

Limits on Inferring T-cell Specificity from Partial Information

James Henderson,^{1,2} Yuta Nagano,^{1,3} Martina Milighetti,^{1,4} and Andreas Tiffeau-Mayer^{1,2}

¹*Division of Infection and Immunity*

²*Institute for the Physics of Living Systems*

³*Division of Medicine*

⁴*Cancer Institute*

University College London

A key challenge in molecular biology is to decipher the mapping of protein sequence to function. To perform this mapping requires the identification of sequence features most informative about function. Here, we quantify the amount of information (in bits) that T-cell receptor (TCR) sequence features provide about antigen specificity. We identify informative features by their degree of conservation among antigen-specific receptors relative to null expectations. We find that TCR specificity synergistically depends on the hypervariable regions of both receptor chains, with a degree of synergy that strongly depends on the ligand. Using a coincidence-based approach to measuring information enables us to directly bound the accuracy with which TCR specificity can be predicted from partial matches to reference sequences. We anticipate that our statistical framework will be of use for developing machine learning models for TCR specificity prediction and for optimizing TCRs for cell therapies. The proposed coincidence-based information measures might find further applications in bounding the performance of pairwise classifiers in other fields.

Mapping the amino acid sequence of a particular T-cell receptor (TCR) to its antigen specificity is a holy grail of systems immunology [1–3]. The T-cell receptor endows T-cells with the ability recognize snippets of pathogenic material presented on the surface of antigen presenting cells by major histocompatibility complexes (MHC) [4]. TCRs are specific, meaning a given T-cell will only activate in response to a select range of antigen stimuli. Coverage of the vast antigen space explored by evolving pathogens is enabled by immense sequence variation within the TCR [5, 6], in particular within six hypervariable loops of the heterodimeric receptor, named complementary determining regions (CDRs).

The immense diversity of TCRs implies that many have no experimentally determined ligands [7]. Emerging computational approaches predict the specificity of such orphan TCRs by their sequence similarity to annotated TCRs [1, 3, 8, 9]. However, which level of partial matching is sufficient for reliable prediction has remained unclear. Moreover, there is substantial interest in understanding for which immunological questions knowledge of paired receptor chains obtainable by single-cell sequencing is worth the trade-off with the higher throughput achievable by bulk sequencing [10], and which TCR features are most informative for machine learning applications [3, 11].

Here, we address these important open questions by putting universal limits on the accuracy with which TCR specificity can be predicted from partial information. Our work takes inspiration from a long history of successful applications of information theory to the study of complex biological input-output relationships from neural coding [12–14] and transcriptional regulation [15, 16] to pattern formation during embryo development [17–19]. Following recent applications of information theory to TCR repertoires by us [20] and others [21], our analysis builds on a fundamental insight from evolution-

ary biology: patterns of sequence conservation in protein families provide clues about functionally relevant properties. In the immunological context, this means that TCR features that are important for specific recognition of a particular epitope will be most highly conserved among epitope-specific TCRs relative to their global diversity (Fig. 1).

In our current work, we provide the first comprehensive map of how much information each section of the paired chain TCR sequence provides about its specificity. To provide such a map, we make use of two recent datasets that have sequenced TCRs specific to a dozen viral MHC class I epitopes [1, 22]. We overcame statistical limitations of prior analyses of pairs of residues [20, 21, 23] using coincidence-based measures of repertoire diversity [24]. These measures can be estimated from smaller samples than traditional measures based on Shannon entropy [25–27]. The information-theoretic approach naturally allowed us to identify synergies between different TCR sections in determining antigen specificity. Importantly, our quantification of coincidence information is underpinned by theory that directly links achievable classification accuracy to the coincidence information gained from a partial match and prior beliefs about the prevalence of epitope-specific T cells in a repertoire.

I. AN INFORMATION-THEORETIC APPROACH TO T CELL SPECIFICITY

A. Coincidence analysis for features

We have recently introduced a coincidence-based statistical framework to measure antigen-driven selection in TCR repertoires [24]. The main idea of this work was to quantify clonal convergence by counting how often pairs of independently recombined clonal lineages in a sample

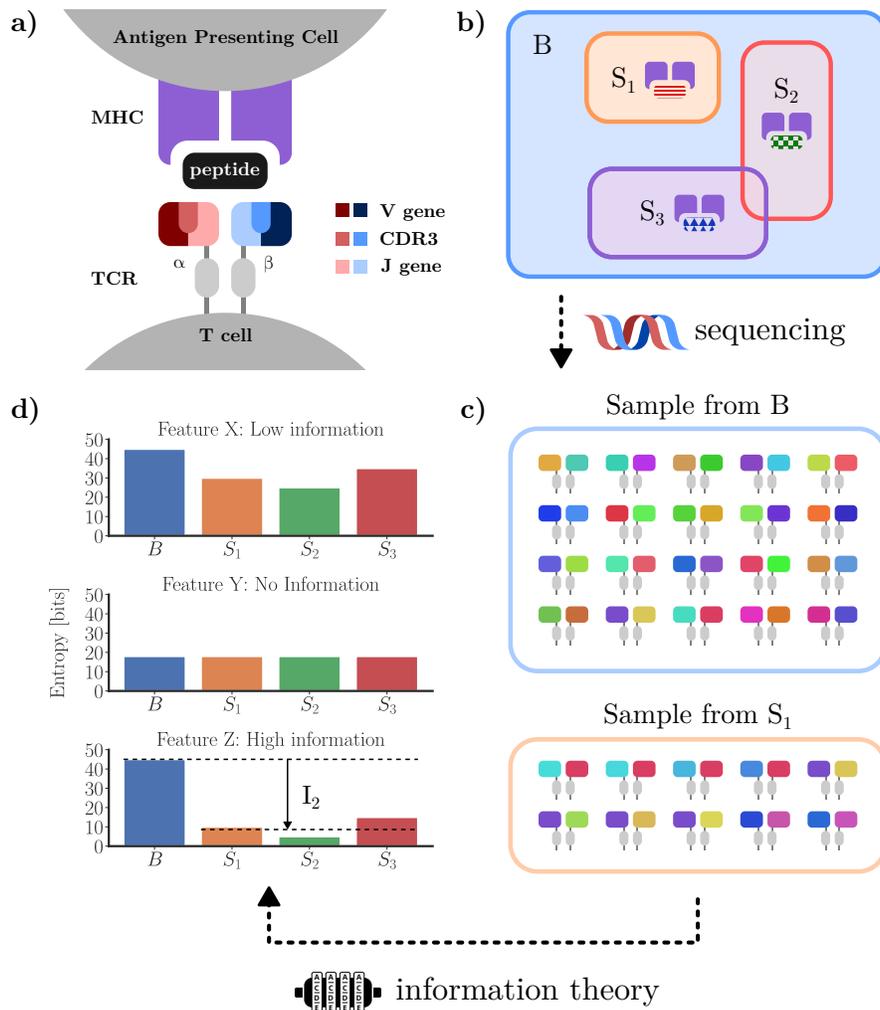


FIG. 1. **Overview of analysis methodology.** **a)** Sketch of T-cell receptor structure highlighting the V, CDR3 and J regions and their interaction with MHC-bound peptides. The TCR is composed of two chains, most commonly α and β chains. Each chain in turn is comprised of a V (variable), J (joining) and C (constant) gene, with the addition of a D (diversity) gene in the β chain. Within each chain, the CDR1 and CDR2 amino acid loops are coded for by the V gene while the CDR3 regions are at the V(D)J intersection, which is additionally diversified through the random insertion and deletion of nucleotides at gene template junctions. **b)** An abstracted view of TCR sequence space. The set B includes all possible TCRs. The subsets S_i represent TCRs specific to particular ligands. **c)** Sequencing TCR from either the whole repertoire or epitope-specific subsets gives us samples from their respective distributions. **d)** The number of pairs which match in a particular feature may then be recorded to compute a probability of coincidence. The logarithm of the probability of coincidence gives a measure of the entropy of the feature. Our information theoretic approach quantifies the change in entropy between background TCRs and sets of specific TCRs of different features (top to bottom). Features which experience a large reduction in entropy (bottom) are the most informative for predicting the epitope specificity of a sequence.

have TCRs that are more similar to each other than some threshold level. Here, we pursue a conceptually related but novel approach that considers near-coincidences as coincidences on the level of coarse-grained TCR features. A feature may be a gene segment choice at a given locus, an amino acid at a particular residue, or a physical property of a hypervariable loop such as its charge or length. Features may also contain other features such as the α chain containing the $V\alpha$, $J\alpha$, and the $CDR3\alpha$ as component features.

Mathematically, a feature is a random variable that

maps the sample space of all TCR sequences to a discrete set of possible categories. We denote the distribution of feature values for randomly drawn TCRs from a repertoire by $P(X)$. The probability that two independent draws return the same outcome, i.e. the probability of coincidence of X , is then defined by

$$p_C[P(X)] = \sum_x P(x)^2, \quad (1)$$

where $P(x)$ represents $P(X = x)$ and the sum runs over all possible outcomes of X . We recall that in ecology

$p_C[P(X)]$ is referred to as the Simpson's diversity index of X with $D(X) = 1/p_C[P(X)]$ being an effective number of distinct species in a population [28]. Intuitively, we expect the most informative features to be those whose diversity is most reduced among TCRs specific to the same epitope when compared with background TCRs. In the following we will make this intuition mathematically precise using a coincidence-based formulation of information theory.

B. Coincidence entropy

A central quantity in information theory is entropy. The entropy of a probability distribution $P(X)$ is given in its form proposed by Shannon in 1948 as [29]

$$H[P(X)] = \sum_x P(x) \log P(x). \quad (2)$$

Entropy represents the average amount of information lacking about the outcome of a measurement of random variable X . It is usually calculated with the logarithm taken to base 2 such that its units are in bits and all logarithms in this following should be understood as logarithms taken with respect to this base. In 1961, Renyi showed that by relaxing one of the Shannon-Khinchin axioms from which the mathematical form of entropy is uniquely derived (strong additivity), a more general expression for entropy may be obtained [30, 31]

$$H_\alpha[P(X)] = \frac{1}{1-\alpha} \log \left(\sum_x P(x)^\alpha \right), \quad (3)$$

where α is referred to as the order of the Renyi entropy. The family of Renyi entropies include Shannon's entropy measure as the limit of $\alpha \rightarrow 1$.

We may note that the probability of coincidence introduced in the previous subsection provides a measure for the Renyi entropy of order $\alpha = 2$

$$H_2[P(X)] = -\log p_C[P(X)]. \quad (4)$$

The Renyi entropy of order 2 is known as collision entropy in cryptography and may also be motivated from an optimal code length perspective with non-linearly weighted length penalties [32]. Here we use the term *coincidence entropy* to stress its relation to coincidence-counting among sample pairs.

C. Coincidence mutual information

We have previously used the coincidence ratio $p_C[P(X|\Pi)]/p_C[P(X)]$ between specific and background TCRs as a measure of antigen-driven selection [24], where $p_C[P(X|\Pi)]$ is the probability of coincidence among epitope-specific TCRs averaged over a collection

of epitopes, Π , and $p_C[P(X)]$ the probability of coincidence among background TCRs. Different definitions of conditional Renyi entropy for $\alpha \neq 1$ have been proposed. Here we follow [33, 34] and define

$$H_2[P(X|Y)] = -\log p_C[P(X|Y)], \quad (5)$$

where $p_C[P(X|Y)]$ is an average of $p_C[P(X|y)]$ over all outcomes of Y

$$p_C[P(X|Y)] = \sum_y \rho_2(y) p_C[P(X|y)], \quad (6)$$

with weighting factors

$$\rho_2(y) = \frac{P(y)^2}{\sum_y P(y)^2}. \quad (7)$$

Detailed justification for these definitions is provided in Appendix A. This definition allows us to express the coincidence probability ratio in terms of coincidence entropies

$$\log \left(\frac{p_C[P(X|\Pi)]}{p_C[P(X)]} \right) = H_2[P(X)] - H_2[P(X|\Pi)]. \quad (8)$$

We note that for Shannon entropy this difference defines the mutual information between X and Π [29], which motivates the following definition of *coincidence mutual information*

$$I_2(X, \Pi) = \log \left(\frac{p_C[P(X|\Pi)]}{p_C[P(X)]} \right). \quad (9)$$

Importantly, our definition of conditional entropy maintains additivity $H_2[P(X, Y)] = H_2[P(X)] + H_2[P(Y|X)]$, where $P(X, Y)$ is the joint distribution of the random variables X and Y . As a correlate it follows that coincidence mutual information is symmetric, $I_2(X, Y) = I_2(Y, X)$ – as is its Shannon counterpart – so it tells us not only how much information we gain about sequence features upon learning their epitope specificity, but also, by symmetry, how much information a sequence feature provides about its epitope specificity. Coincidence mutual information thus provides a natural way to score the importance of a TCR feature in determining specificity, which we will refer to as the feature *relevancy*.

D. Describing the interactions between features with redundancy and synergy

The connection between coincidence analysis and information theory naturally allows us to apply additional notions from information theory [35, 36] to describe how multiple features work in tandem to provide antigen specificity. First, *conditional mutual information*

$$I_2(X, \Pi|Y) = H_2[P(X|Y)] - H_2[P(X|\Pi, Y)], \quad (10)$$

describes the remaining information provided by feature X given that the value of a second feature Y is already

known. Here, $H_2[P(X|\Pi, Y)]$ indicates conditioning on both epitope specificity and feature Y . If $I_2(X, \Pi|Y) = 0$ then we refer to X as a fully *redundant* feature in the context of Y . As a trivial example, knowledge of the complete primary sequence of the full α chain makes any information provided by CDR3 α redundant, and so on.

Second, *interaction information*

$$I_{2,\text{int}}(X, Y|\Pi) = I_2([X, Y], \Pi) - I_2(X, \Pi) - I_2(Y, \Pi) \quad (11)$$

describes how much additional information both features provide in conjunction (Appendix B). Here, $I_2([X, Y], \Pi)$ is the relevancy of the feature produced by combining the two features X and Y . If $I_{2,\text{int}}(X, Y|\Pi) > 0$ then there is *synergy* between the two features.

II. BOUNDING CLASSIFICATION ACCURACY OF PARTIAL TCR MATCHES

A. Pairwise classification odds

There are well-known connections between information measures and achievable classification errors both in the Shannon [37] and Renyi case [38, 39]. In the following we derive how TCR classification accuracy using partial feature matches with a reference sequence is bounded when only partial information is available. We consider a classification setting, where the task is to identify spiked-in TCR sequences specific to a particular epitope π in an otherwise naive repertoire. We will derive how posterior classification odds depend on feature relevancy and prior beliefs, i.e. the fraction of spiked-in sequences $P(\pi)$. Mathematically, in this setting the presence of a TCR sequence σ is due to either of two generative processes:

$$P(\sigma) = P(\pi)P(\sigma|\pi) + P(B)P(\sigma|B), \quad (12)$$

where $P(B) = (1 - P(\pi))$, and where $P(\sigma|\pi)$ describes the distribution of TCR sequences specific to epitope π and $P(\sigma|B)$ the distribution of background TCR sequences according to V(D)J recombination.

To recapitulate the empirical procedure of matching TCR sequences to a database of known binders, we consider the following one-shot classification strategy: We classify a query sequence as having been generated from $P(\sigma|\pi)$, if it matches in a feature X with a reference sequence randomly drawn from $P(\sigma|\pi)$. Using the odds formulation of Bayes' theorem, we may express the posterior odds of correct classification as

$$\frac{P(\pi|x = x')}{P(B|x = x')} = \frac{P(x = x'|\pi) P(\pi)}{P(x = x'|B) P(B)}. \quad (13)$$

Here, $P(x = x'|\pi) = p_C[P(X|\pi)]$ is the probability of a match in feature X if both sequences were truly drawn from distribution $P(\sigma|\pi)$, while $P(x = x'|B)$ is the probability of a match in feature X for a query drawn from $P(\sigma|B)$ and a reference drawn from $P(\sigma|\pi)$. Under the

assumption that the propensity of a TCR for specific binding is independent of its recombination probability [24], one can show that $P(x = x'|B) = p_C[P(X)]$ (Appendix C 1). Therefore

$$\frac{P(\pi|x = x')}{P(B|x = x')} = \frac{p_C[P(X|\pi)] P(\pi)}{p_C[P(X)] P(B)} \quad (14)$$

This expression can be generalized for mixtures of multiple epitope groups, in which case the average odds over epitopes (Appendix C 2) can be expressed as

$$\left\langle \frac{P(\pi|x = x')}{P(B|x = x')} \right\rangle = \frac{p_C[P(X|\Pi)]}{p_C[P(X)]} \left\langle \frac{P(\pi)}{P(B)} \right\rangle, \quad (15)$$

where $p_C[P(X|\Pi)]$ is the conditional probability of coincidence defined previously and the averages for the odds are taken over $P(\pi)/(1 - P(B))$. By the definition of coincidence mutual information (Eq. 9) we can rewrite the last equation as

$$\mathbb{O}_{\text{post}} = 2^{I_2(X, \Pi)} \mathbb{O}_{\text{prior}}, \quad (16)$$

which links the average posterior odds \mathbb{O}_{post} to average prior odds $\mathbb{O}_{\text{prior}}$ via coincidence mutual information. Each bit of coincidence mutual information between X and Π corresponds to a two-fold gain in posterior odds.

B. When is partial information sufficient?

Eq. 16 captures an important Bayesian intuition about classification: Correct classification depends not only on how much information we have available, but also on our prior belief. Here, our prior belief about the likelihood that any particular sequence is specific should reflect the total fraction of spiked-in sequences. If we are searching for a needle in a haystack, this is when $\mathbb{O}_{\text{prior}}$ is small, we need to use more highly informative features for correct classification. Mathematically, a minimal prior odds of $2^{-I_2(X, \Pi)} T$ is needed to ensure that the average posterior odds exceeds a threshold value T .

Expressed in terms of prior probabilities

$$P_{\text{prior}}(I_2) \geq \frac{T 2^{-I_2}}{1 + T 2^{-I_2}}, \quad (17)$$

is needed if only I_2 bits are available for classification. To illustrate this result, we performed in-silico simulations with a toy model of TCR specificity (Appendix C 4). These simulations showed close agreement between predicted values for $P_{\text{prior}}(I_2)$ and those obtained through numerical simulation (Fig. S1).

Note that sequences drawn from $P(\sigma|B)$ may also be specific to π . Therefore, $P(\pi)$ and $P(\pi|x = x')$ are not exactly equal to the fraction of sequences specific to π and the posterior probability of specificity, respectively. However, as shown in Appendix C 3 in most cases of practical interest, where $P(\pi)$ exceeds the background frequency of sequences specific to a given epitope, this distinction is irrelevant.

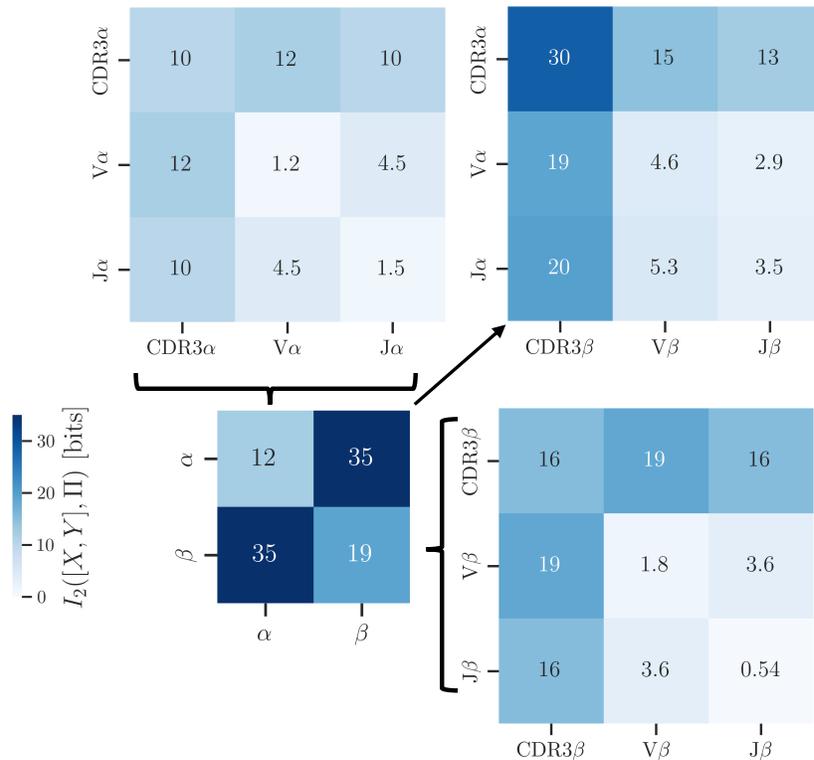


FIG. 2. **Coincidence mutual information between TCR sections and antigen specificity.** Relevancy scores of various sections of the T-cell receptor sequence. The off-diagonal values indicate the amount of coincidence information that combinations of features provide. The top right hand grid shows the relevancy of combination of features where one is from the α chain and the other the β chain. Interaction information and conditional mutual information between features can be computed by taking the difference between the off-diagonals and the sum of the corresponding diagonal values. In particular, positive interaction information is observed between the α and β chains and the CDR3 and V regions indicating synergy between these features while negative interaction information is seen between the CDR3 and J regions indicating redundancy.

III. APPLICATION OF THE METHODOLOGY TO TCR SEQUENCE DATA

We applied our framework to a curated set of multimer-sorted TCRs from CD8⁺ T cells with specificity to viral antigens (described in detail in Appendix E). We combined TCRs specific to nine SARS-CoV-2 epitopes from Minervina *et al.* [22] with TCRs specific to three epitopes from other viruses epitopes from Dash *et al.* [1] (Table I). To obtain a background TCR dataset we randomly paired TCR α and TCR β sequences generated by a computational model of VDJ recombination [40].

A. A decomposition of TCR specificity into its component parts

To provide a top-down decomposition of the information content of the TCR, we computed the relevancy of different sections of the TCR for its specificity, as well as their combinations (Fig. 2). We first analyzed the information provided by the α and β chains alone, which re-

capitulated the expected greater relevancy of the β chain (19 bits) than the α chain (12 bits). By Eq. 17 the information provided by each chain bounds prior probabilities needed for accurate classification using single chain matches. A β chain match requires a prior probability $P_{\text{prior}} \geq 3 \cdot 10^{-5}$ for a 95% posterior confidence. In contrast, an α chain match allows reliable classification only for prior probabilities $P_{\text{prior}} \geq 3 \cdot 10^{-3}$. We then broke down the two chains further into their component V and J gene segments and CDR3 amino acid sequence. A CDR3 β match provides 16 bits of information (corresponding to $P_{\text{prior}} \geq 4 \cdot 10^{-4}$) while a CDR3 α match provides only 10 bits of information (corresponding to $P_{\text{prior}} \geq 1 \cdot 10^{-2}$).

To assess variations in feature relevancy across epitopes, we defined *local relevancy* as the information gain for a specific epitope π ,

$$i_2(X, \pi) = \log \left(\frac{p_C[P(X|\pi)]}{p_C[P(X)]} \right). \quad (18)$$

Local relevancy scores revealed a broadly consistent hierarchy of feature relevancy across epitopes (Figs. S2 - S6).

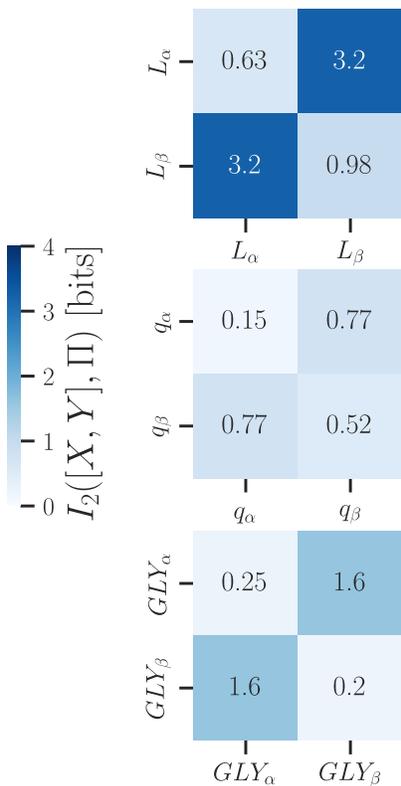


FIG. 3. **Coincidence mutual information between physical properties of the TCR sequence and antigen specificity.** Relevancy scores of CDR3 length, CDR3 net charge and glycine content computed for the α and β chains taken independently and combined. Although each feature has modest relevancy when considered independently, these features all display substantial synergy demonstrating how physical complementarity underlies overall chain pairing constraints.

The analysis also identified variability in local relevancy of features between different epitopes not explained by finite sampling deviations alone in line with our prior findings on a subset of the studied epitopes [27]. We will analyze this variability in more detail in section III D.

B. CDR3 length, net charge and glycine content as features

In addition to analysing features such as the CDR3, V and J regions, we wished to show that our definition of a feature may extend to physical properties of the TCR. To illustrate this, we computed the coincidence mutual information for the length of the CDR3 loop, its net charge and its glycine content (Fig. 3 and Figs. S7 - S9), all features that had been described in the literature as be-

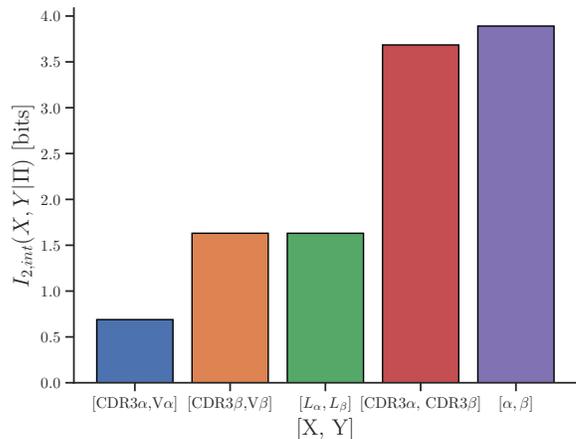


FIG. 4. **Synergistic TCR sequence features.** Interaction information scores for combinations of features computed from Figures 2 and 3. Positive interaction information indicates that two features become more informative in the context of one another and hence have synergy.

ing important for epitope specificity [1, 41]. Our results confirm that each of these summary physical properties of the CDR3 has some relevancy in determining TCR specificity. For instance, CDR3 β net charge is roughly as informative as J β choice. However, no individual CDR3 property captures a substantial proportion of CDR3 information demonstrating the contribution of higher order sequence features to specific binding.

C. Synergy and redundancy between TCR features

Comparing relevancy scores for individual and combined features revealed the pervasiveness of interactions between TCR sections (Fig. 2) and CDR3 features (Fig. 3), where their combined information differed from the sum of their individual relevancies. Fig. 4 summarizes the interaction information between important TCR features.

Our analysis identified substantial synergy between the α and β chains (4 bits). This synergy implies that there are pairing restriction between α and β chains in specific TCRs, which make each chain more informative when considered in its full paired chain sequence context (Appendix B). These results broaden our prior findings to a broader set of epitopes [24], and add to a growing literature investigating TCR α - β pairing rules [20, 24, 41–44]. Pairing restrictions imply that the diversity of TCRs responding to a given epitope is lower than the product of the diversities of responsive α and β chains.

We also analysed the interaction information between the CDR3 of each chain and the corresponding V segment choice. We again identified substantial synergy, presumably reflecting spatial constraints between V-gene encoded framework and CDR1/2 variability and CDR3 choice. In contrast, the interaction information between

the CDR3 and J gene is negative (Figure 2). This is expected as the sequence variability provided by the J gene is contained in the CDR3 region [45], but demonstrates how our framework can identify redundant features without such a prior knowledge. We furthermore discovered substantial synergy between the lengths of CDR3 α and CDR3 β regions. Similarly, glycine-content – which is thought to be involved in CDR3 loop flexibility [41] – also showed synergy across chains. Collectively, these findings identify some of the key physical constraints that underlie the observed global α - β pairing restrictions.

D. Variability in interaction information across epitopes is explained by mixture models

To better understand potential sources of variability of TCR sequence restriction across epitopes we defined additional measures of local sequence variation: *Local conditional mutual information* $i_2(X, \pi|Y) = H_2[P(X|Y)] - H_2[P(X|\pi, Y)]$ and *local interaction information* $i_{2,\text{int}}(X, Y|\pi) = i_2([X, Y], \pi) - i_2(X, \pi) - i_2(Y, \pi)$. We then analysed dependencies across four variables (Fig. 5): Interaction information $i_{2,\text{int}}(\alpha, \beta|\pi)$, α -chain relevancy $i_2(\alpha, \pi)$, β -chain relevancy $i_2(\beta, \pi)$ and paired chain relevancy $i_2([\alpha, \beta], \pi)$. These analyses highlighted strong dependencies between the variables. The more informative an α chain or β chain is for a given epitope, the less α - β interaction information contributes to global diversity restriction (Fig. 5a,b). Moreover, epitopes with more informative α chains also have more informative β chains (Fig. 5c) and more informative full TCR sequences (Fig. 5d).

Unexpectedly, all variables were highly correlated with each other and well-fitted by linear regressions, suggesting the existence of a single underlying degree of freedom that drives the observed variability across epitopes. Based on the clustering of epitope-specific TCRs we had previously proposed mixture of motif models [24], in which epitope-specific TCRs are composed of a number of distinct binding solutions (binding modes or motifs). We asked whether variability in the number of such motifs across epitopes might provide the common degree of freedom explaining the observed correlations. Deriving the expected theoretical relationships between variables (Appendix D 1), we found an increased local interaction information for epitopes with more binding modes and a decrease in individual feature relevance. Across all variable pairs studied in Fig. 5 the mixture model predicted linear relations with slopes of ± 1 , in good agreement with the best fit lines to the empirical data. Intuitively, if an epitope has multiple binding solutions, more α and β chains will be able to bind it, given the right complementary chain (thus lowering the information from each individual chain). At the same time, where many solutions exist a high degree of α - β pairing is expected as most α chains from one binding solution would not be valid with β -chains from another solution (thus increas-

ing the observed synergy between the two chains).

Given the low prevalence of epitope-specific TCRs in a repertoire, we expect the dataset to be a mixture containing some false positive TCRs with no or low affinity to the epitope of interest even if sorting is highly specific. As we show in Appendix D 2 variations in the proportion of false positives across epitopes provide an alternative explanation of the observed dependencies among variables, with high interaction information for epitopes with many false positives. Both models share the common underlying insight that epitope-specific repertoires are mixtures rather than draws from a unimodal distribution – future research might elucidate the contributions of the different underlying mechanisms to the observed variability.

IV. DISTANCE METRICS AND NEAR-COINCIDENCE ENTROPY

A. Generalization of coincidence mutual information to fuzzy matches

As exact matches are rare for complex features, it is of interest to also quantify the information provided by fuzzy feature matches. – As previously explored in Ref. [24], we are not limited to computing the probability of exact coincidences between features but can also consider near-coincidences according to some distance metric. Given a feature X distributed according to $P(X)$ and a distance metric $d(x, x')$ between outcomes x and x' , the probability that two draws from $P(X)$ are at distance $d(x, x') = \Delta$ can be defined as

$$p_C[P(X)](\Delta) = \sum_{x, x'} P(x)P(x')\delta_{d(x, x'), \Delta}, \quad (19)$$

where $\delta_{d(x, x'), \Delta}$ is the Kronecker delta. We use this measure to propose a *near-coincidence entropy* $H_2[P(X)](\Delta) = -\log p_C[P(X)](\Delta)$, and a *near-coincidence conditional entropy* $H_2[P(X|Y)](\Delta) = -\log p_C[P(X|Y)](\Delta)$, where $p_C[P(X|Y)](\Delta)$ once again is an average of $p_C[P(X|y)](\Delta)$ over outcomes of Y using the $\rho_2(y)$ weighting factor. We define a *near-coincidence mutual information*

$$I_2(\Delta_{X, X'}, Y) = H_2[P(X)](\Delta) - H_2[P(X|Y)](\Delta), \quad (20)$$

where $\Delta_{X, X'}$ denotes that this information is computed for near-coincidences in feature X at distance Δ . As this measure deals with pairs of instances of a random variable rather than single instances, this quantity cannot be defined straightforwardly for Shannon entropy but is motivated naturally when using coincidence entropies.

B. Pairwise classification using fuzzy matches

To obtain an interpretation of near-coincidence entropy, we turn once again to pairwise classification. We

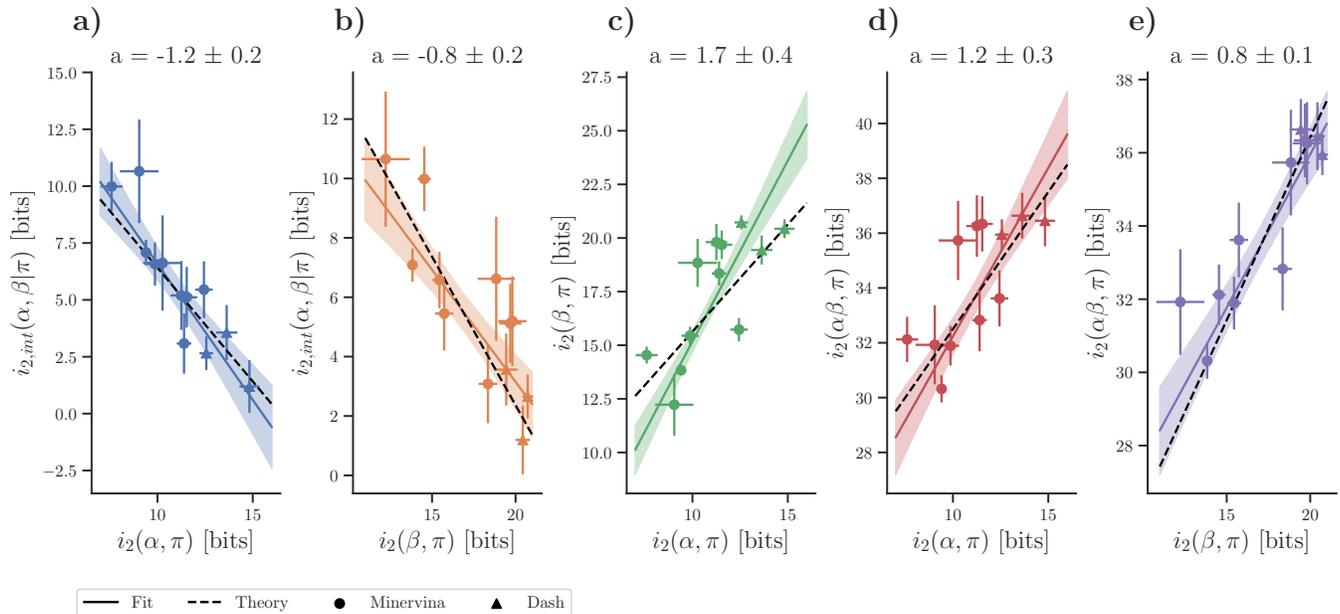


FIG. 5. **Correlation between α - β interaction information and per-chain information across epitopes.** Local interaction information and single chain information across epitopes. Weighted linear fits (solid lines) obtained using orthogonal distance regression were used to quantify the dependence between variables, with regression slopes a displayed above each panel. Epitope-specific interaction information depends negatively on the local informational value of the **a)** α chain and **b)** β . We furthermore find that the **c)** per-chain relevancies are positively correlated with each other as is **d),e)** total information with both single chain relevancies. The observed dependencies between variables agree well with theoretical expectations from a mixture model (dashed lines), in which epitopes differ in the number of distinct binding solutions or contain false positives.

consider the same classification procedure as previously, but based on a fuzzy match where the sequence with unknown specificity is distance Δ from the sequence with known specificity such that $d(x, x') = \Delta$. Similarly to our prior derivations, we find (Appendix C 5)

$$\mathbb{O}_{\text{post}} = 2^{I_2(\Delta_{x, x'}, \Pi)} \mathbb{O}_{\text{prior}}. \quad (21)$$

One bit of near-coincidence mutual information again corresponds to an average two-fold increase in posterior classification odds. As in the case of exact matching, Eq. 21 defines regimes in which fuzzy matches at a given distance are expected to succeed or fail. In Figure 6, we provide an example of this by computing the required prior fraction of specific sequences to obtain a posterior probability of 0.95 with fuzzy CDR3 matches at a certain Levenshtein distance. Inversely, at a given prior odds ratio and target posterior odds ratio we can use these results to compute a critical distance beyond which classification becomes unreliable.

V. DISCUSSION

The ubiquity of information theory lies in its ability to describe complex relationships between data points using a simple quantitative vocabulary. As shown by Shannon [29], entropy provides the most natural measure of

uncertainty and hence changes in entropy directly capture how knowledge of one event increases understanding of another. The application of information theory to the problem of immune receptor specificity has proved highly fruitful in the past. In particular, estimates of residue Shannon entropy aided in identifying potential complementary determining regions of the TCR and immunoglobulin and highlighted that TCRs were the more diverse of these two antigen receptors [46]. Other, more recent studies have employed concepts from information theory such as mutual information to quantify interactions between various sections of the TCR sequence [20, 21]. These previous studies have taken a ‘bottom-up’ approach, computing an upper bound on sequence diversity by summing up the entropy of each constituent amino acid residue or pairs of residues. In part, this ‘bottom-up’ approach has been required due to biases in estimating Shannon entropy in small samples. Although there exist methods for reducing bias in Shannon entropy estimation, these still require resolving higher order distribution moments or essentially resort to coincidence counting and use Renyi entropy to approximate Shannon’s [47, 48]. In this work, we have proposed a ‘top-down’ approach to decomposing TCR specificity firmly rooted in second order Renyi entropy.

Our methodology provides a general framework to assess the role of individual TCR sequence features in determining antigen specificity as well as combinations of

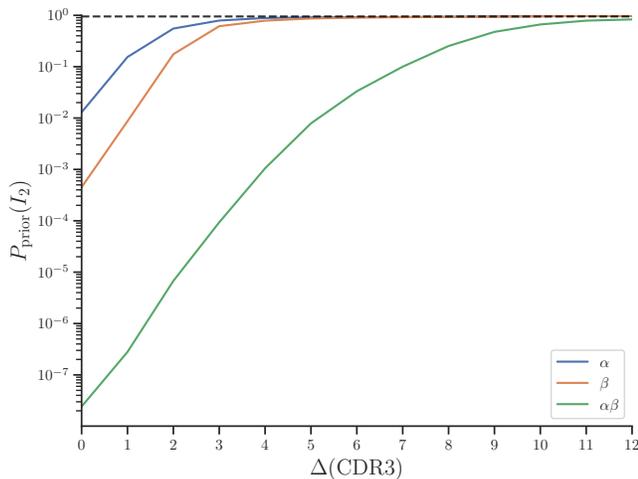


FIG. 6. Information theoretic analysis of fuzzy CDR3 matches. Critical prior probabilities for 95% confidence in classification using a fuzzy match with a Levenshtein distance Δ . Distances are CDR3 α and CDR3 β amino acid edit distances as well as the sum of CDR3 α +CDR3 β edit distance. Levenshtein distances are defined as the minimum number of insertions/deletions and substitutions required to turn one sequence into another.

features by introducing the concepts of relevancy, redundancy and synergy. We first compute the entropy of the full TCR sequence, divide this into its two constituent amino acid chains and then further sub-divide these into their V, J and CDR3 regions. Our results identify the β chain as the most informative of the heterodimeric TCR’s chains and the CDR3 regions to be the most informative regions of each chain. However, we also find that the information these constituent parts provide is far smaller than that of the full TCR sequence. Although these results are unsurprising, with previous work highlighting the higher contribution of the β chain in epitope binding predictions and the importance of paired chain data [23, 49, 50], we provide the first full quantification of the information contained within these regions and, as our methodology has its foundations in coincidence-based statistics, we are able to directly interpret information measures in terms of achievable pairwise classification accuracy. Our work thus paves the way for the development of principled Bayesian methods for interpreting partial sequence matches.

Our results provide clear guides for when a limited amount of TCR sequence information, such as a single chain, is enough to solve an epitope specificity classification problem and when this loss of information may seriously impact predictive performance. We expect these insights to be important for experimental design, to decide whether the time and cost trade-off of single cell sequencing over bulk are worth the increase in information paired chain information might provide.

We have also shown how the vocabulary of information theory can be applied to TCR near-coincidence analy-

sis, which we have introduced in recent work [24]. Our framework predicts pairwise classification performance when using fuzzy matches at a given threshold TCR distance. This approach may be used to define relevant data regimes in which current or future distance metrics [1, 51] may be usefully applied and allows setting critical distances for classifying or clustering sequences [52].

Our ‘top-down’ approach allows us to compute interaction information, which describes synergistic and redundant relationships between TCR sequence features. We observe positive interaction information, synergy, between the α and β chain as well as the CDR3 and V regions, while knowledge of CDR3 regions makes their associated J regions redundant. In addition, we highlight the flexibility of our methodology by assessing the information provided by CDR3 length, net charge and glycine content. We find positive synergy between the lengths of the CDR3 α and CDR3 β amino acid chains and their glycine content. We furthermore show how the relationship between interaction and single chain information across epitopes is compatible with a model in which epitopes vary in the number of distinct binding solutions (or possibly in the rate of false positives). With the steady accumulation of data on more epitopes, we envisage that our approach will help decipher principles underlying sequence space organization of responding TCRs.

The next steps for applying our theoretical approach are numerous. On the practical side, we propose completing the ‘top-down’ approach and performing an analysis of the informational value of the CDR3 sequences residue by residue. This may allow for the identification of informationally dense regions of the CDR3s and for a quantification of more complex allosteric interactions present across the receptor structure. Further extensions of our framework could account for the hierarchy of selective processes shaping the TCR repertoire by varying the background used to compute background entropy. For example, to bound the performance of multi-class classification between a set of known epitopes, it may be more appropriate to quantify the entropy of TCRs across the chosen epitope-specific groups. Likewise, sequence statistics in a naive T cell repertoire could be used as background to account for the imprint of thymic selection. Our information theoretical tools may also be used on problems other than epitope specificity. For example, one may apply them to the study of TCR-HLA associations [53, 54] or TCR sequence to phenotype relationships [44, 55–57].

Linking feature information to classification isn’t a problem unique to the field of protein function, nor is the task of class prediction from pairwise comparisons. Transformer neural networks, the architecture underlying the current rise of large language models, embed data in high dimensional vector spaces [58] and may be trained in a pairwise contrastive manner, such that items from the same class are closer together than items from different classes [59–62]. More generally, metric and representation learning commonly utilise pairwise measures

to tackle problems ranging from sentence embeddings to facial recognition [63–67]. Our pairwise coincidence information measure may be applicable for identifying interpretable informative features in these applications.

To conclude, we have introduced a theoretical framework for mapping the information content of the T-cell receptor sequence with regards to its antigen specificity. Our results confirm prior insights from more limited structural studies regarding the relative importance of the α and β chains [2, 5, 68–71], but also highlight unexpected variability in the synergy between chains across epitopes. As dataset sizes continue to increase the proposed framework will be able to guide the development of machine-learning models for predicting

TCR specificity from sequence by providing means to find interpretable physical features and performance bounds where information is limited.

Acknowledgements. The authors thank Benny Chain, Curtis Callan, Lisa Wagar, Naresha Saligrama, and Sankalan Bhattacharyya for useful discussions over the course of this work. The work of JH and ATM was supported in parts by funding by the Wellcome Leap HOPE Program and the Royal Free Charity. YN and MM were supported by Cancer Research UK studentships under grants BCCG1C8R and A29287, respectively.

-
- [1] P. Dash, A. J. Fiore-Gartland, T. Hertz, G. C. Wang, S. Sharma, A. Souquette, J. C. Crawford, E. B. Clemens, T. H. Nguyen, K. Kedzierska, *et al.*, *Nature* **547**, 89 (2017).
- [2] 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, and M. M. Davis, *Nature* **547**, 94 (2017).
- [3] D. Hudson, R. A. Fernandes, M. Basham, G. Ogg, and H. Koohey, *Nature Reviews Immunology*, 1 (2023).
- [4] M. M. Davis and P. J. Bjorkman, *Nature* **334**, 395 (1988).
- [5] J. Rossjohn, S. Gras, J. J. Miles, S. J. Turner, D. I. Godfrey, and J. McCluskey, *Annual Review of Immunology* **33**, 169 (2015).
- [6] A. Mayer, V. Balasubramanian, T. Mora, and A. M. Walczak, *Proceedings of the National Academy of Sciences* **112**, 5950 (2015).
- [7] M. Goncharov, D. Bagaev, D. Shcherbinin, I. Zvyagin, D. Bolotin, P. G. Thomas, A. A. Minervina, M. V. Pogorelyy, K. Ladell, J. E. McLaren, *et al.*, *Nature methods* **19**, 1017 (2022).
- [8] M. V. Pogorelyy and M. Shugay, *Frontiers in Immunology* **10**, 1 (2019).
- [9] W. D. Chronister, A. Crinklaw, S. Mahajan, R. Vita, Z. n, Z. Yan, J. A. Greenbaum, L. E. Jessen, M. Nielsen, S. Christley, L. G. Cowell, A. Sette, and B. Peters, *Front Immunol* **12**, 640725 (2021).
- [10] S. Valkiers, N. de Vrij, S. Gielis, S. Verbandt, B. Ogunjimi, K. Laukens, and P. Meysman, *ImmunoInformatics* **5**, 100009 (2022).
- [11] Z. S. Ghoreyshi and J. T. George, *Frontiers in Immunology* **14** (2023).
- [12] S. Laughlin, *Zeitschrift für Naturforschung c* **36**, 910 (1981).
- [13] N. Brenner, S. P. Strong, R. Koberle, W. Bialek, and R. R. d. R. v. Steveninck, *Neural computation* **12**, 1531 (2000).
- [14] S. E. Palmer, O. Marre, M. J. Berry, and W. Bialek, *Proceedings of the National Academy of Sciences* **112**, 6908 (2015).
- [15] J. B. Kinney, A. Murugan, C. G. Callan Jr, and E. C. Cox, *Proceedings of the National Academy of Sciences* **107**, 9158 (2010).
- [16] N. M. Belliveau, S. L. Barnes, W. T. Ireland, D. L. Jones, M. J. Sweredoski, A. Moradian, S. Hess, J. B. Kinney, and R. Phillips, *Proceedings of the National Academy of Sciences* **115**, E4796 (2018).
- [17] J. O. Dubuis, G. Tkačik, E. F. Wieschaus, T. Gregor, and W. Bialek, *Proceedings of the National Academy of Sciences* **110**, 16301 (2013).
- [18] M. D. Petkova, G. Tkačik, W. Bialek, E. F. Wieschaus, and T. Gregor, *Cell* **176**, 844 (2019).
- [19] L. McGough, H. Casademunt, M. Nikolić, Z. Aridor, M. D. Petkova, T. Gregor, and W. Bialek, *PRX Life* **2**, 013016 (2024).
- [20] M. Milighetti, *Analysis of T cell receptor sequence and structure to understand the drivers of antigen specificity*, Ph.D. thesis, UCL (University College London) (2023).
- [21] A. M. Xu, W. Chour, D. C. DeLucia, Y. Su, A. J. Pavlovitch-Bedzyk, R. Ng, Y. Rasheed, M. M. Davis, J. K. Lee, and J. R. Heath, *Cell systems* **14**, 273 (2023).
- [22] A. A. Minervina, M. V. Pogorelyy, A. M. Kirk, J. C. Crawford, E. K. Allen, C.-H. Chou, R. C. Mettelman, K. J. Allison, C.-Y. Lin, D. C. Brice, *et al.*, *Nature Immunology* **23**, 781 (2022).
- [23] C. T. Boughter and M. Meier-Schellersheim, *PLOS Computational Biology* **19**, e1011577 (2023).
- [24] A. Mayer and C. G. Callan Jr, *Proceedings of the National Academy of Sciences* **120**, e2213264120 (2023).
- [25] S. k. Ma, *Journal of Statistical Physics* **26**, 221 (1981).
- [26] I. Nemenman, *Entropy* **13**, 2013 (2011).
- [27] A. Tiffeau-Mayer, *arXiv preprint arXiv:2310.03439* (2023).
- [28] E. H. Simpson, *Nature* **163**, 688 (1949).
- [29] C. E. Shannon, *The Bell system technical journal* **27**, 379 (1948).
- [30] A. Rényi, in *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics*, Vol. 4 (University of California Press, 1961) pp. 547–562.
- [31] A. Y. Khinchin, *Mathematical Foundations of Information Theory* (Courier Corporation, 1957).
- [32] L. L. Campbell, *Information and control* **8**, 423 (1965).
- [33] P. Jizba and T. Arimitsu, *Annals of Physics* **312**, 17 (2004).
- [34] V. M. Ilić and M. S. Stanković, *Physica A: Statistical Mechanics and its Applications* **411**, 138 (2014).

- [35] J. R. Vergara and P. A. Estévez, Neural computing and applications **24**, 175 (2014).
- [36] P. L. Williams and R. D. Beer, arXiv preprint arXiv:1004.2515 (2010).
- [37] T. M. Cover and J. A. Thomas, *Elements of Information Theory* (Wiley, Hoboken, NJ, 2005).
- [38] M. Ben-Bassat and J. Raviv, IEEE Transactions on Information Theory **24**, 324 (1978).
- [39] I. Csiszár, IEEE Transactions on information theory **41**, 26 (1995).
- [40] Z. Sethna, Y. Elhanati, C. G. Callan Jr, A. M. Walczak, and T. Mora, Bioinformatics **35**, 2974 (2019).
- [41] K. Yu, J. Shi, D. Lu, and Q. Yang, FEBS Open Bio **9**, 1450 (2019).
- [42] D. S. Shcherbinin, V. A. Belousov, and M. Shugay, PLoS computational biology **16**, e1007714 (2020).
- [43] T. Dupic, Q. Marcou, A. M. Walczak, and T. Mora, PLoS computational biology **15**, e1006874 (2019).
- [44] J. A. Carter, J. B. Preall, K. Grigaityte, S. J. Goldfless, E. Jeffery, A. W. Briggs, F. Vigneault, and G. S. Atwal, Frontiers in immunology **10**, 1516 (2019).
- [45] M.-P. Lefranc, C. Pommie, M. Ruiz, V. Giudicelli, E. Foulquier, L. Truong, V. Thouvenin-Contet, and G. Lefranc, Developmental & Comparative Immunology **27**, 55 (2003).
- [46] J. J. Stewart, C. Y. Lee, S. Ibrahim, P. Watts, M. Shlomchik, M. Weigert, and S. Litwin, Molecular immunology **34**, 1067 (1997).
- [47] A. Chao, Y. Wang, and L. Jost, Methods in Ecology and Evolution **4**, 1091 (2013).
- [48] I. Nemenman, F. Shafee, and W. Bialek, Advances in neural information processing systems **14** (2001).
- [49] I. Springer, N. Tickotsky, and Y. Louzoun, Frontiers in immunology **12**, 664514 (2021).
- [50] P. Meysman, J. Barton, B. Bravi, L. Cohen-Lavi, V. Karnaukhov, E. Lilleskov, A. Montemurro, M. Nielsen, T. Mora, P. Pereira, *et al.*, ImmunoInformatics **9**, 100024 (2023).
- [51] R. Ehrlich, L. Kamga, A. Gil, K. Luzuriaga, L. K. Selin, and D. Ghersi, BMC bioinformatics **22**, 1 (2021).
- [52] K. Mayer-Blackwell, S. Schattgen, L. Cohen-Lavi, J. C. Crawford, A. Souquette, J. A. Gaevvert, T. Hertz, P. G. Thomas, P. Bradley, and A. Fiore-Gartland, Elife **10**, e68605 (2021).
- [53] M. R. Ortega, M. Pogorelyy, A. Minervina, P. Thomas, A. Walczak, and T. Mora, bioRxiv preprint 2024.01.25.577228 (2024).
- [54] H. J. Zahid, R. Taniguchi, P. Ebert, I.-T. Chow, C. Gooley, J. Lv, L. Pisani, M. Rusnak, R. Elyanow, H. Takamatsu, *et al.*, bioRxiv preprint 2024.04.01.587617 (2024).
- [55] S. A. Schattgen, K. Guion, J. C. Crawford, A. Souquette, A. M. Barrio, M. J. Stubbington, P. G. Thomas, and P. Bradley, Nature biotechnology **40**, 54 (2022).
- [56] N. Ceglia, Z. M. Sethna, Y. Elhanati, B. Burman, A. Chow, D. Zamarin, S. DeWolf, S. Shahid, V. Bojilova, N. Rusk, *et al.*, bioRxiv, 2022 (2022).
- [57] J. Textor, F. Buytenhuijs, D. Rogers, È. M. Gauthier, S. Sultan, I. M. Wortel, K. Kalies, A. Fährnich, R. Pagel, H. J. Melichar, *et al.*, Cell Systems **14**, 1059 (2023).
- [58] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, Advances in neural information processing systems **30** (2017).
- [59] C. Sun, F. Baradel, K. Murphy, and C. Schmid, arXiv preprint arXiv:1906.05743 (2019).
- [60] R. Singh, S. Sledzieski, B. Bryson, L. Cowen, and B. Berger, Proceedings of the National Academy of Sciences **120**, e2220778120 (2023).
- [61] T. Gao, X. Yao, and D. Chen, arXiv preprint arXiv:2104.08821 (2021).
- [62] A. Neelakantan, T. Xu, R. Puri, A. Radford, J. M. Han, J. Tworek, Q. Yuan, N. Tezak, J. W. Kim, C. Hallacy, *et al.*, arXiv preprint arXiv:2201.10005 (2022).
- [63] K. Musgrave, S. Belongie, and S.-N. Lim, in *Computer Vision—ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXV 16* (Springer, 2020) pp. 681–699.
- [64] A. Bellet, A. Habrard, and M. Sebban, arXiv preprint arXiv:1306.6709 (2013).
- [65] X. Wang, X. Han, W. Huang, D. Dong, and M. R. Scott, in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (2019) pp. 5022–5030.
- [66] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton, in *International conference on machine learning* (PMLR, 2020) pp. 1597–1607.
- [67] T. Wang and P. Isola, in *International conference on machine learning* (PMLR, 2020) pp. 9929–9939.
- [68] K. C. Garcia, M. Degano, R. L. Stanfield, A. Brunmark, M. R. Jackson, P. A. Peterson, L. Teyton, and I. A. Wilson, Science **274**, 209 (1996).
- [69] J.-B. Reiser, C. Darnault, C. Grégoire, T. Mosser, G. Mazza, A. Kearney, P. A. van der Merwe, J. C. Fontecilla-Camps, D. Housset, and B. Malissen, Nature immunology **4**, 241 (2003).
- [70] Y. He, P. Agnihotri, S. Rangarajan, Y. Chen, M. C. Kerzic, B. Ma, R. Nussinov, R. A. Mariuzza, and J. Orban, Journal of molecular biology **432**, 166697 (2020).
- [71] M. Milighetti, J. Shawe-Taylor, and B. Chain, Frontiers in Physiology **12**, 730908 (2021).
- [72] Y. Nagano and B. Chain, Frontiers in Immunology **14**, 1276106 (2023).
- [73] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, Í. Polat, Y. Feng, E. W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, P. van Mulbregt, and SciPy 1.0 Contributors, Nature Methods **17**, 261 (2020).

Appendix A: Defining conditional coincidence entropy

We follow Refs. [33, 34] and define Renyi conditional entropies as

$$H_\alpha[P(X|Y)] = f^{-1} \left(\sum_y \rho_\alpha(y) f(H_\alpha[P(X|y)]) \right), \quad (\text{A1})$$

where $\rho_\alpha(y)$ represents a generalised weighting given by

$$\rho_\alpha(y) = \frac{P(y)^\alpha}{\sum_y P(y)^\alpha}. \quad (\text{A2})$$

For $\alpha \neq 1$, f can be any invertible function positive in $[0, \infty)$ which is a linear transform of

$$f(x) = 2^{(1-\alpha)x}. \quad (\text{A3})$$

Using this definition of conditional entropy ensures that entropy is additive [33, 34]:

$$H_\alpha[P(X, Y)] = H_\alpha[P(X)] + H_\alpha[P(Y|X)], \quad (\text{A4})$$

where $H_\alpha[P(X, Y)]$ represents the joint Renyi entropy of two random variables defined as

$$H_\alpha[P(X, Y)] = \frac{1}{1-\alpha} \log \left(\sum_{x,y} P(x, y)^\alpha \right). \quad (\text{A5})$$

For simplicity, we will use the following definition for the conditional Renyi entropy of order $\alpha = 2$

$$H_2[P(X|Y)] = -\log \left(\sum_y \rho_2(y) 2^{-H_2[P(X|y)]} \right). \quad (\text{A6})$$

Comparing this expression to the probability of coincidence suggests the following definition for the conditional probability of coincidence averaged over Y

$$p_C[P(X|Y)] = \sum_y \rho_2(y) p_C[P(X|y)], \quad (\text{A7})$$

such that:

$$H_2[P(X|Y)] = -\log p_C[P(X|Y)]. \quad (\text{A8})$$

This definition appears quite natural as it ensures that the conditional probability of coincidence behaves like a regular conditional probability

$$p_C[P(X|Y)] = \frac{p_C[P(X, Y)]}{p_C[P(Y)]}, \quad (\text{A9})$$

and follows Bayes' theorem

$$p_C[P(X|Y)] = \frac{p_C[P(Y|X)] p_C[P(X)]}{p_C[P(Y)]}. \quad (\text{A10})$$

Appendix B: Relations between interaction information and conditional mutual information

Interaction information and conditional mutual information can be related using the additive property of entropy to give

$$I_{2,int}(X, Y|\Pi) = I_2(X, \Pi|Y) - I_2(X, \Pi) \quad (\text{B1})$$

$$= I_2(Y, \Pi|X) - I_2(Y, \Pi). \quad (\text{B2})$$

Therefore, if $I_{2,int}(X, Y|\Pi) > 0$ then X and Y are more relevant when considered in the context of one another than when taken independently. Conversely, for $I_{2,int}(X, Y|\Pi) < 0$ features are partially redundant and becomes less informative in the context of one another. Interaction information may also be expressed in terms of the coincidence mutual information between X and Y to give

$$I_{2,int}(X, Y|\Pi) = I_2(X, Y|\Pi) - I_2(X, Y) \quad (\text{B3})$$

$$= I_2(Y, X|\Pi) - I_2(Y, X). \quad (\text{B4})$$

Features with positive interaction information are those which become more informative about one another in the context of Π . Features with zero interaction information experience no change in their mutual information, while features with negative interaction information have a decrease in their mutual information. Finally, conditional mutual information may be expressed as

$$I_2(X, \Pi|Y) = I_2(X, \Pi) + I_{2,int}(X, Y|\Pi). \quad (\text{B5})$$

So that the information X provides about Π in the context of Y is equal to the information it provides on its own plus a coupling term determined by its interactions with Y . If one feature makes another fully redundant, e.g. $I_2(X, \Pi|Y) = 0$, then the interaction information between them is $I_{2,int}(X, Y|\Pi) = -I_2(X, \Pi)$ such that their coupling completely negates the information provided by X on its own. As explored in [36], definitions of redundancy and synergy may be built from interaction information and conditional mutual information. Interaction information captures a mixture of synergy and redundancy and is positive if synergy outweighs redundancy, negative if redundancy outweighs synergy and zero if they are both equal.

Appendix C: Linking pairwise classification odds to coincidence information

1. Likelihood ratio in terms of coincidence probabilities

We begin with the posterior odds

$$\frac{P(\pi|x = x')}{P(B|x = x')} = \frac{P(x = x'|\pi) P(\pi)}{P(x = x'|B) P(B)}. \quad (\text{C1})$$

Similarly to [24] we may relate the probability distributions associated with π and B using the following expression: $P(x|\pi) = Q^\pi(x)P(x|B)$, where $Q^\pi(x)$ is a selection factor which reweighs the probability of each feature outcome according to some fitness function. To ensure normalisation of $P(X|\pi)$, we require that $\langle Q^\pi(x) \rangle_{P(x|B)} = 1$. $\langle \cdot \rangle_{P(x|B)}$ indicates an expectation value over the distribution $P(X|B)$. $P(x = x'|\pi)$ is the probability of a match in feature X if both sequences were truly drawn from distribution $P(\sigma|\pi)$. We compute this by summing over all possible ways to obtain such a match

$$P(x = x'|\pi) = \sum_{x, x'} \delta_{x, x'} P(x|\pi) P(x'|\pi) = p_C [P(X|\pi)], \quad (\text{C2})$$

where $\delta_{x, x'}$ is the Kronecker delta. $P(x = x'|B)$ is the probability of a match in feature X for a query drawn from $P(\sigma|B)$ and a reference drawn from $P(\sigma|\pi)$. This may be written as

$$P(x = x'|B) = \sum_{x, x'} \delta_{x, x'} P(x|\pi) P(x'|B). \quad (\text{C3})$$

Using the fact that $P(x|\pi) = Q^\pi(x)P(x|B)$ this may be rewritten as

$$P(x = x'|B) = \sum_{x, x'} \delta_{x, x'} Q^\pi(x) P(x|B) P(x'|B), \quad (\text{C4})$$

which may also be written as

$$\begin{aligned} P(x = x'|B) &= \left\langle \sum_{x'} \delta_{x, x'} Q^\pi(x) P(x'|B) \right\rangle_{P(x|B)} \\ &= \langle Q^\pi(x) P(x|B) \rangle_{P(x|B)}. \end{aligned} \quad (\text{C5})$$

Assuming that the background probability of a particular feature is independent of its selection factor, we may decompose the expectation value to arrive at

$$\begin{aligned} P(x = x'|B) &= \langle Q^\pi(x) \rangle_{P(x|B)} \langle P(x|B) \rangle_{P(x|B)} \\ &= p_C[P(X|B)] \end{aligned} \quad (\text{C6})$$

As $\langle Q^\pi(x) \rangle_{P(x|B)} = 1$ by normalisation. We will write $p_C[P(X|B)] = p_C[P(X)]$ as it is the probability of coincidence over a background distribution.

2. Average classification odds over subsets of specific TCRs

The odds ratio for a particular epitope subset given a match in feature X is:

$$\frac{P(\pi|x = x')}{P(B|x = x')} = \frac{p_C[P(X|\pi)]}{p_C[P(X)]} \frac{P(\pi)}{P(B)} \quad (\text{C7})$$

Taking the average of this over the normalised distribution of epitope subsets $P(\Pi)/(1 - P(B))$ yields

$$\sum_{\pi} \frac{P(\pi|x = x')}{P(B|x = x')} \frac{P(\pi)}{1 - P(B)} = \sum_{\pi} \frac{p_C[P(X|\pi)]}{p_C[P(X)]} \frac{P(\pi)}{P(B)} \frac{P(\pi)}{1 - P(B)}. \quad (\text{C8})$$

Recalling the definition of the conditional probability of coincidence, Eq. 6, the right hand side becomes

$$\sum_{\pi} \frac{p_C[P(X|\pi)]}{p_C[P(X)]} \frac{P(\pi)}{P(B)} \frac{P(\pi)}{1 - P(B)} = \frac{p_C[P(X|\Pi)]}{p_C[P(X)]} \frac{\sum_{\pi'} P(\pi')^2}{P(B)(1 - P(B))}, \quad (\text{C9})$$

we therefore obtain in full:

$$\sum_{\pi} \frac{P(\pi|x = x')}{P(B|x = x')} \frac{P(\pi)}{1 - P(B)} = \frac{p_C[P(X|\Pi)]}{p_C[P(X)]} \frac{\sum_{\pi'} P(\pi')^2}{P(B)(1 - P(B))}, \quad (\text{C10})$$

which may be written as:

$$\left\langle \frac{P(\pi|x = x')}{P(B|x = x')} \right\rangle = \frac{p_C[P(X|\Pi)]}{p_C[P(X)]} \left\langle \frac{P(\pi)}{P(B)} \right\rangle. \quad (\text{C11})$$

Where $\langle \rangle$ denotes an average over $\frac{P(\pi)}{1 - P(B)}$.

3. Relating the probability of specificity and the probability of being drawn from a specific subset

We show here how $P(\pi)$ and $P(\pi|x = x')$, which represent the prior and posterior probabilities that a sequence is observed because it was drawn from distribution $P(\sigma|\pi)$ may be related to the prior and posterior probabilities that a sequence is in-fact specific to epitope π . We consider the mixture distribution of sequences

$$P(\sigma) = P(\pi)P(\sigma|\pi) + P(B)P(\sigma|B). \quad (\text{C12})$$

We denote the fraction of sequences in this mixture which would bind to epitope π in a test of specificity as $P(S^\pi)$. We may express this fraction in terms of the mixture proportions

$$P(S^\pi) = P(\pi)P(S^\pi|\pi) + P(B)P(S^\pi|B), \quad (\text{C13})$$

where $P(S^\pi|\pi)$ and $P(S^\pi|B)$ are the fractions of sequences produced by distributions $P(\sigma|\pi)$ and $P(\sigma|B)$ which are specific to π respectively. As $P(\sigma|\pi)$ represents the distribution of sequences specific to π , by definition $P(S^\pi|\pi) = 1$, so

$$P(S^\pi) = P(\pi) + P(B)P(S^\pi|B). \quad (\text{C14})$$

As $P(\pi) + P(B) = 1$, this may be written as

$$P(S^\pi) = P(\pi) + (1 - P(\pi))P(S^\pi|B). \quad (\text{C15})$$

If we assume that the fraction of sequences generated by distribution $P(\sigma|B)$ which are specific to π is much smaller than the proportion $P(\pi)$ in our mixture so $P(S^\pi|B) \ll P(\pi)$, then we may simplify this directly down to:

$$P(S^\pi) \approx P(\pi) \quad (\text{C16})$$

Now substituting $P(S^\pi)$ into our prior odds in Eq. 14, we obtain

$$\frac{P(\pi|x=x')}{P(B|x=x')} = \frac{p_C[P(X|\pi)]}{p_C[P(X)]} \frac{P(S^\pi)}{1 - P(S^\pi)}. \quad (\text{C17})$$

Following similar reasoning, we may write the posterior probability that a sequence sampled from the mixture is specific to epitope π given that it matches in feature X with sequence with known specificity to π as

$$P(S^\pi|x=x') = P(\pi|x=x')P(S^\pi|\pi, x=x') + P(B|x=x')P(S^\pi|B, x=x'). \quad (\text{C18})$$

Once again, $P(S^\pi|\pi, x=x') = 1$ and $P(S^\pi|B, x=x') \leq 1$, therefore

$$P(S^\pi|x=x') \approx P(\pi|x=x') \quad (\text{C19})$$

as long as the posterior odds exceeds one. Taken together, this allows us to re-express Eq. 14 in terms of prior and posterior odds of specificity to π

$$\frac{P(S^\pi|x=x')}{1 - P(S^\pi|x=x')} = \frac{p_C[P(X|\pi)]}{p_C[P(X)]} \frac{P(S^\pi)}{1 - P(S^\pi)}. \quad (\text{C20})$$

4. Simulating pairwise classification with limited information.

a. Generating background and specific TCRs

Here, we followed a previously described procedure for the in silico simulation of epitope-specific repertoires [24]. In short, we simulated a synthetic background repertoire by randomly generating strings of length k from a predefined alphabet of q characters. We defined ‘epitope specific’ TCRs by a set of hard coded rules: For each epitope we generated M motifs, where each motif contained c out of q letters for each of the TCR’s k ‘residues’. To generate the specific TCRs for a motif, we produced all possible combinations of these amino acids. The set of TCRs specific to a given epitope was the full set of TCRs produced by all motifs associated with the epitope.

b. Comparing theoretical performance to true classification accuracy

We computed the relevancy of various ‘features’ of simulated sequences for $q = 20$, $k = 4$, $M = 5$ and $c = 3$. To define features we divided each TCR by its residues. For example, to produce an α and a β chain we divided each TCR into two halves. We then produced a mixture data set into which we mixed a particular fraction of simulated background and specific TCRs. For each TCR we also assigned a label identifying which of these two original sets it came from. We then simulated pairwise classification using feature matching by sampling sequences from the mixture set and looking for matches in a particular feature with a TCR from the true set of specific sequences. Using the labels, we were able to calculate the probability that the sampled sequence was truly specific given that this match occurred. We performed this simulation using a range of features and mixture ratios as shown in Figure S1 and simulation results aligned with the theoretical predictions from coincidence information.

5. Classification odds given a fuzzy match

For both query and reference being drawn from $P(\sigma|\pi)$, the probability of a match is

$$P(d(x, x') = \Delta|\pi) = \sum_{x, x'} \delta_{d(x, x'), \Delta} P(x|\pi) P(x'|\pi) = p_C[P(X|\pi)](\Delta). \quad (\text{C21})$$

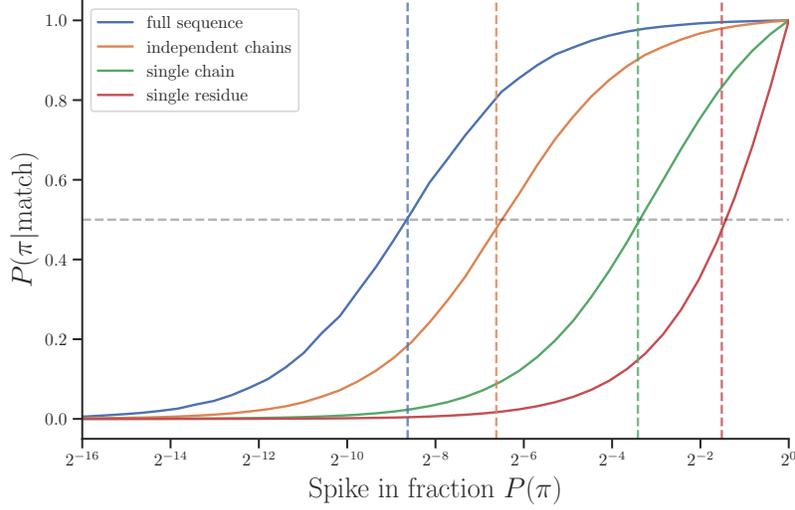


FIG. S1. **Simulated prior vs posterior probabilities given feature match** Prior probability of a sequence being present in a mixture distribution due to being sampled from a spiked in epitope specific distribution, vs posterior probability given a feature match with a specific sequence. Dashed lines indicate theoretical prior probability values for which a posterior probability of 0.5 should be observed given the informational value of a given feature computed from Eq. 17. Solid curves cover classification for a given fraction of spiked in specific sequence in the following scenarios: using matches in the full simulated TCR (single chain), using matches in one half of the TCR (single chain), using matches from a single amino acid (single residue), using matches in both halves of the TCR but considered independently (independent chains).

Now consider the case where the query is truly from $P(\sigma|B)$ while the reference is from $P(\sigma|\pi)$

$$P(d(x, x') = \Delta|B) = \sum_{x, x'} \delta_{d(x, x'), \Delta} P(x|B) P(x'|\pi). \quad (\text{C22})$$

Similarly to the case for exact matching, we may write $P(x'|\pi) = Q^\pi(x')P(x'|B)$ such that

$$P(d(x, x') = \Delta|B) = \sum_{x, x'} \delta_{d(x, x'), \Delta} P(x|B) Q^\pi(x') P(x'|B). \quad (\text{C23})$$

Next, we may rewrite this equation as:

$$P(d(x, x') = \Delta|B) = \left\langle \sum_x \delta_{d(x, x'), \Delta} P(x|B) Q^\pi(x') \right\rangle_{P(x'|B)} \quad (\text{C24})$$

$$= \left\langle Q^\pi(x') \sum_x \delta_{d(x, x'), \Delta} P(x|B) \right\rangle_{P(x'|B)} \quad (\text{C25})$$

We define $n_\Delta(x') = \sum_x \delta_{d(x, x'), \Delta} P(x|B)$ to be the neighbourhood density around x' at distance Δ . Assuming independence of neighbor density $n_\Delta(x')$ and selection factor $Q^\pi(x')$ we have

$$P(d(x, x') = \Delta|B) = \langle Q^\pi(x') \rangle_{P(x'|B)} \langle n_\Delta(x') \rangle_{P(x'|B)}. \quad (\text{C26})$$

By normalization the first term evaluates to 1 so we are left with

$$P(d(x, x') = \Delta|B) = \langle n_\Delta(x') \rangle_{P(x'|B)} = \sum_{x, x'} \delta_{d(x, x'), \Delta} P(x|B) P(x'|B) = p_C[P(X|B)](\Delta), \quad (\text{C27})$$

therefore the likelihood ratio becomes

$$\frac{P(d(x, x') = \Delta|\pi)}{P(d(x, x') = \Delta|B)} = \frac{p_C[P(X|\pi)](\Delta)}{p_C[P(X)](\Delta)}. \quad (\text{C28})$$

Where once again we have used $p_C[P(X)](\Delta) = p_C[P(X|B)](\Delta)$ to denote the background probability of coincidence.

Appendix D: Feature interactions arising from data structure

1. Mixture of motifs model

In the following we explore in a minimal model how local feature relevancy and interaction information depend on the number of distinct binding solutions able to engage a particular epitope. We consider that for a given epitope π there is a distribution of motifs/binding modes, $P(\Phi)$. For a given binding mode, $\Phi = \phi$, there is an associated set of specific sequences and sequence features, $P(X|\phi, \pi)$. The probability of drawing a particular feature $X = x$ from the distribution of sequences specific to π is

$$P(x|\pi) = \sum_{\phi} P(x|\phi, \pi)P(\phi), \quad (\text{D1})$$

The coincidence probability for sequences drawn from $P(X|\pi)$ is

$$p_C[P(X|\pi)] = \sum_x \left(\sum_{\phi} P(x|\phi, \pi)P(\phi) \right)^2 \quad (\text{D2})$$

In [24] we show that such a mixture distribution may be expressed as the sum of a within-motif and a cross-motif term

$$p_C[P(X|\pi)] = p_C[P(\phi)] \langle p_C[P(X|\pi, \phi)] \rangle_{P(\phi|\phi_1=\phi_2=\phi)} + (1 - p_C[P(\phi)]) \langle p_C[P(X|\pi, \phi_1), P(X|\pi, \phi_2)] \rangle_{P(\phi_1, \phi_2|\phi_1 \neq \phi_2)}, \quad (\text{D3})$$

where $\langle p_C[P(X|\pi, \phi)] \rangle_{P(\phi|\phi_1=\phi_2=\phi)} = p_C[P(X|\pi, \Phi)]$ is the average probability of feature coincidence within each motif subset for motifs associated with epitope π weighted by

$$P(\phi|\phi_1 = \phi_2 = \phi) = \frac{P(\phi)^2}{\sum_{\phi'} P(\phi')^2} = \rho_2(\phi). \quad (\text{D4})$$

And $\langle p_C[P(X|\pi, \phi_1), P(X|\pi, \phi_2)] \rangle_{P(\phi_1, \phi_2|\phi_1 \neq \phi_2)}$ is the cross-motif coincidence probability with the probability of coincidence for two distributions defined in general as:

$$p_C[P_i(X), P_j(X)] = \sum_x P_i(x)P_j(x), \quad (\text{D5})$$

with the average computed with respect to the following weighting factor

$$P(\phi_1, \phi_2|\phi_1 \neq \phi_2) = \frac{P(\phi_1)P(\phi_2)}{1 - \sum_{\phi'} P(\phi')^2}. \quad (\text{D6})$$

We will assume that in general this cross-motif coincidence probability is far smaller than the within coincidence probability such that

$$p_C[P(X|\pi)] \approx p_C[P(\phi)]p_C[P(X|\pi, \Phi)]. \quad (\text{D7})$$

To proceed, we will now assume that $p_C[P(\phi)] = 1/M$, where M is number of motifs associated with a given epitope π . We will also assume that for a given M there is a characteristic probability of coincidence $p_C[P(X|\Pi_M)]$ where Π_M denotes the fact that we are considering epitopes with M motifs. We use uppercase Π to highlight that we are theorising that epitopes with equal number of motifs only vary in their coincidence probabilities for particular features due to finite sample size error and hence their true coincidence probabilities should be equal in the limit of infinite data. We will assume that $p_C[P(X|\pi, \Phi)] = p_C[P(X|\Pi_1)]$ such that it is the probability of feature coincidence for epitopes with a single motif. Therefore, for an epitope with M motifs we obtain

$$p_C[P(X|\Pi_M)] \approx \frac{1}{M} p_C[P(X|\Pi_1)]. \quad (\text{D8})$$

The local relevancy of X for an epitope with M motifs is then

$$\begin{aligned} i_2(X, \Pi_M) &= \log \left(\frac{1}{M} \frac{p_C[P(X|\Pi_1)]}{p_C[P(X)]} \right) \\ &= \log \left(\frac{p_C[P(X|\Pi_1)]}{p_C[P(X)]} \right) - \log M \\ &= i_2(X, \Pi_1) - \log M, \end{aligned} \quad (\text{D9})$$

where $i_2(X, \Pi_1)$ is the characteristic local relevancy of feature X for epitopes with a single motif. Therefore, the greater the number of motifs associated with a given epitope, the less locally relevant each feature is expected to become. With such a model we would expect the local relevancy of two features, X and Y , for a given number of motifs to be related by the following equation

$$i_2(X, \Pi_1) - i_2(X, \Pi_M) = i_2(Y, \Pi_1) - i_2(Y, \Pi_M). \quad (\text{D10})$$

Such that

$$i_2(X, \Pi_M) = i_2(Y, \Pi_M) + i_2(X, \Pi_1) - i_2(Y, \Pi_1). \quad (\text{D11})$$

Which is a relationship we observe in Figure 5. Now let us consider how we expect the local interaction information between two features to change with number of motifs

$$i_{2,int}(X, Y|\Pi_M) = i_2([X, Y], \Pi_M) - i_2(X, \Pi_M) - i_2(Y, \Pi_M) \quad (\text{D12})$$

$$= i_2([X, Y], \Pi_1) - i_2(X, \Pi_1) - i_2(Y, \Pi_1) - \log M + 2 \log M \quad (\text{D13})$$

$$= i_{2,int}(X, Y|\Pi_1) + \log M. \quad (\text{D14})$$

We expect the local interaction information between pairs of features to increase with the number of motifs associated to the particular epitope. Therefore, we expect the interaction information between two features and their independent local relevancy for a given number of motifs to be related by the following expressions

$$i_{2,int}(X, Y|\Pi_M) - i_{2,int}(X, Y|\Pi_1) = i_2(X, \Pi_1) - i_2(X, \Pi_M) \quad (\text{D15})$$

$$= i_2(Y, \Pi_1) - i_2(Y, \Pi_M). \quad (\text{D16})$$

So that

$$i_{2,int}(X, Y|\Pi_M) = -i_2(X, \Pi_M) + i_2(X, \Pi_1) + i_{2,int}(X, Y|\Pi_1) \quad (\text{D17})$$

$$= -i_2(Y, \Pi_M) + i_2(Y, \Pi_1) + i_{2,int}(X, Y|\Pi_1). \quad (\text{D18})$$

Which is a relationship observed in Figure 5.

2. False positives

We will briefly show that false positives, that is sequences which are not truly specific to a given epitope but yet appear in a specific TCR sample, lead to similar results to the mixture of motif models. Consider that a sample of sequences is taken in an experiment designed to obtain sequences specific to epitope π and that a distribution of a particular feature X is obtained, $P(X|\tilde{\pi})$. A fraction of features in this distribution are truly specific to π , $P(\pi)$, while a fraction of sequences are false positives, $P(\neg\pi)$. We will denote the subset of sequence features truly specific to epitope π as $P(X|\pi)$ and the subset of sequence feature not specific to π as $P(X|\neg\pi)$. The probability of coincidence for the distribution $P(X|\tilde{\pi})$ is

$$p_C[P(X|\tilde{\pi})] = \sum_x (P(X|\pi)P(\pi) + P(X|\neg\pi)P(\neg\pi))^2. \quad (\text{D19})$$

Expanding out we get

$$p_C[P(X|\tilde{\pi})] = \sum_x P(X|\pi)^2 P(\pi)^2 + \sum_x P(X|\neg\pi)^2 P(\neg\pi)^2 + 2 \sum_x P(X|\pi)P(X|\neg\pi)P(\pi)P(\neg\pi) \quad (\text{D20})$$

$$= P(\pi)^2 p_C[P(X|\pi)] + P(\neg\pi)^2 p_C[P(X|\neg\pi)] + 2P(\pi)P(\neg\pi)p_C[P(X|\pi), P(X|\neg\pi)]. \quad (\text{D21})$$

We will assume that cross coincidences are rare, that coincidences are far more likely between the truly specific sequences and that the fraction of false positives is far smaller than the fraction of true positives such that the following expressions hold: $P(\pi)^2 p_C[P(X|\pi)] \gg 2P(\pi)P(\neg\pi)p_C[P(X|\pi), P(X|\neg\pi)]$ and $P(\pi)p_C[P(X|\pi)] \gg P(\neg\pi)p_C[P(X|\neg\pi)]$. Therefore we obtain

$$p_C[P(X|\tilde{\pi})] \approx P(\pi)^2 p_C[P(X|\pi)].$$

Now considering the local relevancy of X computed from $p_C[P(X|\tilde{\pi})]$

$$\begin{aligned}
 i_2(X, \tilde{\pi}) &= \log \left(\frac{P(\pi)^2 p_C[P(X|\pi)]}{p_C[P(X)]} \right) \\
 &= \left(\frac{p_C[P(X|\pi)]}{p_C[P(X)]} \right) + 2 \log P(\pi) \\
 &= i_2(X, \pi) + 2 \log P(\pi).
 \end{aligned} \tag{D22}$$

If $P(\pi) < 1$, then $\log P(\pi) < 0$ so the relevancy of X will be lower when computed using distribution $P(X|\tilde{\pi})$ than when using $P(X|\pi)$. Similarly to D 1, we identify that when $P(\pi) < 1$ the local interaction information between two features X and Y will be increased

$$\begin{aligned}
 i_{2,int}(X, Y|\tilde{\pi}) &= i_2([X, Y], \tilde{\pi}) - i_2(X, \tilde{\pi}) - i_2(Y, \tilde{\pi}) \\
 &= i_2([X, Y], \pi) - i_2(X, \pi) - i_2(Y, \pi) + 2 \log P(\pi) - 2 \log P(\pi) - 2 \log P(\pi) \\
 &= i_{2,int}(X, Y|\pi) - 2 \log P(\pi).
 \end{aligned} \tag{D23}$$

Therefore the relationships observed in Figure 5 could also be produced by false positives. Analyses of how TCR clustering depends on background sequence space coverage could provide a potential avenue for distinguishing between both models in future work.

Appendix E: Data analysis methods

1. Data collection and pre-processing

TCR sequencing data from Minervina *et al.* [22] and Dash *et al.* [1] was processed as follows: Data was deduplicated on the full nucleotide sequence level for each epitope in order to enable assessment of clonal convergence independently of clonal expansions within individual donors [24]. Only TCRs with full CDR3, V gene and J gene annotations for both the α and β chain were retained, and V and J gene names were standardized using tidytcells [72]. Furthermore, epitope specific sets were only retained in cases where at least one full TCR coincidence was observed. Background TCR sequence data was produced using the OLGA generation model [40]. We generated 1000000 α and β chain sequences and then randomly paired these to produce paired chain background data. The final number of sequences retained is detailed in Table I.

Epitope ID	Data set	Epitope	HLA	Sequence counts
	OLGA	N/A	N/A	1000000
1	Minervina	NQKLIANQF	HLA-B*15:01	148
2	Minervina	DTDFVNEFY	HLA-A*01:01	88
3	Minervina	LLYDANYFL	HLA-A*02:01	53
4	Minervina	PTDNYITTY	HLA-A*01:01	155
5	Minervina	YLQPRTFLL	HLA-A*02:01	288
6	Minervina	FTSDYYQLY	HLA-A*01:01	450
7	Minervina	ALSKGVHFV	HLA-A*02:01	197
8	Minervina	LTDEMIAQY	HLA-A*01:01	398
9	Minervina	TTDPSFLGRY	HLA-A*01:01	1909
10	Dash	NLVPMVATV	HLA-A*02:01	67
11	Dash	GLCTLVAML	HLA-A*02:01	92
12	Dash	GILGFVFTL	HLA-A*02:01	249

TABLE I. TCR sequence data used in this study. The first column shows the epitope ID used throughout figures.

2. Statistical analysis

Coincidence probabilities and their variance were computed using unbiased estimators described in [27]. Conditional probabilities of coincidence were averaged over epitopes using Eq. 6 with an equal weighting assigned to each epitope specific subset. Background entropies of the paired chain TCRs were computed as the sum of α and β chain entropies, exploiting the independent pairing of chains in the computationally constructed background. Mutual information scores and higher order statistics such as synergy were then computed from these entropies using Eqs. 9, 18 and 11. Orthogonal distance regression was performed as implemented in Scipy [73].

Name	Version	Link
Tidytcells [72]	2.00	https://pypi.org/project/tidytcells/
Pyrepseq [24]	1.2.1	https://pypi.org/project/pyrepseq/
OLGA [40]	1.2.4	https://github.com/statbiophys/OLGA
Scipy [73]	1.11.3	https://pypi.org/project/scipy/

TABLE II. Software packages used in this study.

SUPPLEMENTARY FIGURES

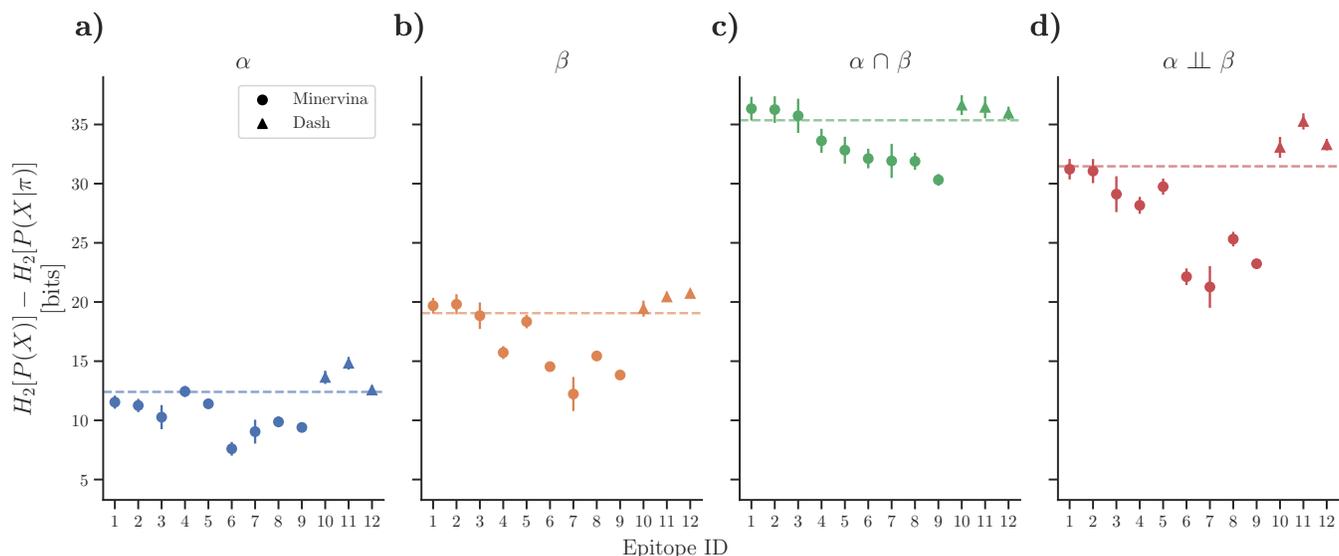


FIG. S2. **Coincidence entropy reduction on the per-epitope level.** Coincidence entropy reduction between specific and background TCRs for the **a)** α and **b)** β chains. **c)** Local relevancy of the full paired TCR chain. $X \cap Y$ represents joint consideration of features X and Y here and in the following. **d)** Sum of local relevancies of the α and β chain. $X \perp\!\!\!\perp Y$ represents separate consideration of features X and Y here and in the following. Epitopes are ordered within each dataset by their full entropy change. Dashed lines indicate the average entropy change shown in Figure 2. Although a large amount of variability between various epitopes not captured by their associated uncertainties is observed, every epitope shows a decrease in both α and β chain entropy. For every epitope, the α and β chain are therefore informative features. The change in entropy of the full TCR sequence $\alpha\beta$ is in all cases greater than the sum of the entropy changes of the α and β chains taken independently, suggesting that synergy between α and β chain information is a general phenomenon.

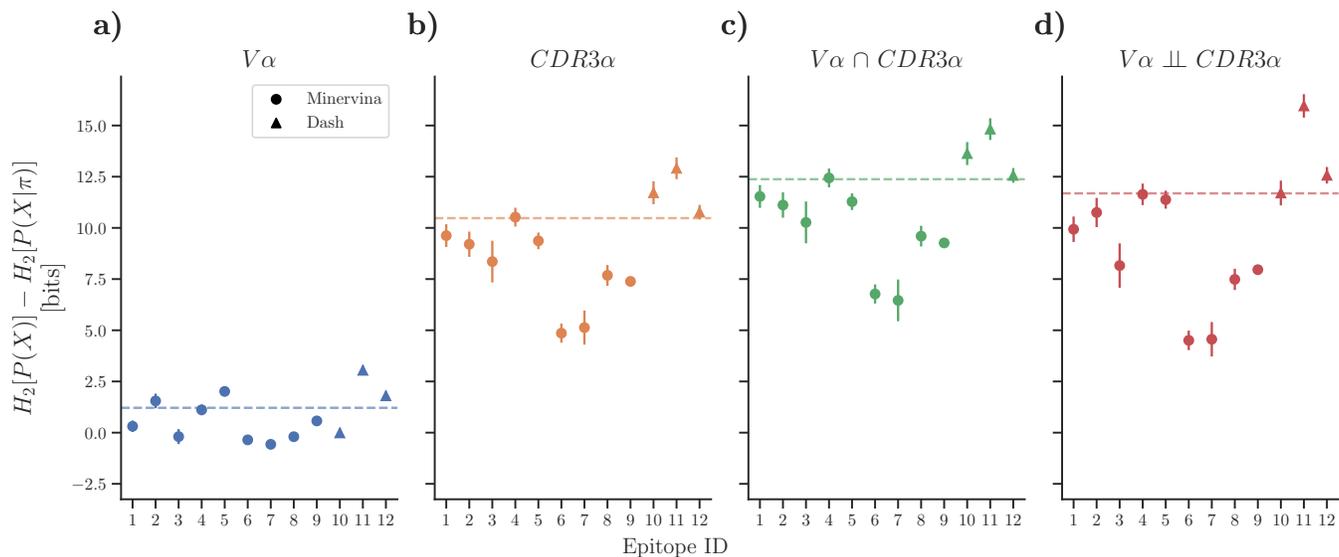


FIG. S3. **Coincidence entropy reduction for $CDR3\alpha$ and $V\alpha$.** Both the $V\alpha$ and $CDR3\alpha$ contributions to specificity vary across epitopes, with some epitopes showing no detectable restriction of $V\alpha$ gene choice. Further details as described for Fig. S2.

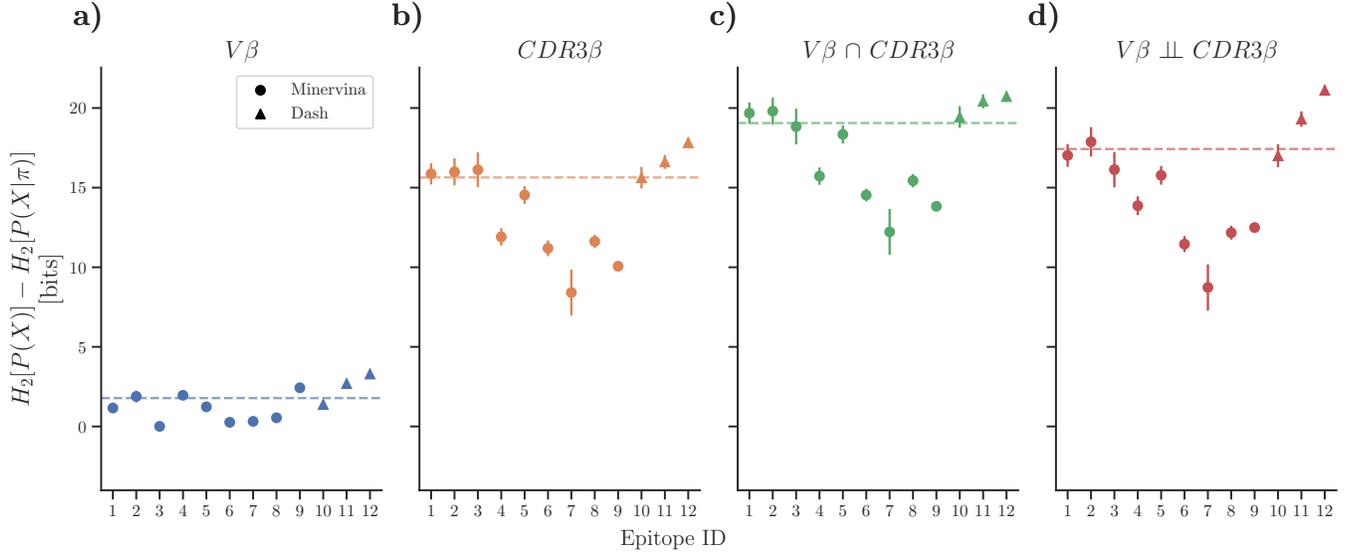


FIG. S4. **Coincidence entropy reduction for CDR3 β and V β .** Both the V β and CDR3 β contributions to specificity vary across epitopes. Most epitopes show small, but positive restriction of V β gene choice. Further details as described for Fig. S2.

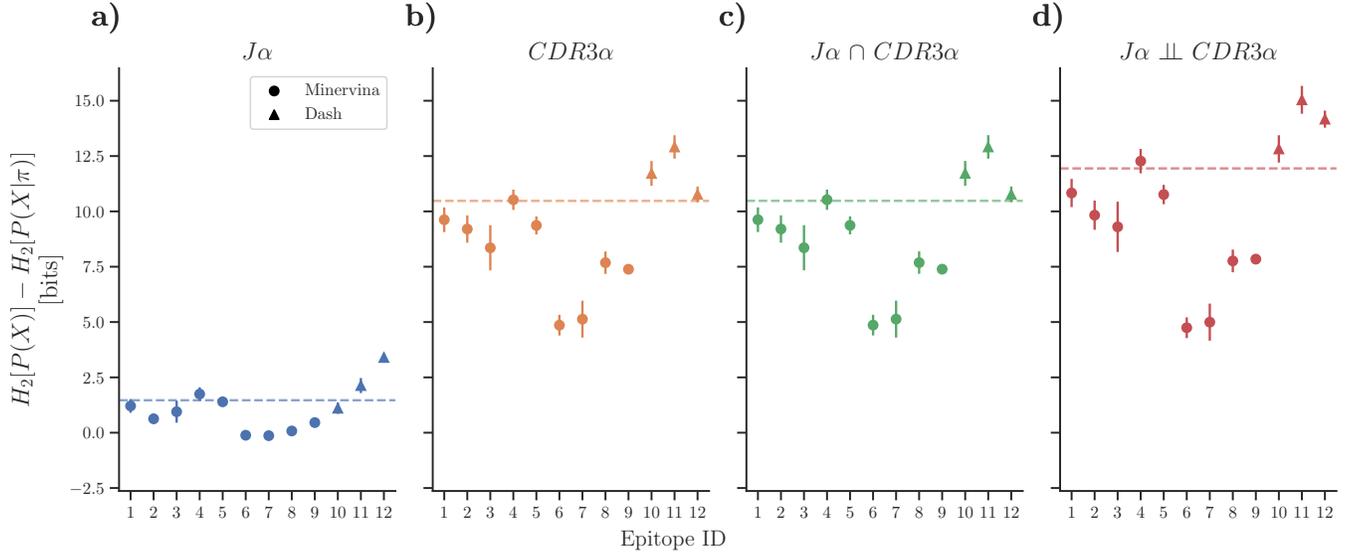


FIG. S5. **Coincidence entropy reduction for CDR3 α and J α .** Comparison of panel c and d reveals negative interaction information between J α and CDR3 α regions suggesting that the CDR3 α makes the J α redundant. Further details as described for Fig. S2.

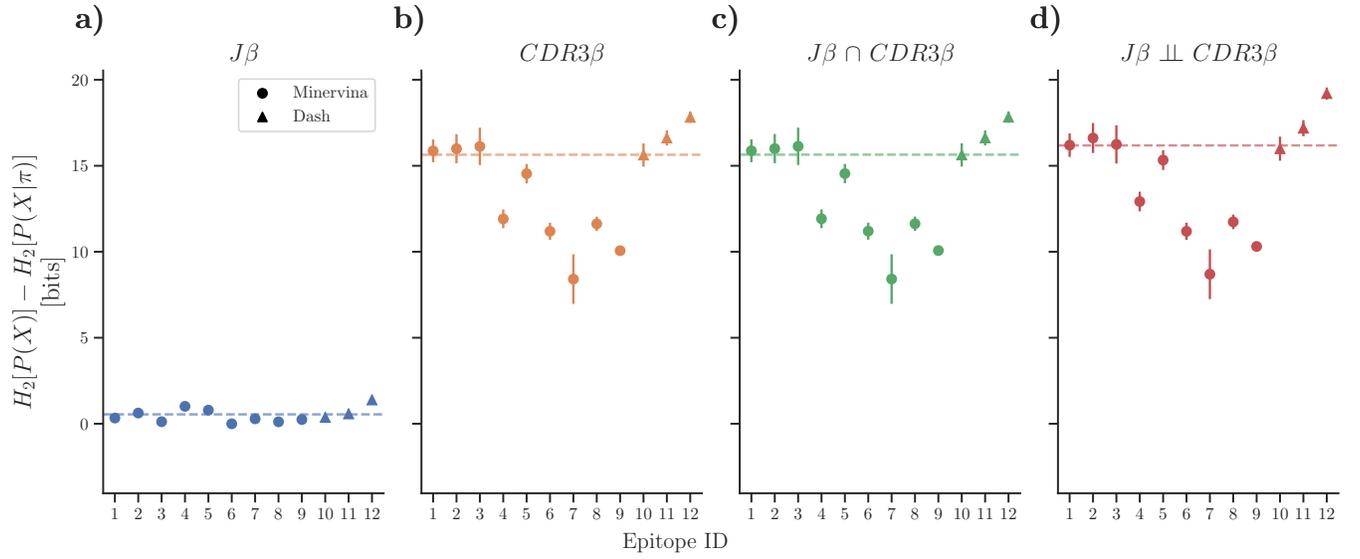


FIG. S6. **Coincidence entropy reduction for CDR3 β and J β .** Comparison of panel c and d reveals negative interaction information between J β and CDR3 β regions suggesting that the CDR3 β makes the J β redundant. Further details as described for Fig. S2.

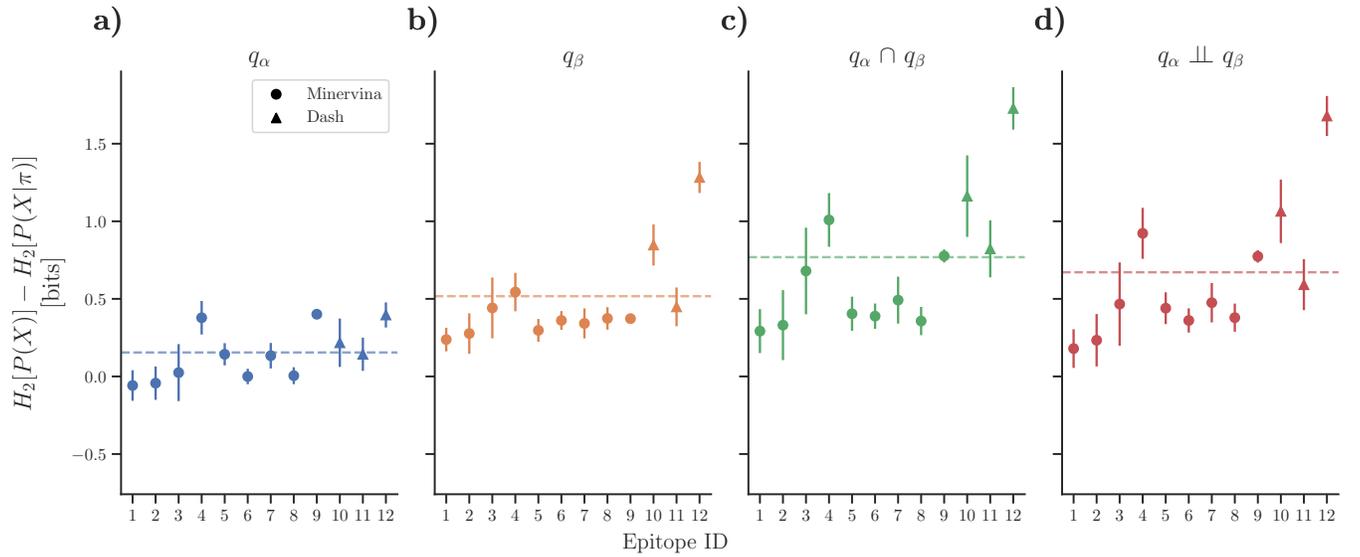


FIG. S7. **Coincidence entropy reduction for CDR3 net charge across epitopes.** While α chain charge has variable relevancy and in some cases is not informative, β chain charge is consistently informative across epitopes. Further details as described for Fig. S2.

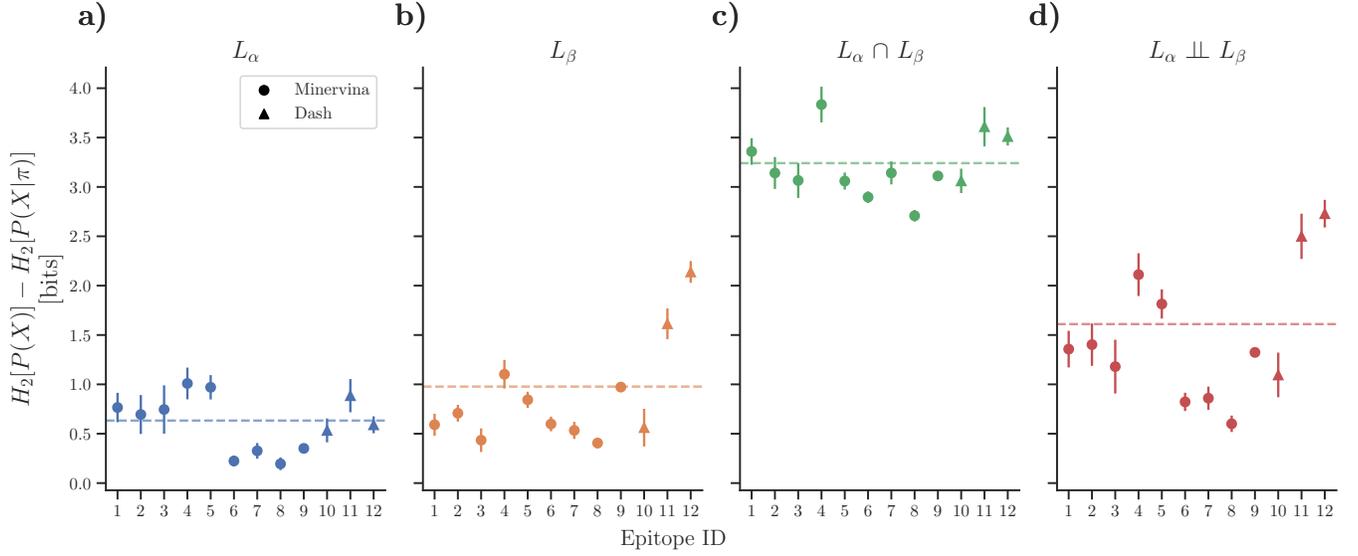


FIG. S8. **Coincidence entropy reduction for CDR3 length across epitopes.** The CDR3 length of both the α and β chains are relevant features for all epitopes. There is substantial α and β length pairing as all epitopes display positive interaction information (synergy). Further details as described for Fig. S2.

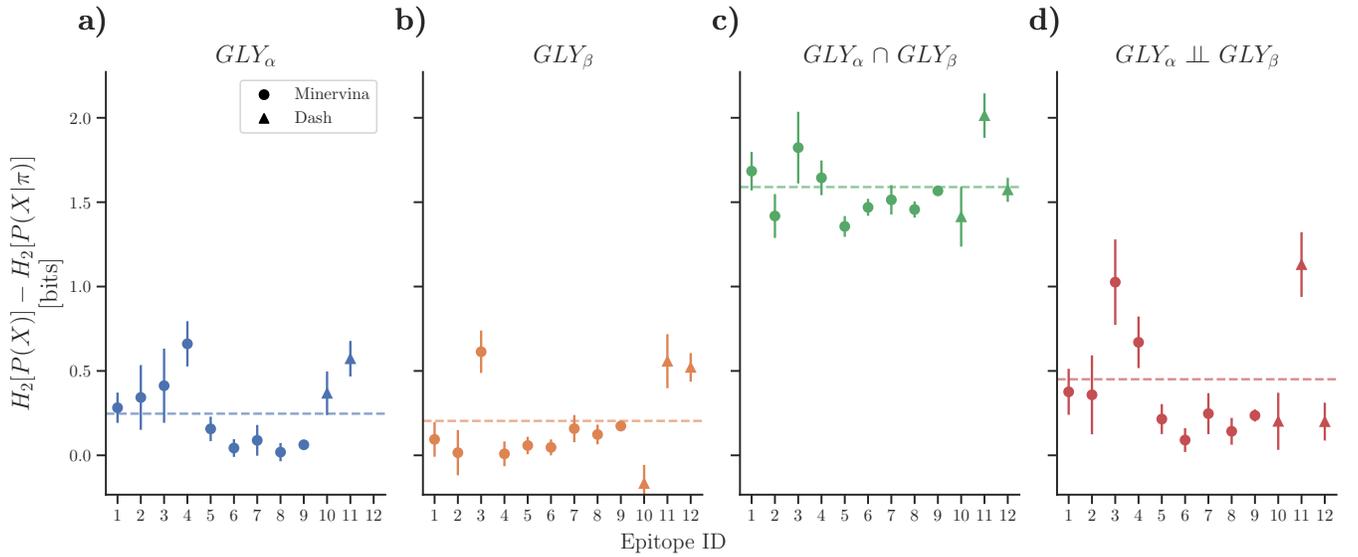


FIG. S9. **Coincidence entropy reduction for CDR3 glycine content across epitopes.** CDR3 glycine content, defined as the number of glycine residues within the CDR3, is only weakly relevant when considered on the single chain level, but there is substantial synergy. Further details as described for Fig. S2.