthesis and antigen presentation pathways. Red arrows follow antigen processing: solid - foreign-antigen direct presentation pathway; dashed - self-antigen

direct presentation pathway; dotted - foreign antigen cross-presentation. **A. MHC class I antigen processing and presentation.** MHCI synthesis is started off by the ribosomes in the endoplasmic reticulum (ER). Additional incorporation of β2-microglobulin into the MHCI structure is aided by a transitional complex with the auxiliary protein calnexin. To protect from unsolicited interactions, the newly synthesised MHCI is complexed with calreticulin and ERp57, and subsequently to tapasin which will assist in epitope binding. Upon transporter associated with antigen processing (TAP) protein activation antigens come through into the ER and simultaneously the MHCI-tapasin-calreticulin complex releases ERp57 and widens the peptide binding cleft which allows for binding of compatible epitopes. The loaded complex is released from ER by endosome encapsulation and transported to the cell membrane to be expressed on cell surface. **Foreign and self antigen processing:** Some pathogens survive internalisation and continue to produce proteins in the cytosol. Alternatively, pathogens may be internalised along with their protein product. These proteins are degraded by the proteosome into peptide fragments, epitopes, and sent to the ER for peptide-MHCI assembly and presentation. Foreign epitopes are shown in orange. Self proteins follow a similar pathway of proteosomal degradation and are sent to the ER for peptide-MHC assembly and self presentation. Self epitopes are shown in blue. All nucleated cells express MHCI and follow these pathways for endogenous antigen presentation. **Cross-presentation:** Exogenous antigens are usually presented on MHCII expressing cells. In order to allow for MHCI presentation of exogenous antigens specialised cells process pathogens as in the MHCII pathway, but present on MHCI complexes. Several pathways might be involved in this process. The pathogen is first internalised and enzymatically degraded in the phagolysosome. The lysosome containing peptide antigens then comes into contact with synthesised MHCI molecule and form the peptide-MHCI complex. One possible pathway is that the generated antigens are transported from the lysosome, through TAP and are loaded onto the MHCI in the ER, following which they are expressed on the cell surface. Another pathway might include a vesicular loading compartment detaching from the ER, carrying the synthesised MHCI molecule, and merging with the epitope carrying lysososme. Upon merging the epitopes could load onto the MHCI and express onto the cell surface. **B. MHC class II antigen processing and presentation.** Pathogens are phagocytosed into the cell interior. Upon merging with a lysosome, proteases cleave the pathogen into short peptide fragments - foreign epitopes, here shown in red. The

same fate befalls the cells own proteins as they undergo degradation by the autophagosome, leaving a phagosome containing short peptides - self epitopes, here shown in blue. Meanwhile, the MHCII protein is synthesised by ribosomes in the ER. Upon assembly, MHCII binds invariant chain, Ii protein. It prevents any unwanted protein binding to the MHCII complex in the ER. The Ii chaperones MHCII out of ER in an endosome. In the endosome, due to slightly acidic conditions the Ii protein degrades leaving class II associated invariant chain peptide, CLIP fragment bound in the MHCII cleft. Upon merging with a epitope containing phagosome, the MHCII comes into contact with foreign and self antigen fragments. Upon binding the peptide-MHCII complex is expressed on the cell surface where it is able to bind $CD4^+$ T cells.

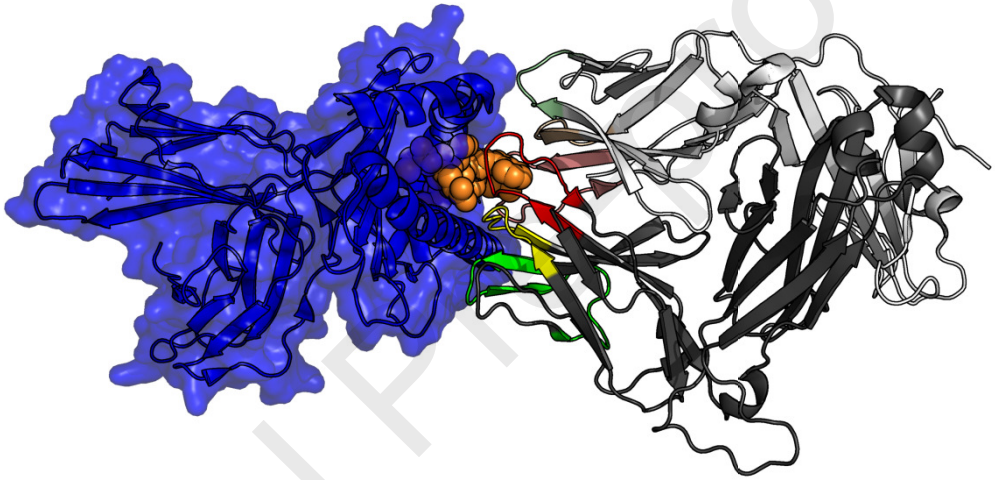

Fig. 2. 3D model of TCR-pMHC complex rendered via PyMOL [28],PDB reference code: 2BNR. MHC (blue) presenting peptide epitope (orange) comes into contact with the TCR (α chain light gray, β chain dark gray). Complementarity determining regions CDR1 (yellow), CDR2 (green) and CDR3 (red) come into contact with the pMHC. CDR3 comes into most contact with the presented peptide, while CDR1 and CDR2 on both chains mostly interact with the MHC.

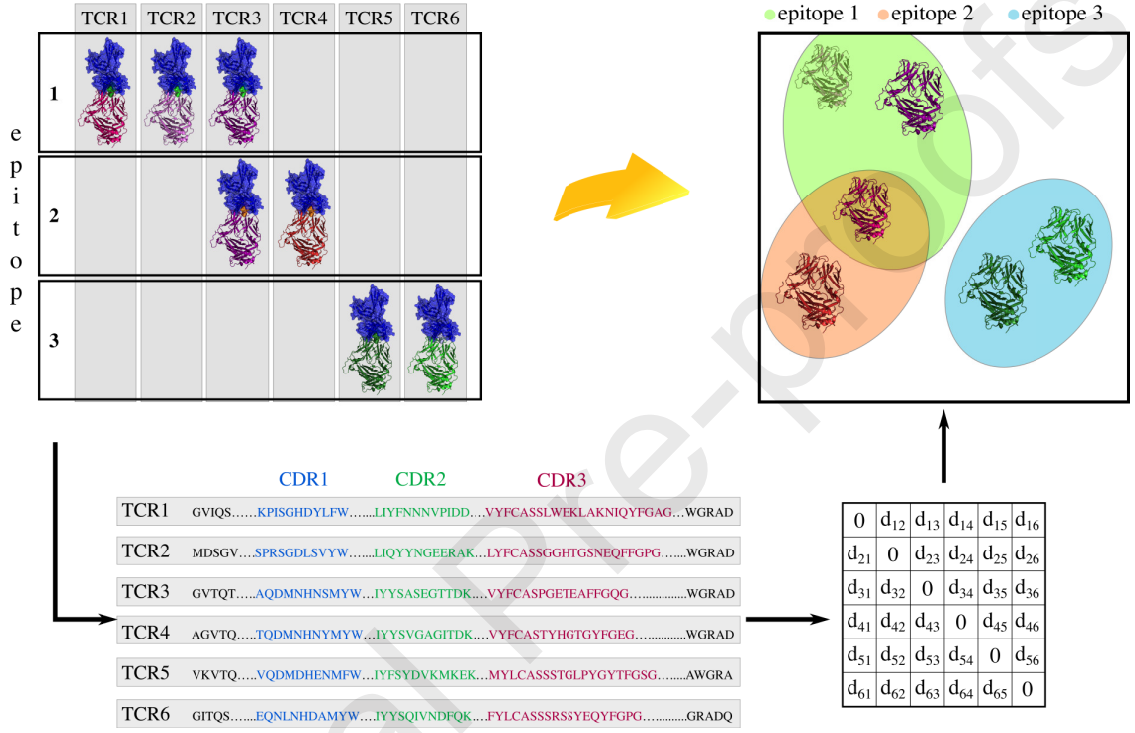

Fig. 3. Graphical representation of attempts to encompass structural and sequence similarity in a suitable clustering distance metric that aims to capture epitope specificity. Binding of six fictional TCRs to three fictional epitopes is depicted on the upper left side. The TCRs are shown in shades of green, purple and red, while epitopes are coloured in green, orange and light purple. If primary sequences of the TCRs are known, sequence comparison can be used to create a distance matrix TCR distance matrix. The matrix could then be used to cluster individual TCRs together based on their sequence similarity, with the goal of clustering by biological similarity i.e. epitope response.

2. Sequencing based approaches

The largest experiments aimed at linking antigen to TCR sequence using multimer technology are now reaching trillions of TCRs [29]. The collaborative approach between Microsoft Healthcare NExT initiative and Adaptive Biotechnologies aims to provide a comprehensive mapping of T cell receptors and their antigen targets covering a multitude of diseases. They aim to unearth biologically and clinically relevant antigens across diseases that can be used for diagnostic purposes from a single blood test. The proof-of-concept study by Emerson et al. [30], outlined initial steps in an diagnostic classification of Cytomegalovirus (CMV) positive and negative individuals by TCR repertoire analysis. A Bayes probabilistic model, based on presence/absence of specific TCRs in 352 CMV negative and 289 CMV negative individuals was used to predict a binary classifier, CMV serostatus. The feature selection and model parameter selection was initially done using cross-validation to provide training and testing sets. The model was also tested on an external validation set of 120 subjects. The authors report excellent classification performance with an AUC of at least 0.93, based on a small (less than 200) set of TCRs over-represented in the CMV+ cohort. This study suggests the potential of TCR sequencing data in disease diagnostics, tracking and treatment in the future. However, it also suggests that very large sequence data sets will be required to provide sufficient power if presence/absence of specific sequences is used, without any attempt to cluster TCR sequences with similar epitope recognition.

2.1. Sequence alignment and clustering approaches

Algorithms which cluster TCRs (or often only CDR3 sequences) exploit similarity measures between TCRs with the aim to identify antigen specificity. In other words, members of a TCR sequence cluster should all recognise the same pMHC. Broadly, the approaches can be divided into those that use global similarities across the whole TCR or CDR3, and local similarities which focus on small amino acid motifs. A common approach of assessing global protein similarity is by sequence alignment and scoring using pre-calculated position specific scoring matrices, such as the BLOSUM [31] and

PAM [32] family of matrices. There exist several alignment algorithms [33, 34, 35] which use a gap introducing penalty and a substitution matrix to align two sequences by their most similar or identical stretches. An important difficulty in alignment of TCRs with known specificities is that TCRs are cross-reactive and may bind multiple very different epitopes. Conversely, a single epitope may be bound by very different TCRs. Moreover, substitution matrices such as BLOSUM and PAM have been derived from studies of evolutionary related proteins. In this case, rather than serving as a measure for evolutionary relatedness of TCRs responding to the same epitope, such matrices provide a useful starting point as a proxy for physico-chemical similarity.

An example of sequence alignment approaches employed in TCR repertoire analysis is the ImmunoMap algorithm [36] (code available at: `github.com/sidhomj/ImmunoMap`). Sidhom et al. evaluate the $CD8^+$ T cell response, from naïve and tumour bearing B6 mice, *in vitro* which bind either self tumor-associated antigen (Kb-TRP2) or a foreign tumor-associeted (Kb-SIY) antigen tetramer nanoparticle artificial antigen presenting cells (nanoAPCs). After β chain sequencing, they create a distance matrix between CDR3 regions using a PAM10 scoring matrix and a large gap penalty and further perform hierarchical clustering on the basis of this distance matrix. The novelty of their clustering approach comes from the visualisation of the dendogram, where the authors add intuitive endings to the branches corresponding to clone sizes. This approach revealed that in the naïve mice response to the self antigen, the expanded T cells in the repertoire were more unrelated and higher frequency than the T cells against the foreign antigen. In tumour bearing mice, the situation altered slightly in the self response with an observed elevated number of high frequency clones as well as usage of distantly related sequences. Following murine sequence analysis, the method was tested in 34 metastatic melanoma patients undergoing α-PD1 immunotherapy (Nivolumab), from whom Tumour Infiltrating Lymphocytes (TILs) were extracted and sequenced. Repertoires were compared prior- and post-therapy, and the authors report observing distinct features on the ImmunoMap dendogram between responders and non-responders, such as the number of high frequency clones and CDR3 relatedness. This was further corrob-

orated by the dominant motif analysis from the expanded ImmunoMap detected clones, which showed some classification power. Although this analysis doesn't seek to assign TCRs to particular epitopes, it conveys a notion of the importance of CDR3 similarity clustering. It also highlights the complexity of response towards even just a single epitope, as assessed by binding to multimer nanoAPC. This graphical approach proves very useful in displaying properties of repertoires with a single specificity; however, it fails to scale up and give an easily readable representation of repertoires at large.

A more focused effort in TCR clustering reflecting epitope specificity comes in the form of TCRdist by Dash et al [37] (code available at: `github.com/phbradley/tcr-dist`). The authors used tetramer staining and single cell sequencing to obtain 4635 paired α and β TCR sequences from 10 different epitope specific repertoires. They analysed data from 78 mice and humans specific for murine and human cytomegalovirus (CMV), influenza and Epstein-Barr virus antigen epitopes. In order to analyse the data they constructed TCRdist, a distance metric based on both the α and β chain of the receptor. It is a similarity weighted mismatch distance using alignment with BLOSUM62 [31] substitution matrix to calculate similarity between CDR regions. Gap penalties are low for the CDR1 and 2 regions, but increase for the CDR3, stemming from the need to conserve short length motifs in the CDR3 regions which might be responsible for binding. Finally a distance between two TCRs is calculated by summing over scores for each CDR region on both chains, as well as an additional variable loop they term CDR2.5. The CDR3 loop scores on both chains is upweighted in the sum, since it is believed to contain most of the information about epitope binding. Using this TCR distance they proceed to cluster TCRs within each epitope-specific repertoire as well as assign TCR sequences from influenza-infected lungs without prior knowledge of their tetramer specificity using nearest-neighbour-distance classifiers. They managed to correctly assign 81% human and 78% murine sequences to their epitope specific repertoire. To the best of our knowledge this is the first specialised single cell TCR similarity measure which use combined α and β chains. However, one limitation of the clustering evaluation is that the metric has not been evaluated on complex repertoires

originating from responses from multiple epitopes.

Another metric, CDRdist, developed by Thakkar et al [38], takes solely CDR3 sequences into account (code available at: `https://github.com/neerjathakkar/Distinguishing-TCR-Groups`). The authors evaluate performance and separately apply their metric on $CDR_{\alpha}3$ and $CDR3_{\beta}$ sequences. To evaluate sequence similarity CDRdist uses local alignment and a substantial gap penalty with BLOSUM45 [31] substitution matrix, usually used for more distantly related alignments than with the higher order BLOSUM matrices. Using this combination of parameters they allow for larger physico-chemical diversity, therefore generating longer matching substrings in the alignments. The authors proceeded to analyse data from monozygotic twins previously published by Zvyagin et al [39]. The original analysis showed that the number of identical CDR3s shared between twins was significantly increased compared to non-twin individuals. Thakkar et al broadened the hypothesis from considering identical sequences, to considering similar sequences, and in fact exclude identical CDR3 sequences from consideration. Applying CDRdist to each CDR3 in the repertoires, they evaluated whether the nearest CDR3 neighbour came from a twin, or another individual. As the number or nearest neighbours coming from twins outweighed those coming from other individuals, they reach the conclusion that twins have more shared similar sequences than non-twins. This finding is perhaps not unexpected, but it strengthened the belief that the CDRdist conveys biological meaning, before proceeding to the more difficult task of epitope classification. Following the approach of Dash et al [37], they try to assign CDR3 sequences to their respective antigen specificity groups from the same epitope-specific repertoires used in Dash et al. by using the nearest neighbour distance classifier. The authors report comparable performance to TCRdist using only $CDR_{\beta}3$ sequences, although they are not able to achieve the same result on $CDR_{\alpha}3$s. They achieve similar performance on the epitope-specific repertoires used for creating and evaluating the GLIPH algorithm [40] which is discussed at length further on. The authors also proceed to classify TCRs by which pathology they come from using data from McPAS-TCR catalogues [41]. They perform reasonably well on classification of infectious diseases (influenza, HIV, yellow

fever and hepatitis C), but are not able to classify on cancers, autoimmune diseases and diabetes. Following closely the evaluation techniques of Dash et al. the authors do not evaluate their metrics classification power on a mixed epitope repertoire.

3. Analysis of characteristic short TCR motifs

The identification of short motifs within TCR sequences provides an alternative to the heavily parametrized sequence alignment and scoring approaches presented above. This approach is rooted in the hot spot interaction hypothesis, which states that only short stretches of complementary amino acid residues are responsible for epitope binding affinity [42, 43, 44]. Using short stretches of amino acids of length k (k-mers) in order to evaluate TCR receptor similarity could reduce informational noise, as opposed to comparing entire sequences. By focusing on short motifs, the problem of gaped alignment in TCRs of different lengths is also circumvented. By using k-mers in various forms, researches are able to pinpoint dominant motifs driving TCR-epitope specificity rather than individual expanded clones.

One such approach is employed in the work of Thomas et al [45] (code available as part of the Supplementary information of the same publication). In the study murine $CD4^+$ T cells were bulk sequenced at different time points following immunisation with killed *Mycobacterium tuberculosis*. Every $CDR_\beta 3$ sequence was encoded as the list of all present triplets (k-mers of length 3). Instead of assessing triplet similarity using substitution matrices, the authors encode each triplet as a set of Atchley factors [46], corresponding to a set of physico-chemical properties. The authors then generate a *triplet codebook*, i.e. a reduced set of representative triplets to describe the complete pooled dataset. This is done by pooling and subsampling triplets from all samples, and grouping them by k means clustering. From each of the resulting clusters of similar triplets, a single representative triplet is selected in order to create the final triplet codebook. Each murine repertoire is then represented as a distribution of triplets in the codebook, by assigning each repertoire triplet vector to the most similar triplet in the codebook. Finally the repertoire representation is converted into a feature vector,

used for classification using hierarchical clustering and Support Vector Machine (SVM) analysis [47]. Both techniques could classify immunised and non-immunised mice, but repertoires taken at different time points from immunised mice were not distinguishable. Although this study does not concern TCR-epitope classification, it highlights the importance of conserved characteristic motifs in assessing epitope responses. The authors note that their results reinforce imporance of diversity of the TCR repertoire, seeing as many private TCRs contribute to the T cell response to the same antigen in genetically identical mice.

A subsequent study combined both global similarity metrics, and local amino acid motifs by Glanville et al. [40] This study evaluated publicly available CDR3s with known specificities, as well as their own pMHC tetramer sorted human $CD4^+$ and $CD8^+$ data (code available at: https://github.com/immunoengineer/gliph). They trained the GLIPH (Grouping of Lymphocyte Interactions by Paratope Hotspots) algorithm to search for enriched conserved TCR motifs of length 2, 3 and 4 within TCR multimer repertoires in the $CDR_\beta 3$ region. The distance metric then combines global and local TCR sequence similarity (CDR3s differing up to 1 amino acid and shared enrichment of motifs, respectively), V gene usage, CDR3 length bias, structural peptide antigen contact propensity and other features, with variable weightings for the different methods. GLIPH was evaluated on a mixture of 8 specificities, where it grouped 94% of the clustered TCRs together with others of same specificity. Another evaluation was performed on $CD4^+$ *Mycobacterium Tuberculosis* specific T cells from 22 patients with latent *M. tuberculosis* infection. Clusters with TCRs shared between 3 individuals or more were examined, and found that 16 specificity groups that were shared between at least 3 individuals included at least 4 uniquely derived βTCR clones. This showed that enrichment of motifs can organise TCRs within or across individuals. Most importantly, the authors state that GLIPH can be used independently of knowing epitope specificity to elucidate novel clusters within repertoires it has not been exposed to previously. Even though GLIPH was validated across patients, it is yet unclear whether or not it will be able to cluster TCRs based on their epitope preference in a mixed epitope

repertoire with unknown specificities.

4. Summary and Outlook

In order to evaluate the performance of the sequence based methods we performed a preliminary comparison using data obtained from VDJdb database taking all human βTCRs paired with their epitope specificities with a VDJdb confidence score above 1. This dataset was split into training and testing datasets based on epitope similarity, so that there are no shared epitopes between the two. The testing set finally consisted of 830 TCRs with known specificity towards one of 28 epitopes. We assessed each method as binary classifiers, based on their ability to cluster together TCRs with identical specificity, and measured their accuracy in terms of Area Under the Roc curve (AUC) [48]. The AUC is 1 for a perfect prediction, and 0.5 for a random prediction. TCRdist was not evaluated as it is calculated considering paired α and β TCR chains simultaneously. Immunomap and CDRdist performed comparably, with an AUC of 0.6449 and 0.6502, respectively. However, when we performed an agglomerative ('bottom-up") hierarchical clustering [49] approach the methods did not reveal any epitope specific clusters. These results are not surprising since both of these methods are based on sequence alignment and scoring techniques on the $CDR_{\beta}3$ region, which is both variable in length and sequence. As mentioned in the introduction, TCRs with very different sequences can bind to the same epitope, and both methods fail at identifying such cases and at forming epitope-specific clusters.

TCRdist contains also information on the CDR1 and CDR2, which come into close contact with the MHC complex. As MHCs also exhibit preference in epitope presentation [50, 51], this provides additional information with respect to methods focused solely on the CDR3 region. Furthermore, TCRdist combines both the alpha and beta chain regions in its analysis, possibly increasing the sensitivity of the method, as both chains are involved in pMHC recognition. On the other hand, this comes at an additional cost, since paired sequencing is still less abundant than bulk sequencing data. Nevertheless, all sequence alignment techniques carry an inherent fault since they can introduce gaps

in the sequences at different positions, rather than focusing on structurally conserved regions in the CDRs that mediate epitope recognition.

The short motif search method has shown remarkable power considering that it does not include entire TCR sequences in the comparison. The short motifs considered are expected to convey a notion of conserved stretches of amino acid sequence coming into contact with the epitope. Which is precisely what the alignment methods are struggling to capture. A difficulty arising in this analysis is that choice of motif length is quite arbitrary. Furthermore, both reviewed analysis focus solely on the $CDR_{\beta}3$ region. Even though GLIPH uses scoring matrices to evaluate similarity of the motifs found in CDR3s, when evaluated on a mixture of eight CDR3 specificities it is not able to cluster all TCRs. Out of the TCRs that were clustered GLIPH is able to group them according to their epitope cluster with 94% accuracy. This remarkable results possibly stems from the fact that epitopes can be evolutionary related, and therefore the short motifs specific to them can in theory reflect this evolutionary similarity. Furthermore, the GLIPH algorithm takes in simultaneously both local motif and global similarity of TCRs capturing more complex characteristics of TCRs. GLIPH is yet to be evaluated on it's predicting power on clustering all the TCRs in the mixture of TCRs with known specificites.

Currently no single tool exists for unequivocal classification of TCR receptor specificity. This is due to two major biological features of the data. Firstly, TCRs are cross-reactive and able to bind multiple antigens with varying affinities. Furthermore, TCR binding is not sufficient to elicit T cell activation. A complex interplay between binding affinity and stability, co-stimulatory signals, and TCR abundancy regulates T cell activation [52]. This underlines the complexity of a T cell antigen response, meaning that clustering to predict epitope specificity might not necessarily show the true state of epitope reactivity. This potentially hampers the intended use of these methods in disease outcome predictions.

Secondly, TCR data, especially CDR3 regions, carry innate redundancy as the termini of CDR3s across individuals share high sequence similarity, that leaves a short

stretch of CDR3 sequence responsible for such a high variability in epitope binding. This similarity comes from V and J genes shared across TCRs and the nature of V(D)J recombination which introduces most sequence variability in the junctions between the individual genes. Upon training a classification method or constructing a similarity metric with the aim of elucidating epitope specificity, much of the dataset will share high similarity with the testing set. Therefore the performance of these methods might plummet dramatically in real-life applications. One possible way to overcome this is by obtaining larger quantities of data than available at present. Higher throughput of technologies which pair TCRs with epitopes, such as multimer technologies, might provide the data necessary to train the more complex machine learning algorithms such as neural networks, to achieve better performance.

Additionally, TCR epitope recognition in reality occurs in three-dimensional space, therefore understanding the complex TCR-pMHC interaction from primary sequence alone is challenging. The importance of including 3D structural information in models for TCR target prediction has already been recognised. [53]. Therefore including TCR structural information into clustering approaches might greatly improve prediction of epitope specificities.

Overall, the rise of availability of bulk and paired $\alpha\beta$ TCR sequencing data offers the opportunity to improve the methods to cluster TCRs and predict their epitope specificities. As TCR data becomes more abundant, the need for higher computing power will rise too. Currently, methods are usually limited to assessing samples of up to 100,000 unique TCR sequences at a time, with subsampling techniques readily employed to increase the analysis speed. When we reach the aspired goal for the amount of TCR epitope annotated data, the machines currently available to most researchers will not carry sufficient computational power to perform such tasks. However, technological advances will ensue, which will allow even more computing power to be readily available to a wide population of scientists and empower researchers for even larger scale data analysis.

References

[1] I. Z. Mamedov, O. V. Britanova, I. V. Zvyagin, M. A. Turchaninova, D. A. Bolotin, E. V. Putintseva, Y. B. Lebedev, D. M. Chudakov, Preparing Unbiased T-Cell Receptor and Antibody cDNA Libraries for the Deep Next Generation Sequencing Profiling, Frontiers in Immunology 4 (DEC) (2013) 456. doi:10.3389/fimmu.2013.00456.
URL http://journal.frontiersin.org/article/10.3389/fimmu.2013.00456/abstract

[2] R. Leinonen, H. Sugawara, M. Shumway, o. b. o. t. I. N. S. D. Collaboration, The Sequence Read Archive, Nucleic Acids Research 39 (suppl_1) (2010) D19–D21. doi:10.1093/nar/gkq1019.
URL https://doi.org/10.1093/nar/gkq1019

[3] M. Shugay, D. V. Bagaev, I. V. Zvyagin, R. M. Vroomans, J. C. Crawford, G. Dolton, E. A. Komech, A. L. Sycheva, A. E. Koneva, E. S. Egorov, A. V. Eliseev, E. Van Dyk, P. Dash, M. Attaf, C. Rius, K. Ladell, J. E. McLaren, K. K. Matthews, E. Clemens, D. C. Douek, F. Luciani, D. van Baarle, K. Kedzierska, C. Kesmir, P. G. Thomas, D. A. Price, A. K. Sewell, D. M. Chudakov, VDJdb: a curated database of T-cell receptor sequences with known antigen specificity, Nucleic Acids Research 46 (D1) (2017) D419–D427. doi:10.1093/nar/gkx760.
URL https://doi.org/10.1093/nar/gkx760

[4] R. Gowthaman, B. G. Pierce, TCR3d: The T cell receptor structural repertoire database, Bioinformatics 35 (24) (2019) 5323–5325. doi:10.1093/bioinformatics/btz517.
URL https://doi.org/10.1093/bioinformatics/btz517

[5] immuneACCESS Data (2020).
URL https://clients.adaptivebiotech.com/immuneaccess

[6] A. Memarnejadian, C. E. Meilleur, C. R. Shaler, K. Khazaie, J. R. Bennink, T. D. Schell, S. M. M. Haeryfar, PD-1 Blockade Promotes Epitope Spreading in Anticancer CD8 + T Cell Responses by Preventing Fratricidal Death of Subdominant Clones To Relieve Immunodomination , The Journal of Immunology 199 (9) (2017) 3348–3359. doi:10.4049/jimmunol.1700643.
URL http://www.ncbi.nlm.nih.gov/pubmed/28939757http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC5731479

[7] J. L. Gulley, R. A. Madan, R. Pachynski, P. Mulders, N. A. Sheikh, J. Trager, C. G. Drake, Role of Antigen Spread and Distinctive Characteristics of Immunotherapy in Cancer Treatment., Journal of the National Cancer Institute 109 (4) (2017). doi:10.1093/jnci/djw261.
URL http://www.ncbi.nlm.nih.gov/pubmed/28376158http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC5441294

[8] K. K. M. Murphy, C. Weaver, Janeway's immunobiology, 9th Edition, Garland Science, Taylor & Francis Group, LLC, 2016.

[9] J. G. Abelin, D. B. Keskin, S. Sarkizova, C. R. Hartigan, W. Zhang, J. Sidney, J. Stevens, W. Lane, G. L. Zhang, T. M. Eisenhaure, K. R. Clauser, N. Hacohen, M. S. Rooney, S. A. Carr, C. J. Wu, Mass Spectrometry Profiling of HLA-Associated Peptidomes in Mono-allelic Cells Enables More Accurate Epitope Prediction, Immunity 46 (2) (2017) 315–326. doi:10.1016/j.immuni.2017.02.007.

[10] T. Trolle, C. P. McMurtrey, J. Sidney, W. Bardet, S. C. Osborn, T. Kaever, A. Sette, W. H. Hildebrand, M. Nielsen, B. Peters, The Length Distribution of Class I–Restricted T Cell Epitopes Is Determined by Both Peptide Supply and MHC Allele–Specific Binding Preference, The Journal of Immunology 196 (4) (2016) 1480–1487. doi:10.4049/jimmunol.1501721.
URL http://www.ncbi.nlm.nih.gov/pubmed/26783342http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4744552

[11] D. J. Laydon, C. R. Bangham, B. Asquith, Estimating T-cell repertoire diversity: Limitations of classical estimators and a new approach, Philosophical Transactions of the Royal Society B: Biological Sciences 370 (1675) (8 2015). doi:10.1098/rstb.2014.0291.

[12] M. M. Davis, P. J. Bjorkman, T-cell antigen receptor genes and T-cell recognition (1988). doi:10.1038/334395a0.

[13] M. K. Jenkins, H. H. Chu, J. B. McLachlan, J. J. Moon, On the composition of the preimmune repertoire of T cells specific for Peptide-major histocompatibility complex ligands., Annual review of immunology 28 (1) (2010) 275–94. doi:10.1146/annurev-immunol-030409-101253.
URL http://www.ncbi.nlm.nih.gov/pubmed/20307209

[14] A. K. Sewell, Why must T cells be cross-reactive? (9 2012). doi:10.1038/nri3279.

[15] K. K. Jensen, V. Rantos, E. C. Jappe, T. H. Olsen, M. C. Jespersen, V. Jurtz, L. E. Jessen, E. Lanzarotti, S. Mahajan, B. Peters, M. Nielsen, P. Marcatili, TCRpMHC-models: Structural modelling of TCR-pMHC class I complexes, Scientific Reports 9 (1) (2019) 1–12. doi:10.1038/s41598-019-50932-4.

[16] J. Leem, S. H. de Oliveira, K. Krawczyk, C. M. Deane, STCRDab: the structural T-cell receptor database, Nucleic Acids Research 46 (D1) (2017) D406–D412. doi:10.1093/nar/gkx971.
URL https://doi.org/10.1093/nar/gkx971

[17] E. Lanzarotti, P. Marcatili, M. Nielsen, Identification of the cognate peptide-MHC target of T cell receptors using molecular modeling and force field scoring, Molecular Immunology 94 (2018) 91–97. doi:10.1016/j.molimm.2017.12.019.
URL http://www.ncbi.nlm.nih.gov/pubmed/29288899http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC5800965

[18] B. G. Pierce, Z. Weng, A flexible docking approach for prediction of T cell receptor-peptide-MHC complexes, Protein Science 22 (1) (2013) 35–46. doi:10.1002/pro.2181.
URL http://doi.wiley.com/10.1002/pro.2181

[19] I. H. Liu, Y. S. Lo, J. M. Yang, Genome-wide structural modelling of TCR-pMHC interactions, BMC Genomics 14 (Suppl 5) (2013) S5. doi:10.1186/1471-2164-14-S5-S5.
URL http://bmcgenomics.biomedcentral.com/articles/10.1186/1471-2164-14-S5-S5

[20] T. Hoffmann, A. Marion, I. Antes, DynaDom: structure-based prediction of T cell receptor inter-domain and T cell receptor-peptide-MHC (class I) association angles, BMC Structural Biology 17 (1) (2017) 1–19. doi:10.1186/s12900-016-0071-7.

[21] J. D. Sedgwick, P. G. Holt, The ELISA-plaque assay for the detection and enumeration of antibody-secreting cells. An overview, Journal of Immunological Methods 87 (1) (1986) 37–44. doi:10.1016/0022-1759(86)90341-8.

[22] D. D. Anthony, P. V. Lehmann, T-cell epitope mapping using the ELISPOT approach, Methods 29 (3) (2003) 260–269. doi:10.1016/S1046-2023(02)00348-1.

[23] J. D. Altman, P. A. H. Moss, P. J. R. Goulder, D. H. Barouch, M. G. McHeyzer-Williams, J. I. Bell, A. J. McMichael, M. M. Davis, Phenotypic analysis of antigen-specific T lymphocytes. Science. 1996. 274: 94-96, Journal of immunology (Baltimore, Md. : 1950) 187 (1) (2011) 7–9.
URL https://pubmed.ncbi.nlm.nih.gov/21690331

[24] R. S. Andersen, P. Kvistborg, T. Mørch Frøsig, N. W. Pedersen, R. Lyngaa, A. H. Bakker, C. J. Shu, P. T. Straten, T. N. Schumacher, S. R. Hadrup, Parallel detection of antigen-specific t cell responses by combinatorial encoding of MHC multi-

mers, Nature Protocols 7 (5) (2012) 891–902. doi:10.1038/nprot.2012.037.
URL http://www.ncbi.nlm.nih.gov/pubmed/22498709

[25] S. R. Hadrup, A. H. Bakker, C. J. Shu, R. S. Andersen, J. van Veluw, P. Hombrink, E. Castermans, P. thor Straten, C. Blank, J. B. Haanen, M. H. Heemskerk, T. N. Schumacher, Parallel detection of antigen-specific T-cell responses by multidimensional encoding of MHC multimers, Nature Methods 6 (7) (2009) 520–526. doi:10.1038/nmeth.1345.

[26] M. S. Klausen, M. V. Anderson, M. C. Jespersen, M. Nielsen, P. Marcatili, LYRA, a webserver for lymphocyte receptor structural modeling., Nucleic acids research 43 (W1) (2015) 349–55. doi:10.1093/nar/gkv535.
URL http://www.ncbi.nlm.nih.gov/pubmed/26007650http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4489227

[27] A. K. Abbas, A. H. Lichtman, S. Pillai, D. L. M. i. Baker, A. Baker, Cellular and molecular immunology, Elsevier, 2017.

[28] Schrödinger LLC, The PyMOL Molecular Graphics System, Version˜1.8 (11 2015).

[29] TCR-Antigen Map – Adaptive Biotechnologies (2020).
URL https://www.adaptivebiotech.com/partnerships/antigen-map/

[30] R. O. Emerson, W. S. DeWitt, M. Vignali, J. Gravley, J. K. Hu, E. J. Osborne, C. Desmarais, M. Klinger, C. S. Carlson, J. A. Hansen, M. Rieder, H. S. Robins, Immunosequencing identifies signatures of cytomegalovirus exposure history and HLA-mediated effects on the T cell repertoire, Nature Genetics 49 (5) (2017) 659–665. doi:10.1038/ng.3822.

[31] S. Henikoff, J. G. Henikoff, Amino acid substitution matrices from protein blocks, Proceedings of the National Academy of Sciences of the United States of America 89 (22) (1992) 10915–10919. doi:10.1073/pnas.89.22.10915.

[32] M. O. Dayhoff, M. O. Dayhoff, R. M. Schwartz, Chapter 22: A model of evolutionary change in proteins, IN ATLAS OF PROTEIN SEQUENCE AND STRUCTURE (1978).
URL http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.145.4315

[33] S. B. Needleman, C. D. Wunsch, A general method applicable to the search for similarities in the amino acid sequence of two proteins, Journal of Molecular Biology 48 (3) (1970) 443–453. doi:10.1016/0022-2836(70)90057-4.

[34] T. F. Smith, M. S. Waterman, Identification of common molecular subsequences, Journal of Molecular Biology 147 (1) (1981) 195–197. doi:10.1016/0022-2836(81)90087-5.

[35] O. Gotoh, An improved algorithm for matching biological sequences, Journal of Molecular Biology 162 (3) (1982) 705–708. doi:10.1016/0022-2836(82)90398-9.

[36] J. W. Sidhom, C. A. Bessell, J. J. Havel, A. Kosmides, T. A. Chan, J. P. Schneck, ImmunoMap: A bioinformatics tool for T-cell repertoire analysis, Cancer Immunology Research 6 (2) (2018) 151–162. doi:10.1158/2326-6066.CIR-17-0114.

[37] P. Dash, A. J. Fiore-Gartland, T. Hertz, G. C. Wang, S. Sharma, A. Souquette, J. C. Crawford, E. B. Clemens, T. H. O. Nguyen, K. Kedzierska, N. L. La Gruta, P. Bradley, P. G. Thomas, Quantifiable predictive features define epitope-specific T cell receptor repertoires, Nature 547 (7661) (2017) 89–93. doi:10.1038/nature22383.
URL http://www.nature.com/doifinder/10.1038/nature22383

[38] N. Thakkar, C. Bailey-Kellogg, Balancing sensitivity and specificity in distinguishing TCR groups by CDR sequence similarity, BMC Bioinformatics 20 (1) (2019) 1–14. doi:10.1186/s12859-019-2864-8.

[39] I. V. Zvyagin, M. V. Pogorelyy, M. E. Ivanova, E. A. Komech, M. Shugay, D. A. Bolotin, A. A. Shelenkov, A. A. Kurnosov, D. B. Staroverov, D. M. Chudakov,

Y. B. Lebedev, I. Z. Mamedov, Distinctive properties of identical twins' TCR repertoires revealed by high-throughput sequencing, Proceedings of the National Academy of Sciences of the United States of America 111 (16) (2014) 5980–5985. doi:10.1073/pnas.1319389111.

[40] J. Glanville, H. Huang, A. Nau, O. Hatton, L. E. Wagar, F. Rubelt, X. Ji, A. Han, S. M. Krams, C. Pettus, N. Haas, C. S. Arlehamn, A. Sette, S. D. Boyd, T. J. Scriba, O. M. Martinez, M. M. Davis, Identifying specificity groups in the T cell receptor repertoire, Nature 547 (7661) (2017) 94–98. doi:10.1038/nature22976.

[41] N. Tickotsky, T. Sagiv, J. Prilusky, E. Shifrut, N. Friedman, McPAS-TCR: a manually curated catalogue of pathology-associated T cell receptor sequences., Bioinformatics (Oxford, England) 33 (18) (2017) 2924–2929. doi:10.1093/bioinformatics/btx286.
URL `http://www.ncbi.nlm.nih.gov/pubmed/28481982`

[42] T. Clackson, J. A. Wells, A hot spot of binding energy in a hormone-receptor interface, Science 267 (5196) (1995) 383–386. doi:10.1126/science.7529940.

[43] S. Ovchinnikov, H. Kamisetty, D. Baker, Robust and accurate prediction of residue-residue interactions across protein interfaces using evolutionary information, eLife 2014 (3) (5 2014). doi:10.7554/eLife.02030.

[44] D. S. Marks, T. A. Hopf, C. Sander, Protein structure prediction from sequence variation (11 2012). doi:10.1038/nbt.2419.

[45] N. Thomas, K. Best, M. Cinelli, S. Reich-Zeliger, H. Gal, E. Shifrut, A. Madi, N. Friedman, J. Shawe-Taylor, B. Chain, Tracking global changes induced in the CD4 T-cell receptor repertoire by immunization with a complex antigen using short stretches of CDR3 protein sequence., Bioinformatics (Oxford, England) 30 (22) (2014) 3181–8. doi:10.1093/bioinformatics/btu523.

URL http://www.ncbi.nlm.nih.gov/pubmed/25095879http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4221123

[46] W. R. Atchley, W. M. Fitch, A. D. Fernandes, T. Drüke, A natural classification of the basic helix-loop-helix class of transcription factors., Proceedings of the National Academy of Sciences of the United States of America 94 (10) (1997) 5172–6. doi:10.1073/pnas.94.10.5172.
URL http://www.ncbi.nlm.nih.gov/pubmed/9144210http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC24651http://www.pnas.org/cgi/doi/10.1073/pnas.94.10.5172

[47] C. Cortes, V. Vapnik, Support-vector networks, Machine Learning 20 (3) (1995) 273–297. doi:10.1007/bf00994018.

[48] J. A. Swets, Measuring the accuracy of diagnostic systems, Science Science 240 (4857) (1988) 1285–1293. doi:10.1126/science.3287615.

[49] F. Nielsen, Hierarchical Clustering, in: Undergraduate Topics in Computer Science, Springer, Cham, 2016, pp. 195–211. doi:10.1007/978-3-319-21903-5_8.

[50] V. Jurtz, S. Paul, M. Andreatta, P. Marcatili, B. Peters, M. Nielsen, NetMHCpan-4.0: Improved Peptide–MHC Class I Interaction Predictions Integrating Eluted Ligand and Peptide Binding Affinity Data, The Journal of Immunology 199 (9) (2017) 3360–3368. doi:10.4049/jimmunol.1700893.

[51] V. I. Jurtz, L. E. Jessen, A. K. Bentzen, M. C. Jespersen, S. Mahajan, R. Vita, K. K. Jensen, P. Marcatili, S. R. Hadrup, B. Peters, M. Nielsen, NetTCR: sequence-based prediction of TCR binding to peptide-MHC complexes using convolutional neural networks, bioRxiv (2018) 433706doi:10.1101/433706.
URL https://www.biorxiv.org/content/early/2018/10/02/433706

[52] J. Gálvez, J. J. Gálvez, P. García-Peñarrubia, Is TCR/pMHC Affinity a Good Estimate of the T-cell Response? An Answer Based on Predictions

From 12 Phenotypic Models, Frontiers in Immunology 10 (MAR) (2019) 349. doi:10.3389/fimmu.2019.00349.
URL https://www.frontiersin.org/article/10.3389/fimmu.2019.00349/full

[53] E. Lanzarotti, P. Marcatili, M. Nielsen, T-Cell Receptor Cognate Target Prediction Based on Paired α and β Chain Sequence and Structural CDR Loop Similarities, Frontiers in Immunology 10 (AUG) (2019) 2080. doi:10.3389/fimmu.2019.02080.
URL https://www.frontiersin.org/article/10.3389/fimmu.2019.02080/full