

Molecular mimicry as a mechanism of viral immune evasion and autoimmunity

Received: 21 March 2024

Accepted: 18 October 2024

Published online: 30 October 2024

Cole Maguire¹, Chumeng Wang¹, Akshara Ramasamy¹, Cara Fonken¹,
Brinkley Morse¹, Nathan Lopez¹, Dennis Wylie² & Esther Melamed¹✉

Mimicry of host protein structures, or ‘molecular mimicry’, is a common mechanism employed by viruses to evade the host’s immune system. Short linear amino acid (AA) molecular mimics can elicit cross-reactive antibodies and T cells from the host, but the prevalence of such mimics throughout the human virome has not been fully explored. Here we evaluate 134 human-infecting viruses and find significant usage of linear mimicry across the virome, particularly those in the *Herpesviridae* and *Poxviridae* families. Furthermore, host proteins related to cellular replication and inflammation, autosomes, the X chromosome, and thymic cells are enriched as viral mimicry targets. Finally, we find that short linear mimicry from Epstein-Barr virus (EBV) is higher in auto-antibodies found in patients with multiple sclerosis than previously appreciated. Our results thus hint that human-infecting viruses leverage mimicry in the course of their infection, and that such mimicry may contribute to autoimmunity, thereby prompting potential targets for therapies.

Over the centuries, viruses have co-evolved with their hosts, particularly to gain traits that increase their virulence such as faster replication times, longer infectious periods, and strategies to evade the host immune system¹. One notable trait is molecular mimicry, in which pathogens “mimic” protein structures of their host². Due to mechanisms generally preventing formation of auto-reactive immune responses^{3,4}, mimicked proteins are thought to be protective to pathogens by limiting the number of pathogenic epitopes able to be targeted by the host, thereby impeding the host’s immune response². However, despite the biological safeguards against autoimmunity, mimicry can elicit T cell and antibody responses that are cross-reactive to both the mimicking and mimicked proteins, with this cross-reactivity thought to possibly underlie autoimmune pathologies^{5–8}.

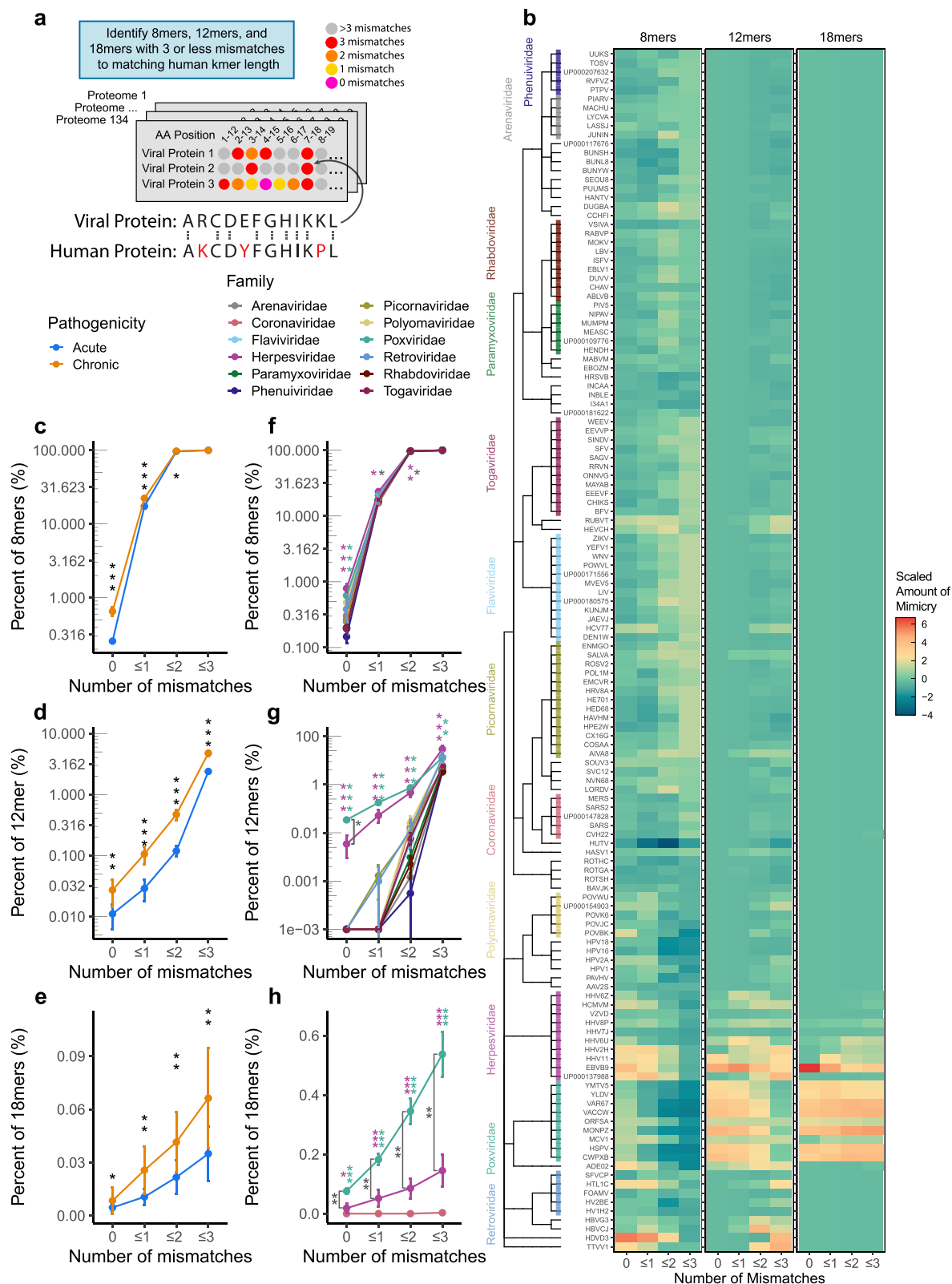
Previous work has been done to quantify mimicry across a wide range of viral proteomes with an emphasis on determining mimicry at the tertiary and quaternary protein structure level⁹. However, adaptive immune responses generally target small regions of proteins (called epitopes), with T cells primarily responding to linear protein epitopes of 8–12 or 18–24 amino acids (AAs) long^{10–13}, and up to 50% of antibodies binding linear epitopes of 4–12 AAs^{14,15}. Thus, mimicry at the level of primary protein structure (i.e. linear sequence) needs to be

evaluated to better understand how viral mimicry may be perceived by the adaptive immune system. While prior studies have evaluated linear AA mimicry in limited cohorts of pathogens^{16–18} or searched specifically for top mimics that may contribute to auto-immunity¹⁹, it remains unclear how the abundance of linear mimicry differs across a wider range of the human virome.

Interestingly, the “molecular mimicry trade-off hypothesis” posits that mimicry may not always confer a net advantage to the virus²⁰. Thus, to adopt mimicry, the advantages gained by the virus must outweigh any associated drawbacks, such as extended replication times due to longer protein sequences or compromised protein functionality due to mutations introduced to achieve mimicry²⁰. However, short linear mimicry at the size of an immune epitope may be able to offer substantial evasion of the adaptive immune system while minimizing detrimental effects to either viral protein function or length. As a result, short linear mimicry may possibly offer one optimization to the molecular mimicry trade-off and be advantageous to viruses.

In this study, we conduct a comprehensive screen of 134 human-infecting viral proteomes to quantify the prevalence of short linear molecular mimicry across the human virome. Our analysis reveals distinct patterns of mimicry across viral families and identifies specific

¹Department of Neurology, Dell Medical School, The University of Texas at Austin, Austin, TX, USA. ²Center for Biomedical Research Support, The University of Texas at Austin, Austin, TX, USA. ✉e-mail: esther.melamed@austin.utexas.edu



host cellular processes, thymic cell processes and genomic regions that are preferentially targeted by molecular mimicry. We also explore the potential connection between viral mimicry and autoimmune disorders, using multiple sclerosis as a case study. Our findings shed light on the intricate dynamics of host-pathogen interactions and viral adaptation mechanisms. These findings reveal that viruses employ diverse evolutionary strategies in their interactions with hosts, and

that such adaptations may influence autoimmune processes, thus suggesting new directions for therapeutic approaches.

Results

Elevated mimicry in poxviridae and herpesviridae

To systematically investigate the rate of molecular mimicry in human infecting viruses, we queried 134 viral proteomes for the matching

Fig. 1 | Molecular mimicry across human infecting viruses. **a** Experimental Schema and example of 12mer AA proteomic alignment. In total, 134 human-infecting viral proteomes were screened to identify 8mer, 12mer, and 18mer AA k-mers with 3 or less mismatches to a matching human k-mer. **b** Heatmap of scaled percentages of mimicry for 8mers, 12mers, and 18mers with 3 or less mismatches, with viruses aligned by taxonomy. Percents of k-mers, were scaled separately for each mismatch and k-mer length combination. Percentage of viral **c** 8mers, **d** 12mers, and **e** 18mers with 0, ≤ 1 , ≤ 2 , and ≤ 3 mismatches between acute and chronically infecting viruses. Percentage of viral **f** 8mers, **g** 12mers, and **h** 18mers with 0, ≤ 1 , ≤ 2 , and ≤ 3 mismatches between different viral families. Viral families only shown in (**f–h**) if containing at least 5 species. For **c–h** black stars indicate comparisons between

chronic and acute viruses, magenta and green stars indicate comparisons of the Herpesviruses and Poxviruses against all other viruses, dark gray stars indicate comparison between Herpesviruses and Poxviruses. Kruskal–Wallis test was used for multigroup comparison, and Wilcoxon rank-sum test (two-sided) for pairwise comparison. Error bars denote mean \pm standard error of the mean. All p values adjusted for multiple hypothesis testing using Benjamini–Hochberg corrections ($^*p_{\text{adj}} \leq 0.05$, $^{**}p_{\text{adj}} \leq 0.01$, $^{***}p_{\text{adj}} \leq 0.001$). For **c–e** chronic viruses $n = 25$ and acute viruses $n = 103$, and for **f–h** *Arenaviridae* $n = 5$, *Coronaviridae* $n = 5$, *Flaviviridae* $n = 12$, *Herpesviridae* $n = 10$, *Paramyxoviridae* $n = X$, *Phenuiviridae* $n = 5$, *Picornaviridae* $n = 13$, *Polyomaviridae* $n = 5$, *Poxviridae* $n = 9$, *Retroviridae* $n = 5$, *Rhabdoviridae* $n = 9$, *Togaviridae* $n = 11$. Source data are provided as a Source Data file.

percentage of amino acid (AA) 8mers, 12mers, and 18mers between human and viral proteins with up to three mismatches (Fig. 1a). The evaluated AA length (k-mer) was motivated by the known length of T cell recognized epitopes which range on average from 8–12 AA for CD8⁺ T cells and 18–24 AA for CD4⁺ T cells^{10–13}. We found that the phylogenetic relationship between pathogens and the relative amount of mimicry was similar for 8mers, 12mers, and 18mers between closely related viral species (Fig. 1b).

In comparing acute and chronically infecting viruses (infection less than or greater than 21 days respectively), chronically infecting viruses had significantly elevated rates of mimicry for 0, ≤ 1 , and ≤ 2 mismatches for all evaluated k-mers (8, 12, and 18 AAs) and for ≤ 3 mismatches in 12 and 18mers (Fig. 1c–e). On further evaluation by viral family, while both *Herpesviridae* and *Poxviridae* had significantly higher levels of 8mer mimicry at 0 mismatches, only *Herpesviridae* demonstrated significance for ≤ 1 and ≤ 2 mismatches (Fig. 1f). In turn, for 12mer and 18mers, *Herpesviridae* and *Poxviridae* were both significant for 0, ≤ 1 , ≤ 2 , and ≤ 3 mismatches compared to all other viral families (Fig. 1g, h). Notably, *Poxviridae* had significantly more 18mers with 0, ≤ 1 , ≤ 2 , and ≤ 3 mismatches compared to *Herpesviridae* (Fig. 1h), suggesting that *Poxviridae* may be more adept at mimicking longer AA sequences. In line with these findings, we also observed that classifying viruses by genome structure (Baltimore classification, e.g. dsDNA, ssDNA, \pm ssRNA) revealed elevated rates of mimicry in dsDNA viruses (driven largely by Herpesviruses and Poxviruses, Supplementary Fig. 1A–C).

To confirm that these results were not driven by embedded eukaryotic linear motifs (e.g. ATP docking motif), we leveraged the Eukaryotic Linear Motif (ELM) database²¹ to repeat the analysis but exclude any k-mer that contained even a single amino acid of an ELM (Supplementary Fig. 1D–I). This analysis replicated our findings and demonstrated Herpesviruses and Poxviruses were significantly more likely to contain mimicry under 8, 12, and 18 length k-mers, even when filtering for these functional protein motifs (Supplementary Data 3). Additionally, when compared to the mouse proteome, Herpesviruses demonstrated elevated mimicry specific to humans (Supplementary Fig. 1J).

Furthermore, the observed elevation in mimicry across diverse viruses was not limited to specific viral proteins (e.g. known virokinases) but rather was more broadly distributed throughout the proteomes (Supplementary Fig. 2A, B). Additionally, relative amino acid usage did not greatly differ between viruses and was unlikely to explain the significant differences in mimicry (Supplementary Fig. 2C). Thus, most of the mimicry observed was not due to global sequence similarity, rather spread throughout proteins at a low frequency, except in the case of a few known proteins which have been adopted and modified by the virus from the host.

Permutation testing reveals significant rates of mimicry in herpesviruses and poxviruses

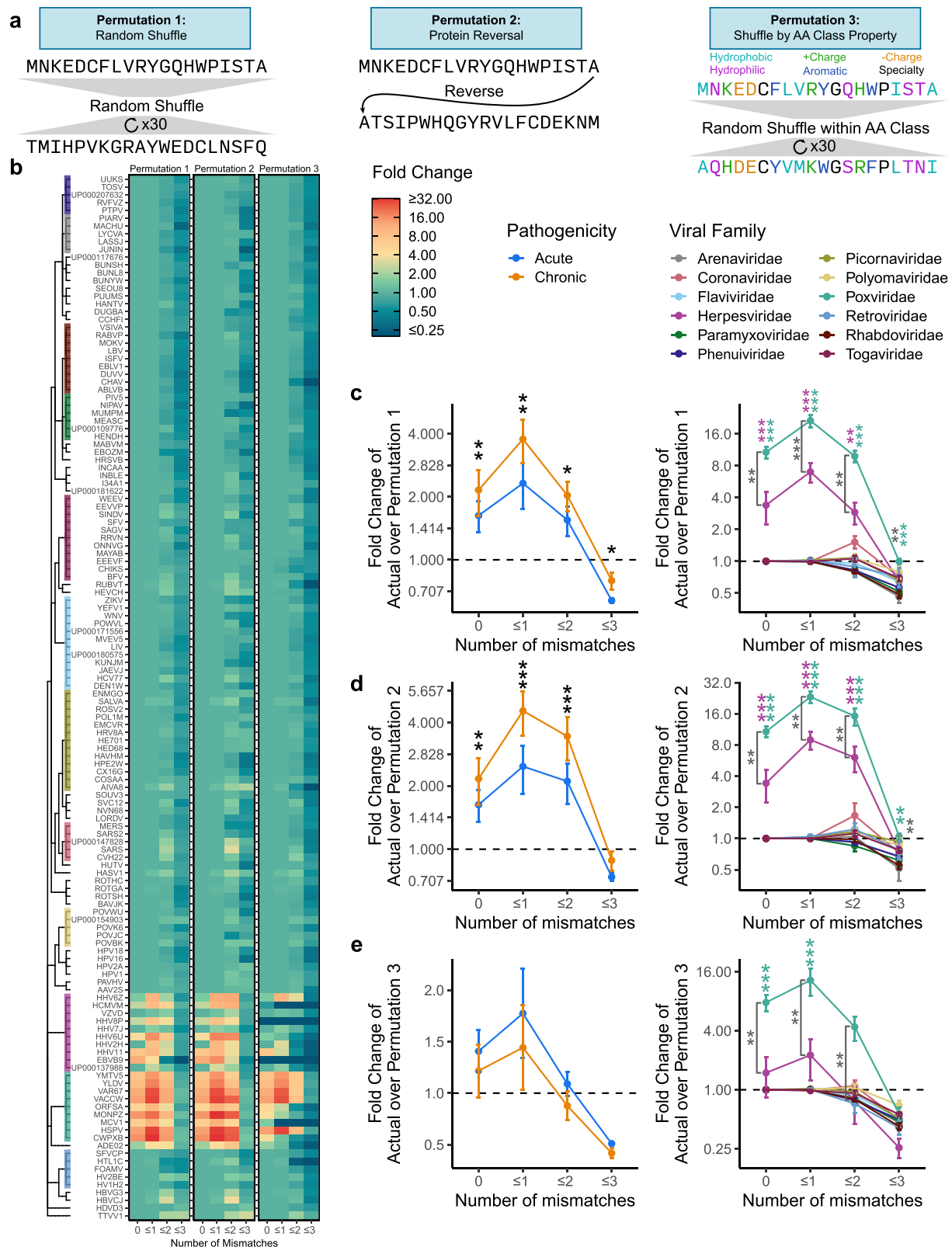
To identify whether the rate of detected mimicry was above random chance, we conducted three permutation tests using 12mers: (1) scrambled viral proteins (permutation 1), (2) reversed viral proteins (permutation 2), and (3) scrambled viral proteins by amino acid

(permutation 3) (Fig. 2a, b). We chose to utilize 12mers over 8 and 18mers for two reasons: first, most viral 8mers had a human 8mer match at ≤ 2 mismatches and second, 18mers were too sparse for detection across most viral families (Fig. 1f, h). When the actual proteomes were compared to the mean of the randomly scrambled proteomes, chronic viruses and the specific families of *Poxviridae* and *Herpesviridae* were confirmed to have significant levels of mimicry at 0, ≤ 1 , and ≤ 2 mismatches (Fig. 2c). However, this permutation strategy does not fully capture the fact that protein sequences are not truly random. Thus, to evaluate the robustness of the permuted results, as an alternative strategy, we also evaluated the frequency of mimicry in viral proteomes with reversed protein sequences. This permutation also confirmed that the rate at which we observed viral 12mers with 0, ≤ 1 , and ≤ 2 mismatches was significantly above random chance in chronic viruses and specifically in *Poxviridae* and *Herpesviridae* (Fig. 2d). Finally, as the most stringent permutation, we scrambled AAs by class (e.g. hydrophobic, hydrophilic, positively charged, etc.), which conserved overall protein structure of AA properties. We observed a similar trend to the first two permutations, with *Herpesviridae* significant for 12mers with 0 mismatches and *Poxviridae* significant for 12mers with 0 and ≤ 1 mismatches (Fig. 2e). The weakened significance compared to the first two permutations likely reflects the overly stringent nature of this permutation as several AAs were intentionally not shuffled (e.g. Cystine, Glycine, and Proline), and AAs in each class ended up shuffled back into their original position at a higher frequency compared to the first two permutations.

Characteristics of peptide mimicking by viral family

Having confirmed mimicry was significantly higher in *Herpesviridae* and *Poxviridae*, we next evaluated the AA length of observed mimicry. Due to the nature of the rolling window approach used in the alignment, we could identify when several k-mers in a row aligned to the same protein resulting in a k-mer “run” (example shown in Fig. 3a). These runs allowed us to identify how long the potential mimicry sequence could be (e.g. three 12mers in a row for a single protein forming a single 14mer, five 12mers in a row forming a single 16mer). Using this approach, several viral families were identified as having longer runs on average including *Herpesviridae* and *Poxviridae* (Fig. 3b and Supplementary Fig. 3). Notably, there were several very long k-mers (>30 AA) that reflect known full protein mimics such as vIL-10, a viral analog of an anti-inflammatory cytokine²².

Additionally, we also identified some viral families that had more “multi-mimics” where a single viral 12mer had more than one human 12mer with 3 or less mismatches (Fig. 3c, d). This result reflects that these 12mers are mimicking a human motif scattered across multiple genes that may perhaps be more advantageous to mimic. Interestingly, when evaluating the mean rate of multi-mimics versus the percentage of viral 12mers participating in mimicry (i.e. percent of 12mers with 3 or less mismatches to a human 12mer) for each viral protein from across all proteomes in the cohort, we observed that multi-mimics are more frequently found in proteins with lower percentages of mimicry (Fig. 3e). Notably, overall *Herpesviridae* and *Picornaviridae* proteins display a greater degree of multi-mimicking (Fig. 3d, e).



Viral mimics target key human cellular pathways

Next, we evaluated whether the human mimicked proteins were overenriched in specific biological pathways. Using the KEGG database²³, we identified several pathways that were significantly enriched across different viral proteomes (Fig. 4), with the pathways primarily relating to cellular replication and inflammation (Figs. 5a and S4). Notably, many pathways' significance was a result of "multi-

mimics" that reflect mimicry of central human motifs in the pathways. When enrichment testing was repeated with only viral mimics that mimicked five or greater different gene products, many pathways were still significant (Fig. 4).

Leveraging the Human Protein Atlas, we also evaluated the rate at which proteins from specific cell types, tissues, or organ systems were mimicked, though no strong association was observed

Fig. 2 | Permutation confirms significant mimicry in poxviruses and herpesviruses. **a** Example schema of the three permutation strategies (random, protein reversal, AA class shuffle). **b** A heatmap of the fold change of actual mimicry over mimicry in permutations 1–3, aligned by viral taxonomy. Results of the **c** random permutation (permutation 1), **d** protein reversal permutation (permutation 2), and **e** AA class shuffle permutation (permutation 3) for chronic vs. acute viruses and by viral family. Viral families only shown in (c–e) if containing at least 5 species. (Black stars indicate comparisons between chronic and acute viruses, magenta and green stars indicate comparisons of the Herpesviruses and Poxviruses against all other viruses respectively, dark gray stars indicate comparison between Herpesviruses

and Poxviruses. Kruskal–Wallis test was used for multigroup comparison, and Wilcoxon rank-sum test (two-sided) for pairwise comparison. Error bars denote mean \pm standard error of the mean. All p values adjusted for multiple hypothesis testing using Benjamini–Hochberg corrections (* $p_{\text{adj}} \leq 0.05$, ** $p_{\text{adj}} \leq 0.01$, *** $p_{\text{adj}} \leq 0.001$). For c–e chronic viruses $n = 25$ and acute viruses $n = 103$, *Arenaviridae* $n = 5$, *Coronaviridae* $n = 5$, *Flaviviridae* $n = 12$, *Herpesviridae* $n = 10$, *Paramyxoviridae* $n = X$, *Phenuiviridae* $n = 5$, *Picornaviridae* $n = 13$, *Polyomaviridae* $n = 5$, *Poxviridae* $n = 9$, *Retroviridae* $n = 5$, *Rhabdoviridae* $n = 9$, *Togaviridae* $n = 11$. Source data are provided as a Source Data file.

(Supplementary Fig. 5). Instead, pathogens mimicked from a myriad of proteins non-specific to tissues or cell types, in line with our finding that mimicry was dispersed throughout viral proteomes (Supplementary Fig. 2A, B).

Viruses' mimicry targets proteins from autosomes and the X chromosome

To further investigate the potential of evolutionary pressure on viral molecular mimicry, we evaluated chromosome location of the mimicked human proteins. As the Y chromosome is only carried in males, we hypothesized that viruses would display preference for mimicking autosome and X chromosome proteins found in the entire population. Using the reversed proteome as a reference, viruses had significantly greater mimicry to proteins from the autosomes and the X chromosome (Fig. 5b). The result was replicated with the randomly shuffled and AA class shuffled permutations (Supplementary Fig. 6).

Greater viral mimicry in latently expressed proteins of EBV compared to lytic proteins

To analyze whether the timing of protein expression during the viral life cycle (Fig. 6a) may explain degree of mimicry, we evaluated one of the top mimicking pathogens, EBV, as expression of its latent and lytic proteins have been well characterized²⁴. Utilizing previously reported gene sets²⁴, we found that latent stage proteins displayed significantly more mimicry than lytic stage proteins (Fig. 6b). This observation was not replicated in HHV8 or CMV, however these viruses still maintained overall high levels of mimicry in both their latent and lytic proteins (Supplementary Fig. 7).

Enrichment of thymic negatively selected proteins

We hypothesized that some viruses may intentionally acquire mimicry to protect their epitopes from being targeted by the adaptive immune system, as thymic negative selection would deplete T cells capable of responding to human mimicked epitopes. To test this hypothesis, we evaluated the proportion of the viral mimics that aligned to genes known to be expressed in human thymus antigen presenting cell populations (mTECs, CD19+ (B cells), CD141+ (dendritic cells), CD123+ (dendritic cells), or “Any” reflecting the list of genes expressed in any of these cell types), as genes not expressed in the thymus during negative selection have more limited mechanisms preventing formation of auto-reactive T cells (i.e. peripheral tolerance)²⁵. Using Gabrielsen et al.'s²⁶ reported gene expression of human thymic cells, we observed that the viral mimicking sequences disproportionately mimic proteins found in human thymic cells when compared to the frequency of all human proteins expressed in those cell types (Fig. 6c)²⁶. Notably, every virus evaluated in the *Poxviridae* and *Herpesviridae* family maintained a robust relationship with at least one thymic cell type.

Molecular mimicry explains portions of the multiple sclerosis auto-antibodyome

Finally, we sought to evaluate whether our viral mimics could explain auto-antibodies in a pathological autoimmune disease. Due to the strong epidemiological association between multiple sclerosis (MS) and EBV²⁷, we leveraged the recently generated data from Zamecnik

et al.²⁸ to further evaluate how linear mimicry may play a role in the MS autoimmune pathophysiology. Previously, Zamecnik et al.²⁸ screened for auto-antibodies against the human proteome using Phip-seq, a technique used to identify auto-antibodies against 49 AA long linear peptides from the human proteome. The authors found that approximately 8% of MS patients will have an auto-antibody signature that contains a motif found in EBV's Envelope Glycoprotein M and BRRF2, with this signature referred to as an immunogenic cluster (IC Cluster).

In line with Zamecnik et al.'s findings, we observed that even non-IC cluster MS auto-antibodies are more likely to contain EBV mimicry (Fig. 7a). This association is even stronger for the most prevalent auto-antibodies. In fact, 87% (20/23) of the top MS auto-antibodies found post MS diagnosis outside of the IC-cluster contained an EBV 8mer with ≤ 2 mismatches compared to 72% in the healthy control auto-antibodies (Fig. 7b). Furthermore, we identified that these additional MS auto-antibodies were not fully concomitant with the IC signature, suggesting that EBV mimics between MS patients may differ (Supplementary Fig. 8).

Discussion

In this study, we evaluated 134 proteomes from all existing human infecting viruses for short linear molecular mimicry to human proteins, comprising one of the largest viral screens to date. These results broaden the current knowledge on the rates of mimicry, as prior studies have primarily focused on limited number of pathogens or mimics predicted to bind to common HLA haplotypes^{16–19}. Furthermore, our research reveals underlying cellular and biological patterns associated with mimicry, highlighting the complexity and molecular consequences of host-pathogen interactions. Overall, our results both capture a more holistic picture of viral mimicry and viral evolutionary adaptations relative to prior studies and offer potential avenues for targeted therapeutic strategies in autoimmune diseases.

Notably, we found that viruses within the *Herpesviridae* (e.g. EBV, CMV, HSV, Varicella Zoster) and *Poxviridae* (e.g. Monkeypox, Orf virus, Molluscum contagiosum) families have significantly more short linear mimicry relative to all other viruses. Additionally, we demonstrate that the observed elevated rates of mimicry in Herpesviruses and Poxviruses are significantly over-represented compared to permuted forms of their proteomes, suggesting an evolutionary pressure for mimicry. In line with evolutionary pressure, we observe that mimicry to proteins from the autosomes and X chromosome was significantly greater than proteins from the Y chromosome, which would offer less evolutionary benefit to the virus due to only existing in approximately half of the human population. Furthermore, we identified that viruses disproportionately mimicked proteins expressed during negative selection in the thymus (a key biological process in apoptosis of auto-reactive T cells during their development)^{4,25}, in concordance with our hypothesis that mimicry might improve evasion of the adaptive immune system. In addition, we observe that for one of the most common herpesviruses (EBV), mimicry is significantly higher in latently expressed proteins compared to lytic proteins, further suggesting evasion of the adaptive immune system may be more advantageous during the latent stage of infection. These findings collectively raise an intriguing question about the co-evolution of viruses with the

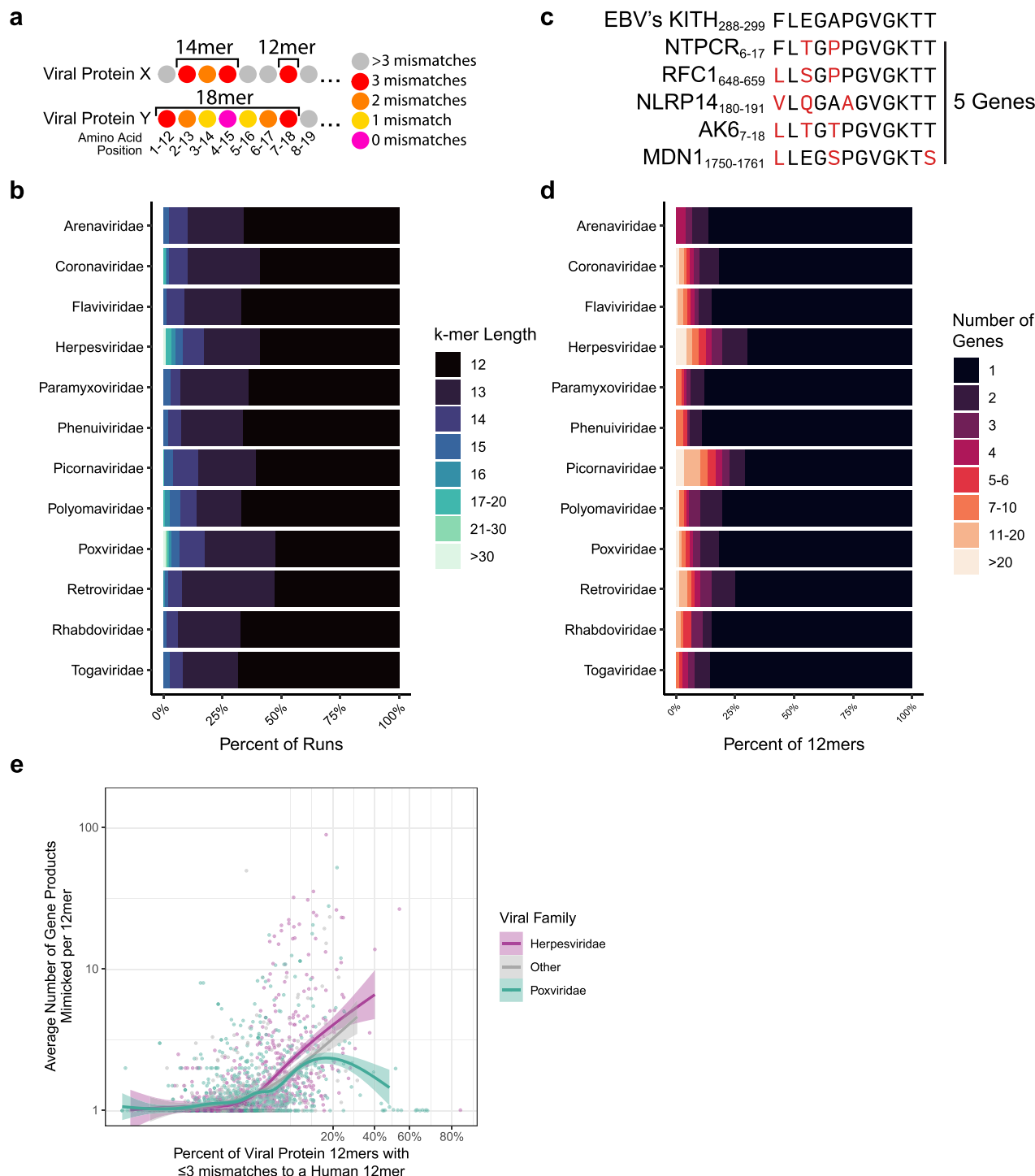


Fig. 3 | Viral mimicry is diverse in length and differentially targets human protein motifs. a An example of an 18mer and 14mer comprised of 12mers. **b** Average percent of mimics at varying k-mer lengths. **c** An example of multi-mapping in which a single viral 12mer aligns to multiple human proteins with the criteria of 3 or less mismatches. **d** Percent of viral 12mers and their corresponding

number of mimicked human genes (multi-mapping). **e** Average number of host gene products mimicked, plotted against the percent of the viral protein that participates in mimicry (defined as a 12mer with 3 or less mismatches). Trendlines represent fit and 95% confidence interval from a generalized additive model. Source data are provided as a Source Data file.

human genome and whether there may be consequences from viral mimicry to the host.

Interestingly, *Herpesviridae*, which are chronically infecting viruses, have been previously associated with autoimmune diseases (e.g. EBV in MS²⁷, CMV in systemic lupus erythematosus²⁹), with molecular mimicry hypothesized as one of the major mechanisms leading to cross-reactive autoimmune responses^{5–8}. Similarly, infection with

acute respiratory and gastrointestinal viruses has also been observed more frequently in weeks prior to autoimmune disease onset^{30–33}; however, the mechanistic link between acute infecting pathogens and autoimmunity remains unknown, unlike with *Herpesviridae*. Of note, recent work has shown that during infection with COVID-19, ~41% of patients will have reactivation of EBV, ~28% HSV1/HSV2, ~25% CMV, and ~34% HHV6³⁴. Thus, it is possible that herpesviruses, as opportunistic

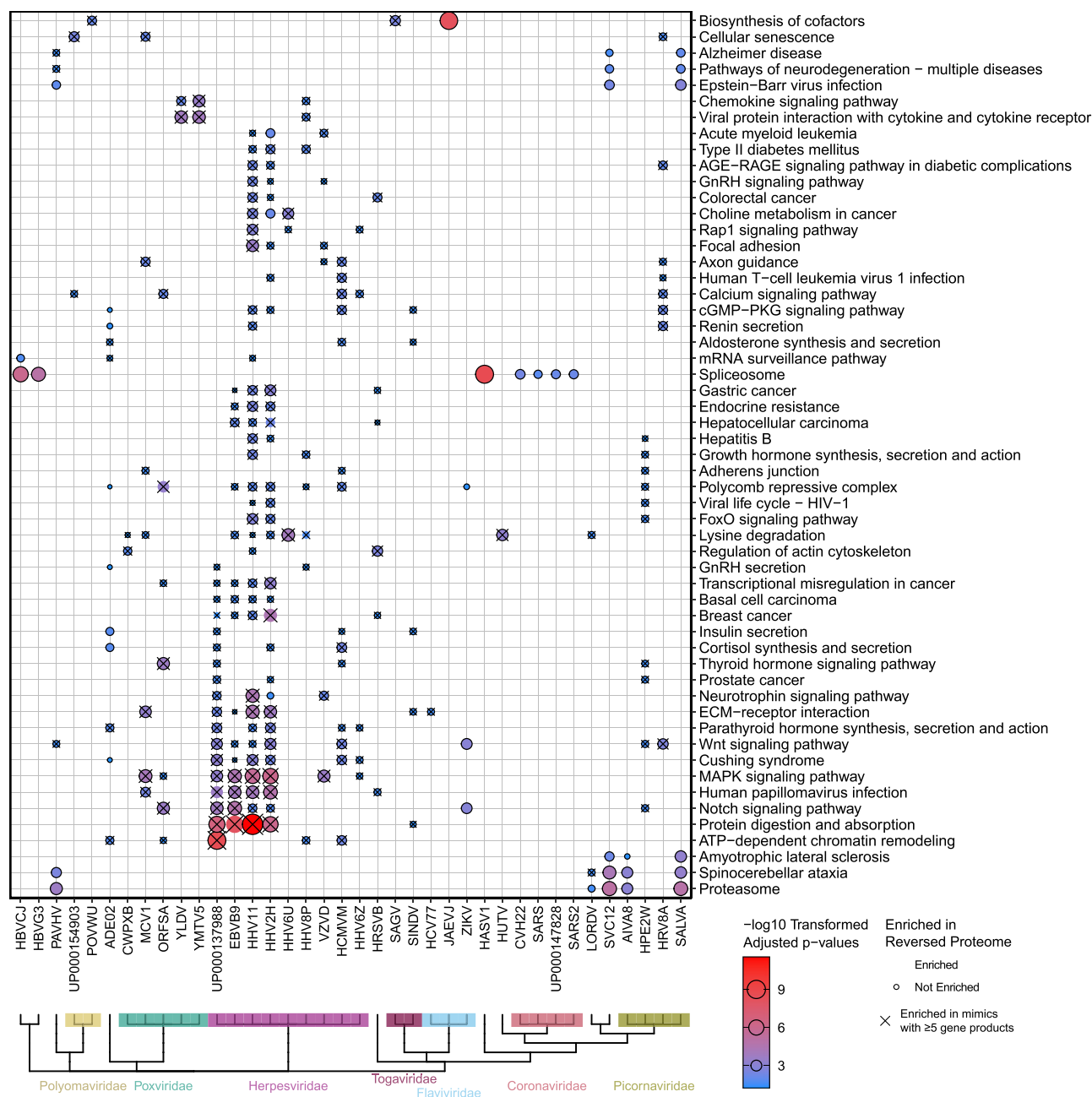


Fig. 4 | Viral mimicry is concentrated towards proteins from key human biological pathways and leverages mimicry of common motifs. Hypergeometric enrichment testing of KEGG pathways for human proteins that are mimicked by each virus. Significant enrichment is displayed as a dot if the adjusted p -value ≤ 0.05

(calculated using Benjamini–Hochberg correction) and is outlined if that biological enrichment was not observed in the reverse proteome permutation (permutation 2). Only pathways significant in at least 3 viruses are shown, with all viruses and significant pathways shown in Supplementary Fig. 4.

pathogens, may take advantage of other acute infections to exit latency, which may explain why many acute viral infections appear to be epidemiologically associated with autoimmunity.

Similarly to *Herpesviridae*, members of the *Picornaviridae* family also have a greater percentage of mimicry that targets human protein motifs compared to other viral families. As *Picornaviridae* infections often precede initiation of auto-immune diseases, such as type-1 diabetes³¹, mimicry by these viruses to human protein motifs may carry a greater autoimmune risk and explain why both *Herpesviridae* and *Picornaviridae* are epidemiologically more associated with autoimmune diseases.

In comparison, poxviruses have not been associated with autoimmune pathology, despite also displaying significant mimicry.

This may be explained by poxviruses mimicking a less diverse number of proteins, focusing instead on functional mimics to modulate host response (e.g. NFkB-I), which often are longer and more accurate mimics^{35–39}. In contrast, herpesviruses display widespread distribution of short and low accuracy mimics (as evident by herpesviruses being significantly higher than poxviruses for 12mers with ≤ 3 mismatches, but not ≤ 2 mismatches), which might be sufficient to extort host negative selection to improve evasion of the adaptive immune system. Furthermore, due to establishing a latent infection, the immune system will more frequently encounter and activate against herpetic viruses, possibly increasing the likelihood that a mimicked region is targeted. Conversely, poxviruses are more rarely encountered in the human population and

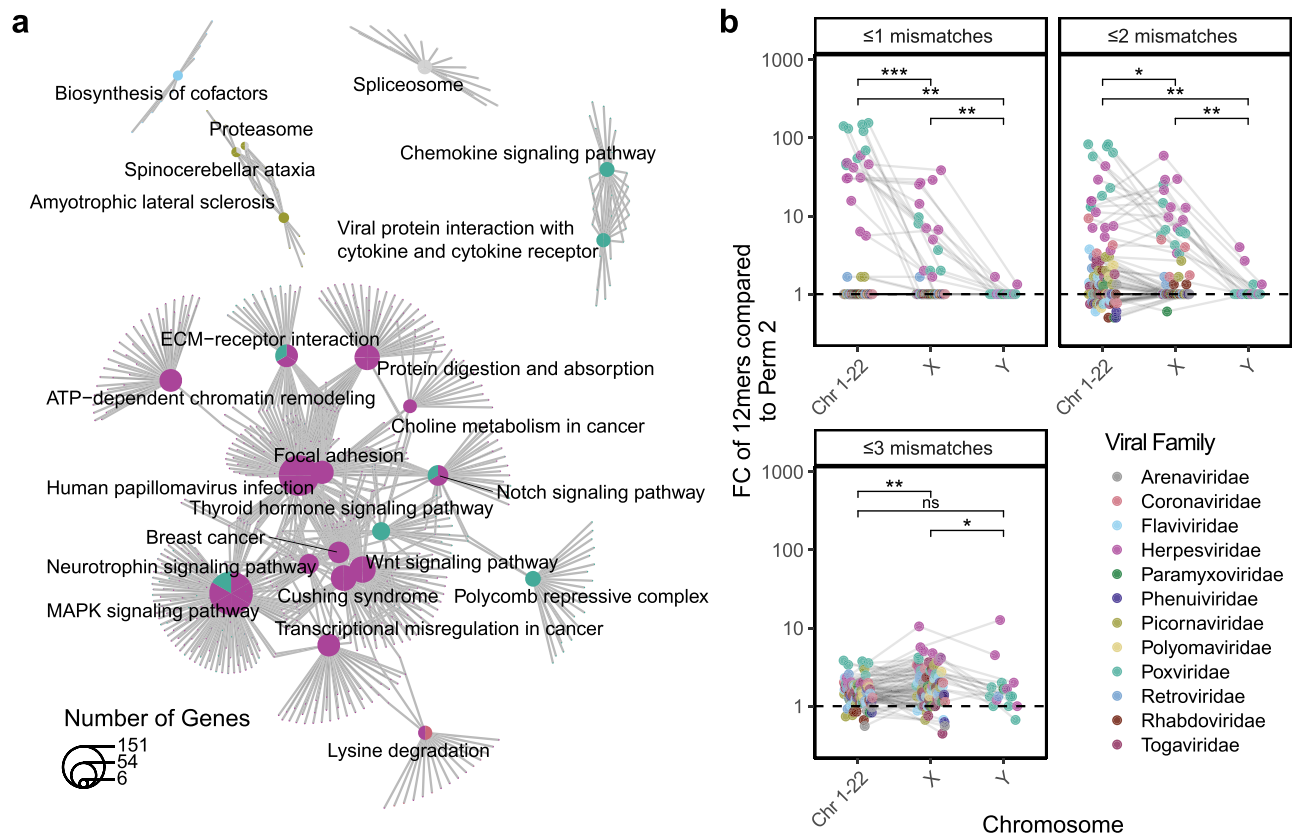


Fig. 5 | Viral mimicry targets critical cellular pathways of cellular replication and inflammation and avoids proteins from the Y chromosome. a Shared overlap between the significant pathway from Fig. 4 reveal broad roles of inflammation and cellular replication amongst the pathways. Pathways and genes are represented by a pie chart, colored by the proportion of viruses belonging to each family that were significantly enriched for the pathway, with lines connecting genes to their respective pathways. **b** Fold change of the percent of mimics whose human

counterpart is encoded on either an autosome (Chromosomes 1-22), X, or Y chromosome over the rate in the reversed proteome (permutation 2). For **b**, the Wilcoxon summed-rank test (two-sided) was used for paired pairwise comparison. All p values adjusted for multiple hypothesis testing using Benjamini-Hochberg corrections (* $p_{adj} \leq 0.05$, ** $p_{adj} \leq 0.01$, *** $p_{adj} \leq 0.001$). Source data are provided as a Source Data file.

rarely establish latency, thus representing a less frequently encountered infection^{40,41}.

Previous studies on the development of autoreactive T cells due to mimicry have identified that cross-reactivity may primarily be evoked in non-naïve T cells⁴². It is hypothesized that in the naïve state, the T cell receptor (TCR) is specific to only the viral peptide, but as the T cell transitions into a memory state post-infection and develops a lower threshold for activation, the TCR becomes capable of responding to both the viral and the human mimicked peptide. This hypothesis is intriguing as it may explain why cross-reactive T cells may not be negatively selected during development, and why molecular mimics targeted for autoreactivity often have mismatches in the epitope^{43,44}. Interestingly, this is in line with our observation that herpesviruses have higher levels of short and low accuracy mimicry which may be prime for eliciting autoimmunity.

The high dispersal of mimicry throughout a viruses' proteome at an overall low frequency is of interest, as it potentially provides insights into mechanisms by which mimicry is acquired and maintained. First, low frequency proteome-wide dispersal of mimicry might better avoid interference with protein structure and function, thus addressing the molecular mimicry trade-off hypothesis²⁰. Second, low frequency mimicry also suggests that most of these events are not due to horizontal genetic transfer, except for a few outliers that do reflect horizontal gene acquisition (Fig. 3e). Similarly, we did not observe enrichment of mimicry in proteins specific to the tissues and cell types each virus infects, an observation that would otherwise suggest more

convincing evidence of horizontal gene transfer. The molecular mimicry trade off hypothesis also identifies that adding mimicry to the viral genome may lengthen replication time, an important trait in overall virulence. Interestingly, we observe that viruses with the most mimicry have some of the largest genomes, possibly indicative of less evolutionary pressure on replication time which might enable them to carry more mimicry. Thus, despite the potential drawbacks of mimicry, its usage on a micro scale appears may prove viable.

The observed mimics (12mers) were significantly enriched for key biological pathways, primarily pertaining to cellular replication and immune signaling, suggesting that some of these mimics may be "functional". Functional mimicry is defined as a pathogenic mimic capable of eliciting complete, partial, or modified effects of the original host protein, with some examples including mimicry of IL10 and NFkB-I, which all further modulate host immune responses^{36,45}. Often, functional mimicry is a byproduct of horizontal genetic transfer and has traditionally been studied in the context of mimicry of complete proteins. However, transcription factor and protein binding sites can comprise small portions of a protein, as small as 4 AAs, which the 12mers in this study would encompass. Thus, mimicry on a smaller scale might also offer some functional effects. Interestingly, the significant enrichment we observe towards cellular replication and immune signaling pathways was often the result of mimicry of key motifs found throughout multiple proteins in the pathway. Thus, some of these mimics might offer functional benefit, particularly in modulating signaling associated with the original mimicked protein.

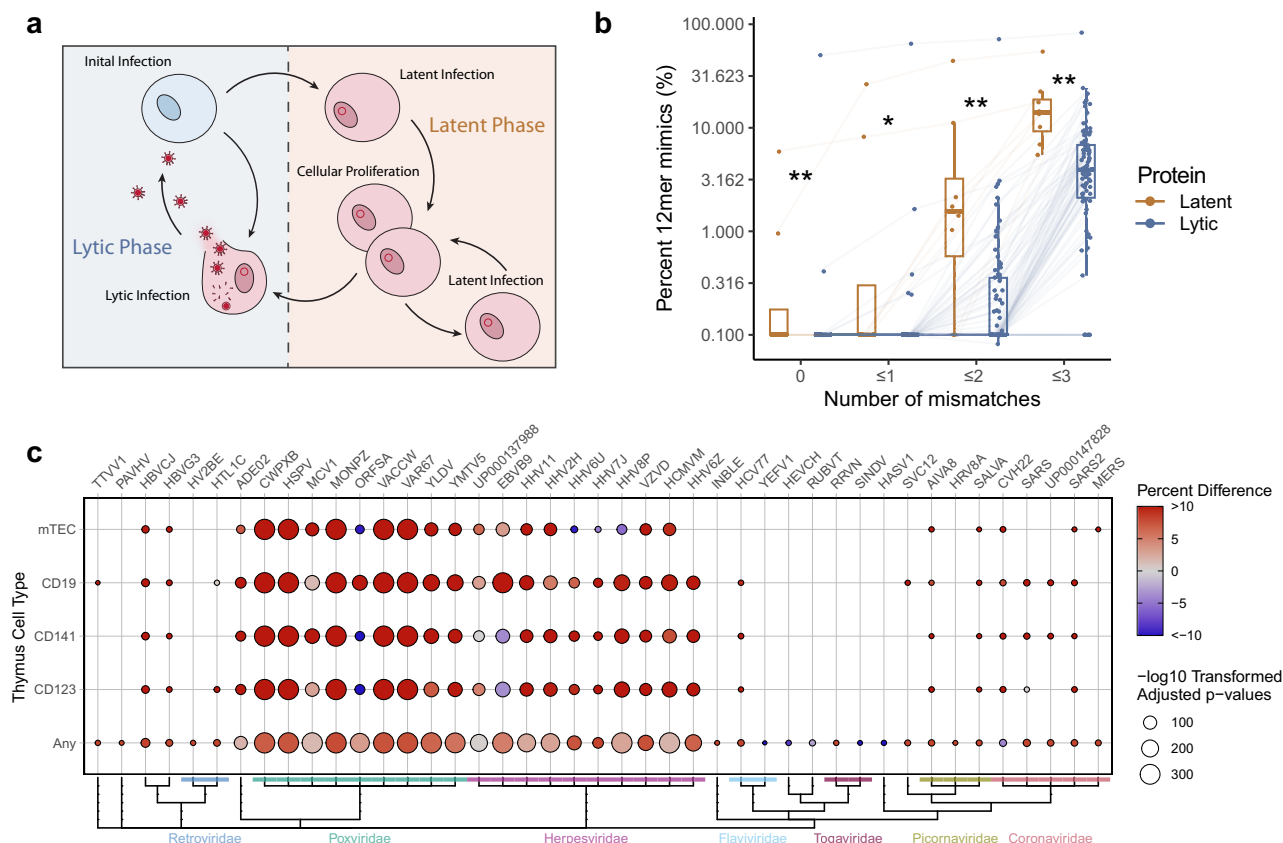


Fig. 6 | EBV latent proteins have elevated mimicry and viral mimicry targets negatively selected proteins. a Schematic of latent vs. lytic stages of viral replication. **b** Percent of 12mers with 0, ≤1, ≤2, and ≤3 mismatches in latent and lytic EBV proteins. Each point denotes a single EBV protein connected across the mismatch levels with boxplots denoting the median as the midline, the bounds of the box defining the 25th and 75th percentile, and the whiskers denoting the 1.5 times the interquartile range. For comparison of latent vs. lytic genes in (**b**), the Wilcoxon summed-rank test was used with Benjamini–Hochberg corrections (* $p_{adj} \leq 0.05$

and ** $p_{adj} \leq 0.01$). For **b** latent genes $n = 8$ and lytic genes $n = 84$. **c** Percent difference of mimics expressed in mTEC, CD19+ B cells, CD141+ dendritic cells, CD123+ dendritic cells, or “any” of these cells in the human thymus (all antigen presenting cells), compared to the percentage of all human genes expressed by the cell type (as determined by RNA expression in Gabrielsen et al.²⁶). Adjusted p value calculated using fisher exact test (two-sided) with Benjamini–Hochberg corrections. Source data are provided as a Source Data file.

Finally, we demonstrate the direct translatability of our results to autoimmunity in MS, one of the most common autoimmune diseases²⁸. The recent study by Zamecnik et al.²⁸ detailed an auto-antibody profile in ~8% of MS cases, believed to stem from EBV-induced molecular mimicry. Here, we extend these findings and demonstrate that even the rare MS auto-antibodies not in this signature, found in only ~2–4% of MS patients (and absent in controls), also display over enrichment of EBV mimicry, suggesting that EBV mimicry may play a larger role in the auto-antibodyome of MS than previously thought.

Our study also had several limitations. First, we specifically focus on linear mimicry, and as a result we may have missed structural 3D mimics, which might be important, particularly to cross-reactive antibody binding. Second, we use the number of mismatches in AA 8mer, 12mer, and 18mers to identify approximate patterns of molecular mimicry; however, mismatches in key positions of the k-mer may differentially affect HLA binding, thus total number of mismatches may not best reflect mimicry in T cell presentation. Yet, the consistent trend of elevated rates of mimicry across all three k-mer lengths and across various levels of mismatches supports the idea that the observed mimicry is robust. Third, the evolutionary gain of molecular mimicry was not directly confirmed in this study due to the limited availability of historical data for validation. Fourth, these results use the consensus viral and human proteomes which do not fully capture genetic diversity in the population.

Molecular mimicry may also play a role in human endogenous retroviruses (HERVs) which comprise ~8% of the human genomes and are frequently expressed in many chronic diseases, such as autoimmunity and cancer^{46,47}. However, due to HERVs being a part of the human genome and being expressed in the thymus⁴⁸, there is some central tolerance, although incomplete, against HERVs⁴⁹. Thus, in future studies, it will be critical to elucidate how adaptive tolerance is maintained against different HERVs.

The relevance of our findings may also extend to explain the emergence of autoimmunity post CAR-T and immune checkpoint inhibitor therapies^{50–55}. Recently advances have enabled more personalized treatments, with CAR-T therapy capable of targeting personalized “neo-antigens” (novel protein sequences that occur as a result of genetic mutations or incorrect protein splicing) from sequenced tumor biopsies^{56–58}. Due to the founding sequence being human, neo-antigens may pose a similar threat to molecular mimicry, and result in cross-reactive immune responses. Furthermore, the concomitant use of checkpoint inhibitors with CAR-T may exacerbate cross-reactivity due to the weakened lymphocyte activation thresholds^{59,60}. Similarly, neo-antigens may exist in unique disease-specific isoforms of proteins^{61,62}, often occurring at novel splice junctions, and thereby contribute to molecular mimicry in different autoimmune syndromes.

Lastly, given our data and emerging role of viruses in autoimmune diseases, a better understanding of molecular mimicry may also inform future therapies and interventions in autoimmunity. For

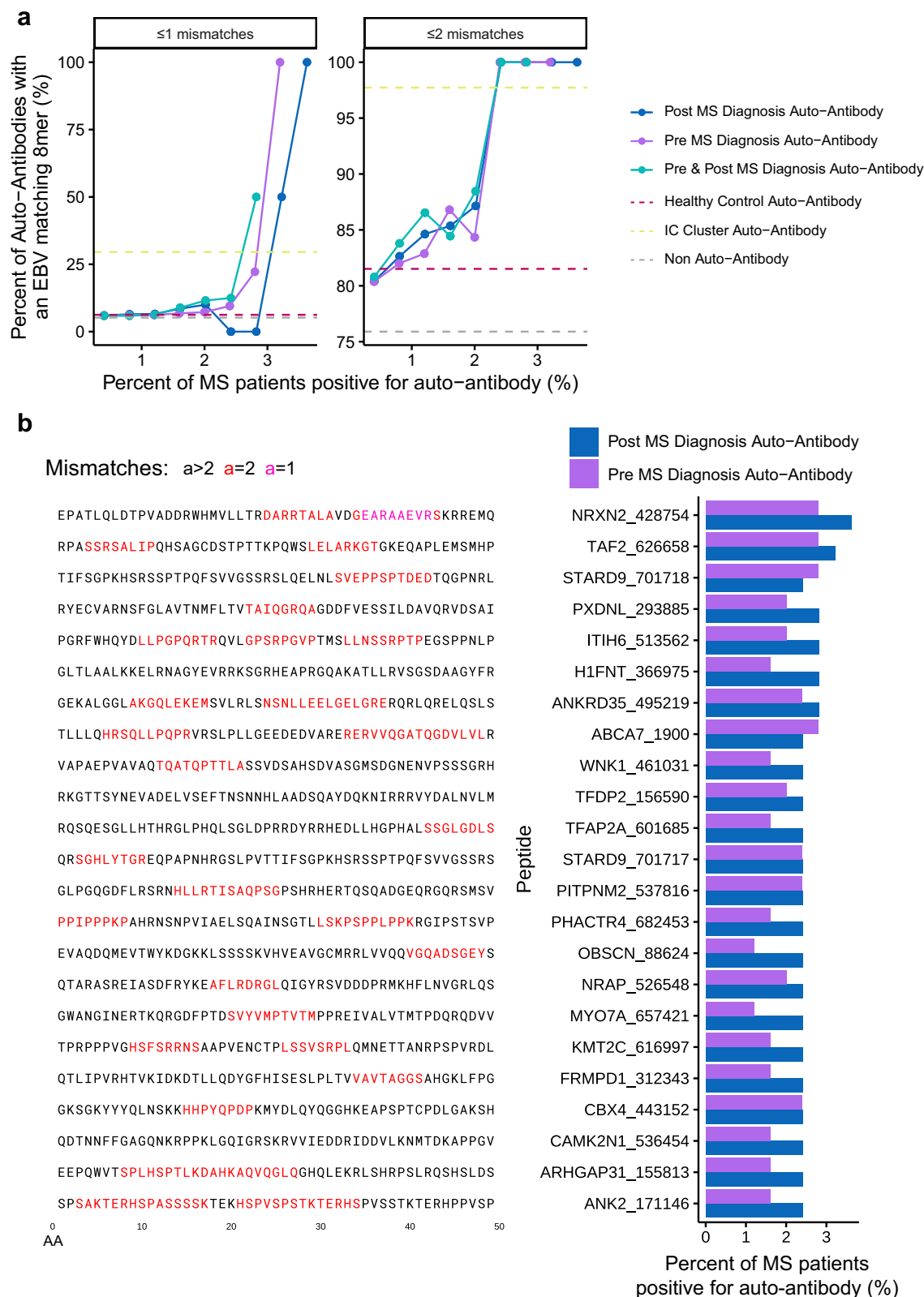


Fig. 7 | EBV mimics are more frequent in MS auto-antibodies. **a** Frequency of EBV mimicry (defined as EBV 8mers with ≤ 1 (left) or ≤ 2 mismatches (right)) in auto-antibody targets in MS patients pre and post diagnosis. Dotted lines show the rate of peptides with an EBV mimic for auto-antibodies found in either in no participant, in healthy controls, or in the previously identified IC Cluster (Zamecnik et al.²⁸).

b Left: Peptide sequences of the top non-IC Cluster and most frequent post MS diagnosis auto-antibodies. Right: Percentage of MS patients positive for the top non-IC Cluster auto-antibodies at both the pre and post diagnosis timepoints. Source data are provided as a Source Data file.

Table 1 | Key datasets and resources used in this study

Dataset/Key resource	Citation	Link to data	Version	Notes
Viral Zone: List of Human infecting viruses	Masson et al. ⁶⁵	https://viralzone.expasy.org/678		
NCBI Taxonomy: Phylogenetic Information of Pathogens	Schoch et al. ⁶⁶	https://www.ncbi.nlm.nih.gov/taxonomy		
Eukaryotic Linear Motif Database	Kumar et al. ⁷¹	http://elm.eu.org/elms	Version: May 13th, 2024, 4:35 p.m.	
KEGG: Biological Pathway Gene Sets	Kanehisa and Goto ²³	https://www.genome.jp/kegg/	Release 109.0 (2024/02)	
Human Protein Atlas: Cell Type, Tissue, and Organ System Protein Expression	Uhlén et al. ⁶⁷	https://www.proteinatlas.org/download/normal_tissue.tsv.zip	Version 22.0 [2022-12-07]	Tissue expression was further grouped into organ system with this mapping provided with the source code.
UniProt: Pathogen and Human Proteomes	UniProt Consortium ⁷¹	https://www.uniprot.org/		See Supplementary Data 1 for full list of proteomes used.
List of EBV Latent Proteins	Kang et al. ²⁴	https://www.nature.com/articles/emm201484/tables/1		Table 1 of Kang et al. ²⁴
Human Thymus Gene Expression Dataset	Gabrielsen et al. ²⁶	https://ndownloader.figstatic.com/files/15667655		Supplementary Data 3 from Gabrielsen et al. ²⁶
Multiple Sclerosis Phip-seq Dataset	Zamecnik et al. ²⁸	https://github.com/UCSF-Wilson-Lab/MS_DoD_and_ORIGINS_study_data		

example, in MS, clinical interventions for EBV management are currently being evaluated including infusion of EBV specific autologous T cells and repurposing of anti-virals such as tenofovir/emtricitabine to limit EBV replication^{63,64}. Thus, identifying pathogens as risk factors in specific autoimmune diseases could pave the way for tailored therapeutic strategies that incorporate both the autoimmune disease-specific therapy as well as targeted management of the implicated viral infection(s).

Methods

Proteome retrieval

The full list of human infecting viruses was retrieved from ViralZone⁶⁵ (Supplementary Data 1) with phylogeny obtained from NCBI Taxonomy⁶⁶. Chronic infecting pathogens were defined as pathogens whose infection on average persists past 21 days (Supplementary Data 1). All pathogen proteomes used were retrieved from UniProt (Supplementary Data 1) and compared to the reference human proteome (UP000005640) using suffix array kernel sorting to identify matching 8, 12 or 18 amino acid (AA) fragments (8mer, 12mer, and 18mers) with up to 3 mismatches in silico (Fig. 1a–c). Only the canonical protein sequences were used for the pathogen proteomes (i.e. no isoforms were used). The 8mer, 12mer, and 18mer length were chosen for their respective association with processing and presentation of peptides on MHC-I (8mer and 12mer) and MHC-II (12mer and 18mer) complexes to CD8+ and CD4+ T cells respectively^{10–13}. In the case of an unknown AA (X) in the human or viral proteome, it was conservatively called a mismatch to all AAs. Proteins that have been deleted from UniProt as of September 12th, 2023 were removed from the previously downloaded data from May 2022. Additionally, duplicated proteins in either the viral or human proteomes were removed. When calculating the percent of 8mer, 12mer, or 18mers that had 3 or less mismatches (Fig. 1), a mimic was not counted if both the human and pathogen proteins had the same enzyme commission number or were reported as having the same chemical reaction in UniProt due to concern of homology between similarly functioning proteins (this filter was only used for Fig. 1 analyses). A supplementary analysis that also excluded k-mers that contained any instance of a eukaryotic linear motifs (ELMs) from being included in mimicry calculations is reported in Supplementary Fig. 1, and supported the observations of the non-ELM filtered analysis. Summarized data and statistical results comparing the rate of 8mer, 12mer, and 18mers are reported in Supplementary Data 2. Summarized data and statistical results for the supplemental analysis comparing the rate of 8mer, 12mer, and 18mers while filtering out for ELMs are reported in Supplementary Data 3.

Permutation testing of 12mers

To evaluate whether the rate of 12mers was above the expected random chance three permutation strategies were utilized. The first strategy randomly shuffled all proteins in each viral proteome 30 times and fold change (FC) was calculated between the actual proteome and the average of the 30 shuffles. The second strategy reversed all proteins in each viral proteome, for which we calculated the FC over the actual proteome. The third strategy was a quasi-random shuffle with AA shuffled by class defined in Fig. 2a, in which all proteins in each viral proteome were shuffled 30 times and FC was calculated of the actual proteome to the average of the 30 shuffles. Statistical testing results for Fig. 2 are provided in Supplementary Data 4.

Enrichment analysis of 12mers

The hypergeometric test was used to evaluate significantly enriched KEGG (v 109.0) pathways mimicked by individual pathogen proteomes using the R package clusterProfiler²³. Only pathways significantly mimicked by at least three viruses are shown in Fig. 4, with all results shown in Supplementary Fig. 4 and reported in Supplementary Data 5. Enrichment of cell type and cellular location of proteins used data

from the human protein atlas (v 22.0), with results reported in Supplementary Fig. 5⁶⁷. Enrichment of mimics to proteins from specific chromosomes and from genes expressed by human mTEC cells in the thymus was evaluated with the Fisher's exact test, using genes reported in Gabrielsen et al. with results reported in Supplementary Data 6²⁶.

Analysis of Phip-seq autoantibodies

Data was retrieved from Zamecnik et al.²⁸. Reads per 100,000 (rpK) were used to calculate FC for each peptide of MS patients based on the control mean and distribution separately for the pre-diagnosis and post-diagnosis timepoint. Peptides with FC > 10 were called as hits.

Suffix array kernel sorting was used against the Phip-seq peptide library to identify 12mers with 3 or less mismatches and 8mers with 3 or less mismatches (following the same protocol as used for the proteome analysis except using the specific human proteome used in the production of the Phip-seq library). The rate of 12mer mimics in "hit" peptides was quantified and compared to non-hit peptides and correlated with prevalence in all samples.

Statistical analysis

Computational analyses were performed using the Biomedical Research Computing Facility at UT Austin, Center for Biomedical Research Support RRID#: SCR_021979 and Texas Advance Computing Center (TACC) The University of Texas at Austin. K-mer alignment was conducted using Julia 1.6.0⁶⁸. All other analyses were conducted using R 4.3.1⁶⁹. The non-parametric tests of Kruskal–Wallis for multiple group comparisons and Wilcoxon rank-sum test for pairwise group comparisons were used for all testing unless otherwise stated. All *p* values were adjusted using Benjamini–Hochberg multiple test correction and are referred to as “*p*.adj” or “adjusted *p*.values”.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

Publicly available datasets used in this study are reported in Table 1. The data generated in this study have been deposited at GitHub, <https://github.com/melamedlab/MolecularMimicry> (<https://doi.org/10.5281/zenodo.11411891>)⁷⁰. Source data are provided with this paper.

Code availability

All code used in this study is available at <https://github.com/melamedlab/MolecularMimicry> (<https://doi.org/10.5281/zenodo.11411891>)⁷⁰.

References

- Tortorella, D., Gewurz, B. E., Furman, M. H., Schust, D. J. & Ploegh, H. L. Viral subversion of the immune system. *Annu. Rev. Immunol.* **18**, 861–926 (2000).
- Gowthaman, U. & Eswarakumar, V. P. Molecular mimicry. *Virulence* **4**, 433–434 (2013).
- Chen, J. W. et al. Positive and negative selection shape the human naive B cell repertoire. *J. Clin. Invest.* **132**, e150985 (2022).
- Palmer, E. Negative selection-clearing out the bad apples from the T cell repertoire. *Nat. Rev. Immunol.* **3**, 383–391 (2003).
- Cusick, M. F., Libbey, J. E. & Fujinami, R. S. Molecular mimicry as a mechanism of autoimmune disease. *Clin. Rev. Allergy Immunol.* **42**, 102–111 (2012).
- Smatti, M. K. et al. Viruses and autoimmunity: a review on the potential interaction and molecular mechanisms. *Viruses* **11**, 762 (2019).
- Zhao, Z.-S., Granucci, F., Yeh, L., Schaffer, P. A. & Cantor, H. Molecular mimicry by herpes simplex virus-type 1: autoimmune disease after viral infection. *Science* **279**, 1344–1347 (1998).
- Sabbatini, A., Bombardieri, S. & Migliorini, P. Autoantibodies from patients with systemic lupus erythematosus bind a shared sequence of SmD and Epstein-Barr virus-encoded nuclear antigen EBNA 1. *Eur. J. Immunol.* **23**, 1146–1152 (1993).
- Lasso, G., Honig, B. & Shapira, S. D. A sweep of Earth's virome reveals host-guided viral protein structural mimicry and points to determinants of human disease. *Cell Syst.* **12**, 82–91.e3 (2021).
- Chang, S. T., Ghosh, D., Kirschner, D. E. & Linderman, J. J. Peptide length-based prediction of peptide–MHC class II binding. *Bioinformatics* **22**, 2761–2767 (2006).
- Wieczorek, M. et al. Major histocompatibility complex (MHC) class I and MHC class II proteins: conformational plasticity in antigen presentation. *Front. Immunol.* **8**, 292 (2017).
- Trolle, T. et al. The length distribution of class I-restricted T cell epitopes is determined by both peptide supply and MHC allele-specific binding preference. *J. Immunol.* **196**, 1480–1487 (2016).
- Burrows, S. R., Rossjohn, J. & McCluskey, J. Have we cut ourselves too short in mapping CTL epitopes? *Trends Immunol.* **27**, 11–16 (2006).
- Buus, S. et al. High-resolution mapping of linear antibody epitopes using ultrahigh-density peptide microarrays. *Mol. Cell Proteom.* **11**, 1790–1800 (2012).
- Qi, H. et al. Antibody binding epitope mapping (AbMap) of hundred antibodies in a single run. *Mol. Cell Proteom.* **20**, 100059 (2021).
- Doxey, A. C. & McConkey, B. J. Prediction of molecular mimicry candidates in human pathogenic bacteria. *Virulence* **4**, 453–466 (2013).
- Lebeau, G. et al. Zika E glycan loop region and Guillain–Barré syndrome-related proteins: a possible molecular mimicry to be taken in account for vaccine development. *Vaccines* **9**, 283 (2021).
- Adiguzel, Y. Molecular mimicry between SARS-CoV-2 and human proteins. *Autoimmun. Rev.* **20**, 102791 (2021).
- Begum, S. et al. Molecular mimicry analyses unveiled the human herpes simplex and poxvirus epitopes as possible candidates to incite autoimmunity. *Pathogens* **11**, 1362 (2022).
- Hurford, A. & Day, T. Immune evasion and the evolution of molecular mimicry in parasites. *Evolution* **67**, 2889–2904 (2013).
- Kumar, M. et al. ELM—the Eukaryotic Linear Motif Resource—2024 update. *Nucleic Acids Res.* **52**, D442–D455 (2024).
- Slobedman, B., Barry, P. A., Spencer, J. V., Avdic, S. & Abendroth, A. Virus-encoded homologs of cellular interleukin-10 and their control of host immune function. *J. Virol.* **83**, 9618–9629 (2009).
- Kanehisa, M. & Goto, S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28**, 27–30 (2000).
- Kang, M.-S. & Kieff, E. Epstein–Barr virus latent genes. *Exp. Mol. Med.* **47**, e131 (2015).
- Xing, Y. & Hogquist, K. A. T cell tolerance: central and peripheral. *Cold Spring Harb. Perspect. Biol.* **4**, a006957 (2012).
- Gabrielsen, I. S. M. et al. Transcriptomes of antigen presenting cells in human thymus. *PLoS ONE* **14**, e0218858 (2019).
- Bjornevik, K. et al. Longitudinal analysis reveals high prevalence of Epstein-Barr virus associated with multiple sclerosis. *Science* **375**, 296–301 (2022).
- Zamecnik, C. R. et al. An autoantibody signature predictive for multiple sclerosis. *Nat. Med.* 1–9 <https://doi.org/10.1038/s41591-024-02938-3> (2024).
- Hrycek, A., Kuśmierz, D., Mazurek, U. & Wilczok, T. Human cytomegalovirus in patients with systemic lupus erythematosus. *Autoimmunity* **38**, 487–491 (2005).
- Israeli, E., Agmon-Levin, N., Blank, M., Chapman, J. & Shoenfeld, Y. Guillain–Barré syndrome—a classical autoimmune disease triggered by infection or vaccination. *Clin. Rev. Allergy Immunol.* **42**, 121–130 (2012).
- Oikarinen, M. et al. Enterovirus infections are associated with the development of celiac disease in a birth cohort study. *Front. Immunol.* **11**, 604529 (2020).

32. Yazdanpanah, N. & Rezaei, N. Autoimmune complications of COVID-19. *J. Med. Virol.* **94**, 54–62 (2022).
33. Gómez-Rial, J., Rivero-Calle, I., Salas, A. & Martín-Torres, F. Rotavirus and autoimmunity. *J. Infect.* **81**, 183–189 (2020).
34. Banko, A., Miljanovic, D. & Cirkovic, A. Systematic review with meta-analysis of active herpesvirus infections in patients with COVID-19: old players on the new field. *Int. J. Infect. Dis.* **130**, 108–125 (2023).
35. Boys, I. N., Johnson, A. G., Quinlan, M. R., Kranzusch, P. J. & Elde, N. C. Structural homology screens reveal host-derived poxvirus protein families impacting inflammasome activity. *Cell Rep.* **42**, 112878 (2023).
36. Albarnaz, J. D. et al. Molecular mimicry of NF- κ B by vaccinia virus protein enables selective inhibition of antiviral responses. *Nat. Microbiol.* **7**, 154–168 (2022).
37. Howard, J., Justus, D. E., Totmenin, A. V., Shchelkunov, S. & Kotwal, G. J. Molecular mimicry of the inflammation modulatory proteins (IMPs) of poxviruses: evasion of the inflammatory response to preserve viral habitat. *J. Leukoc. Biol.* **64**, 68–71 (1998).
38. Gubser, C. et al. A new inhibitor of apoptosis from vaccinia virus and eukaryotes. *PLoS Pathog.* **3**, e17 (2007).
39. Mansur, D. S. et al. Poxvirus targeting of E3 ligase β -TrCP by molecular mimicry: a mechanism to inhibit NF- κ B activation and promote immune evasion and virulence. *PLoS Pathog.* **9**, e1003183 (2013).
40. Baxby, D. Poxviruses. in *Medical Microbiology* (ed. Baron, S.) (University of Texas Medical Branch at Galveston, 1996).
41. Efridi, W. & Lappin, S. L. Poxviruses. In *StatPearls* (StatPearls Publishing, 2024).
42. Amrani, A. et al. Expansion of the antigenic repertoire of a single T cell receptor upon T cell activation. *J. Immunol.* **167**, 655–666 (2001).
43. Zamvil, S. S., Spencer, C. M., Baranzini, S. E. & Cree, B. A. C. The gut microbiome in neuromyelitis optica. *Neurotherapeutics* **15**, 92–101 (2018).
44. Lanz, T. V. et al. Clonally expanded B cells in multiple sclerosis bind EBV EBNA1 and GlatCAM. *Nature* **603**, 321–327 (2022).
45. Rojas, J. M., Avia, M., Martín, V. & Sevilla, N. IL-10: a multifunctional cytokine in viral infections. *J. Immunol. Res.* **2017**, 6104054 (2017).
46. Grabski, D. F., Hu, Y., Sharma, M. & Rasmussen, S. K. Close to the bedside: a systematic review of endogenous retrovirus and their impact in oncology. *J. Surg. Res.* **240**, 145–155 (2019).
47. Balada, E., Vilardell-Tarrés, M. & Ordi-Ros, J. Implication of human endogenous retroviruses in the development of autoimmune diseases. *Int. Rev. Immunol.* **29**, 351–370 (2010).
48. Passos, V., Pires, A. R., Foxall, R. B., Nunes-Cabaço, H. & Sousa, A. E. Expression of human endogenous retroviruses in the human thymus along T cell development. *Front. Virol.* **2**, (2022).
49. Alcazer, V., Bonaventura, P. & Depil, S. Human endogenous retroviruses (HERVs): shaping the innate immune response in cancers. *Cancers* **12**, 610 (2020).
50. Khalid, F. et al. Neurological adverse effects of immune checkpoint inhibitors and chimeric antigen receptor T cell therapy. *World J. Oncol.* **14**, 109–118 (2023).
51. Marini, A. et al. Neurologic adverse events of immune checkpoint inhibitors: a systematic review. *Neurology* **96**, 754–766 (2021).
52. Gritsch, D. & Valencia-Sanchez, C. Drug-related immune-mediated myelopathies. *Front. Neurol.* **13**, 1003270 (2022).
53. Hottinger, A. F. Neurologic complications of immune checkpoint inhibitors. *Curr. Opin. Neurol.* **29**, 806–812 (2016).
54. Oliveira, M. C. B., de Brito, M. H. & Simabukuro, M. M. Central nervous system demyelination associated with immune checkpoint inhibitors: review of the literature. *Front. Neurol.* **11**, 538695 (2020).
55. Ramos-Casals, M. et al. Immune-related adverse events of checkpoint inhibitors. *Nat. Rev. Dis. Prim.* **6**, 38 (2020).
56. Leko, V. et al. Identification of neoantigen-reactive T lymphocytes in the peripheral blood of a patient with glioblastoma. *J. Immunother. Cancer* **9**, e002882 (2021).
57. Zhou, W., Yu, J., Li, Y. & Wang, K. Neoantigen-specific TCR-T cell-based immunotherapy for acute myeloid leukemia. *Exp. Hematol. Oncol.* **11**, 100 (2022).
58. Zhu, Y., Qian, Y., Li, Z., Li, Y. & Li, B. Neoantigen-reactive T cell: an emerging role in adoptive cellular immunotherapy. *MedComm* **2**, 207–220 (2021). (2020).
59. Grosser, R., Cherkassky, L., Chintala, N. & Adusumilli, P. S. Combination immunotherapy with CAR T cells and checkpoint blockade for the treatment of solid tumors. *Cancer Cell* **36**, 471–482 (2019).
60. Yoon, D. H., Osborn, M. J., Tolar, J. & Kim, C. J. Incorporation of immune checkpoint blockade into chimeric antigen receptor T cells (CAR-Ts): combination or built-in CAR-T. *Int. J. Mol. Sci.* **19**, 340 (2018).
61. Park, J. & Chung, Y.-J. Identification of neoantigens derived from alternative splicing and RNA modification. *Genomics Inf.* **17**, e23 (2019).
62. Huang, P., Wen, F., Tuerhong, N., Yang, Y. & Li, Q. Neoantigens in cancer immunotherapy: focusing on alternative splicing. *Front. Immunol.* **15**, 1437774 (2024).
63. Nantes University Hospital. An open single-center, phase I proof of concept trial to assess the safety and feasibility of adoptive cell therapy with autologous EBV-specific cytotoxic T lymphocytes (CTL) in patients with a first clinical episode highly suggestive of multiple sclerosis. <https://clinicaltrials.gov/study/NCT02912897> (2023).
64. Michael, L. Effects of antiviral therapies on Epstein-Barr virus replication. <https://clinicaltrials.gov/study/NCT05957913> (2023).
65. Masson, P. et al. ViralZone: recent updates to the virus knowledge resource. *Nucleic Acids Res.* **41**, D579–D583 (2013).
66. Schoch, C. L. et al. NCBI Taxonomy: a comprehensive update on curation, resources and tools. *Database* **2020**, ba0062 (2020).
67. Uhlén, M. et al. Proteomics. Tissue-based map of the human proteome. *Science* **347**, 1260419 (2015).
68. Bezanson, J., Edelman, A., Karpinski, S. & Shah, V. B. Julia: a fresh approach to numerical computing. *SIAM Rev.* **59**, 65–98 (2017).
69. R Core Team. R: A Language and Environment for Statistical Computing (R Foundation for Statistical Computing, 2021).
70. Maguire, C. et al. Molecular mimicry as a mechanism of viral immune evasion and autoimmunity. Zenodo <https://doi.org/10.5281/zenodo.13272863> (2024).
71. Consortium, UniProt UniProt: the Universal Protein Knowledgebase in 2023. *Nucleic Acids Res.* **51**, D523–D531 (2023).

Acknowledgements

This work was supported by NIH R01AI04870-S1 (E.M.), NIAAA K08 T26-1616-11 (E.M.), NIDA 5T32DA018926-18 (C.M.), and institutional Dell Medical School Startup funding (E.M.). Computational servers used in this study were donated by Advance Microdevices. Funding sources did not have a direct role in design, analysis, or approval of this manuscript. We appreciate discussions with John Moore, Dr. Victoria Chu, Dr. Lauren Ehrlich, Dr. Laura Fonken, Dr. Dayne Mayfield, and Dr. Hans Hofmann. We thank Dr. Lauren Ehrlich for allowing access to the Advanced Micro Devices donated computational servers used in this study. We are also grateful for the administrative and technical support from Dell Medical School Neurology Department and the UT Austin Biomedical Research Computing Facility.

Author contributions

C.M. and E.M. conceived this work. D.W. and C.M. conceived the methodology and software. C.M., C.W., A.R., C.F., B.M., and N.L. undertook the formal analysis. C.W., A.R., C.F., B.M., and N.L. obtained

resources and data curation. C.M. was responsible for visualization. E.M. obtained funding. C.M. conducted the project administration. E.M. and D.W. were responsible for supervision. C.M. wrote the original draft. C.M., C.W., A.R., C.F., B.M., N.L., D.W., and E.M. reviewed and edited the final manuscript.

Competing interests

E.M. has received research funding from Babson Diagnostics, honorarium from Multiple Sclerosis Association of America and has served on advisory boards of Genentech, Horizon, Teva and Viela Bio. The other authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-024-53658-8>.

Correspondence and requests for materials should be addressed to Esther Melamed.

Peer review information *Nature Communications* thanks Yorgo Modis and the other anonymous reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2024