Endogenous retroviral signatures predict immunotherapy response in clear cell renal cell carcinoma

Christof C. Smith,[1,2] Kathryn E. Beckermann,[3] Dante S. Bortone,[2,4] Aguirre A. De Cubas,[3] Lisa M. Bixby,[2] Samuel J. Lee,[1,2] Anshuman Panda,[5] Shridar Ganesan,[5] Gyan Bhanot,[5] Eric M. Wallen,[2,6] Matthew I. Milowsky,[2,7] William Y. Kim,[2,6,7,8] W. Kimryn Rathmell,[3] Ronald Swanstrom,[2,9] Joel S. Parker,[2,4,8] Jonathan S. Serody,[1,2,7] Sara R. Selitsky,[2,4] and Benjamin G. Vincent[1,2,7,10]

[1]Department of Microbiology and Immunology, UNC School of Medicine, Chapel Hill, North Carolina, USA. [2]Lineberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA. [3]Division of Hematology and Oncology, Vanderbilt University Medical Center, Nashville, Tennessee, USA. [4]Lineberger Bioinformatics Group, Lineberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA. [5]Rutgers Cancer Institute of New Jersey, New Brunswick, New Jersey, USA; and Department of Physics, Rutgers University, Piscataway, New Jersey, USA. [6]Department of Urology, [7]Division of Hematology/Oncology, Department of Medicine, [8]Department of Genetics, [9]Department of Biochemistry and Biophysics, and [10]Curriculum in Bioinformatics and Computational Biology, UNC School of Medicine, Chapel Hill, North Carolina, USA.

Human endogenous retroviruses (hERVs) are remnants of exogenous retroviruses that have integrated into the genome throughout evolution. We developed a computational workflow, hervQuant, which identified more than 3,000 transcriptionally active hERVs within The Cancer Genome Atlas (TCGA) pan-cancer RNA-Seq database. hERV expression was associated with clinical prognosis in several tumor types, most significantly clear cell renal cell carcinoma (ccRCC). We explored two mechanisms by which hERV expression may influence the tumor immune microenvironment in ccRCC: (i) RIG-I–like signaling and (ii) retroviral antigen activation of adaptive immunity. We demonstrated the ability of hERV signatures associated with these immune mechanisms to predict patient survival in ccRCC, independent of clinical staging and molecular subtyping. We identified potential tumor-specific hERV epitopes with evidence of translational activity through the use of a ccRCC ribosome profiling (Ribo-Seq) dataset, validated their ability to bind HLA in vitro, and identified the presence of MHC tetramer–positive T cells against predicted epitopes. hERV sequences identified through this screening approach were significantly more highly expressed in ccRCC tumors responsive to treatment with programmed death receptor 1 (PD-1) inhibition. hervQuant provides insights into the role of hERVs within the tumor immune microenvironment, as well as evidence that hERV expression could serve as a biomarker for patient prognosis and response to immunotherapy.

Introduction

Human endogenous retroviruses (hERVs) are remnants of exogenous retroviruses integrated into the primate genome over evolutionary time (1). hERVs share genomic similarities to other retroviruses, including the presence of functional and remnant 5′ and 3′ long terminal repeats (LTRs), and *gag*, *pro*, *pol*, and *env* genes. Subsets of recently integrated hERVs still maintain limited translation under physiological and pathological conditions (2–6), including evidence for modulation of melanoma, lymphomas, leukemias, and ovarian, breast, prostate, urothelial, and renal carcinomas (5, 7–14). Although studies have identified the role of specific hERVs in the pathogenesis and progression of these cancers, to date there have been a limited number of pan-cancer studies elucidating the landscape and impact of hERV expression. A recent study by Rooney et al. analyzed features associated with genes important for immune cytolytic activity, finding that one of these associated features was expression of a small subset of hERVs (15). While this study provided evidence that hERV expression associated with an immune phenotype, the exploration of hERVs was limited by a small reference set, no reported mechanism of association or prognostic impact of hERV expression, and no confirmation of a hERV-specific immune population within any tumor type. Thus, the role of hERVs in modulating the tumor immune microenvironment remains largely unexplored, predominately due to a lack of tools for identification of full-length, intact hERVs from sequencing data. To fully understand the role of hERVs in antitumor immunity, a more comprehensive database containing greater numbers of individual full-length hERVs is required. Understanding patterns of hERV expression will allow for greater knowledge of the impact of hERVs on tumor-immune interactions, the design of new prognostic models based on hERV signatures, and further identification of tumor-specific hERV epitopes for targeted tumor vaccinations.

Currently, a limited repertoire of tools are available for hERV quantification. There exist several databases of hERV elements, including HERVd, which contains hERV-like elements, and their genomic locations that have been used for analysis of RNA-Seq data (16–18). Additionally, there are several tools for identification of intra- and intergenic hERV-like elements (19), related transposable elements (20), and interspersed repeats (RepeatMasker) among human transcripts (21). While these resources provide

Authorship note: BGV, SRS, and JSS contributed equally to this work.
Conflict of interest: WYK is the inventor on the BASE47 gene classifier (US patent application WO2015073949A1).

Submitted: April 6, 2018; **Accepted:** August 10, 2018.
Reference information: *J Clin Invest*. 2018;128(11):4804–4820.
https://doi.org/10.1172/JCI121476.

methods to quantify expression of hERV-like elements among transcripts, they do not provide quantification based on an intact, full-length hERV proviral reference. This capability to distinguish and quantify individual hERVs provides a useful tool to classify hERVs into distinct groups based on biological associations in various cancers.

Recently, Vargiu et al. compiled a database of 3,173 intact, full-length hERV sequences and developed a comprehensive method for classifying these sequences into 11 superfamilies (Supplemental Table 1; supplemental material available online with this article; https://doi.org/10.1172/JCI121476DS1) (3). Using this database as a reference, we designed a computational workflow for identifying the expression of specific hERVs from RNA sequencing (RNA-Seq), hervQuant, and quantified hERV expression within the Cancer Genome Atlas (TCGA) pan-cancer dataset. We assessed interactions of specific hERVs with immune and clinical features. Among all cancer types encompassed within the pan-cancer dataset, clear cell renal cell carcinoma (ccRCC, designated by TCGA as KIRC) contained the greatest number of prognostic hERVs. Thus, we explored two mechanisms by which hERV expression may influence the tumor immune microenvironment in ccRCC: (i) activation of RIG-I–like pathway signaling and (ii) hERV epitope-triggered T and B cell activation. Using biological classes of hERV signatures derived from these two mechanisms, we further demonstrated the ability of hERV expression to predict patient survival in a multivariate regression model, independent of traditional clinical staging and molecular subtyping. Last, we used a publicly available ccRCC ribosome profiling (Ribo-Seq) dataset (22) to screen for translation of tumor-specific hERV epitopes, validated their capacity to bind HLA in vitro, and demonstrated the presence of tetramer-positive epitope-specific T cells within ccRCC tumors. We found tumor-specific hERV expression to be associated with clinical response to PD-1 axis inhibition in ccRCC patients, suggesting that hERV expression may provide a biomarker for immunotherapy responsiveness and hERV viral proteins may provide targetable, tumor-specific epitopes. The information gained from hERV expression profiling gives new insight into the role of hERVs within tumor-immune microenvironment interactions and provides evidence for hERV expression–based molecular models for patient prognosis and responsiveness to immunotherapy.

Results

Expression and association of hERVs in TCGA pan-cancer. TCGA pan-cancer hERV expression was determined using hervQuant, described in detail in the Supplemental Notes (Figure 1A and Supplemental Figures 1 and 2). For consistency, only samples sequenced by Illumina NextSeq at 2 × 50 bp were analyzed, resulting in complete removal of ESCA, GBM, OV, and STAD and partial removal of COAD, UCEC, and READ subtypes (see Supplemental Table 2 for tumor abbreviations). All 3,173 reference hERVs were expressed in at least one sample, encompassing all 11 superfamilies and 3 lineages (Supplemental Table 1). Relative hERV expression patterns were strikingly homogenous across all cancer types (Figure 1B and Supplemental Figure 3). Among all cancer types, TGCT demonstrated the greatest mean and median hERV expression, while LIHC, ACC, and UVM ranked last (Supplemental Figure 4). To identify similar hERV expression patterns across models, we calculated the Euclidean distance of mean hERV expression between each cancer type (Figure 1B and Supplemental Figure 5). Tumor types with lowest overall hERV expression (LIHC, ACC, UVM) were closely related by unsupervised clustering and shared very low similarity with all other tumor types. Two large clusters comprised 10 (PCPG, SKCM, CHOL, SARC, THYM, DLBC, PRAD, THCA, KICH, and LGG) and 8 (LUAD, PAAD, BLCA, CESC, MESO, UCEC, UCS, and BRCA) cancer types. While several cancer types demonstrated similar hERV expression patterns based on tissue location (UCEC and UCS, HNSC and LUSC, KIRC and KIRP, and READ and COAD), the clustering observed between various tumor types suggests that hERV expression may be conserved among cancers across a variety of tissues. Notably, two tumor types with immune-privileged tissues of origin (TGCT and UVM) demonstrated lower similarities to all other cancers. Lack of immune interactions within these native tissues may potentially result in unique hERV expression profiles in these tumors, suggesting that shared hERV expression profiles within other tumor types may be shaped by the presence of related tumor immune responses.

Overexpression of specific hERVs within tumors has been attributed to epigenetic demethylation of genes associated with provirus expression, which can be triggered through the use of epigenetic modulatory agents (14, 23–27). hERV expression was highly associated with Illumina Methylation450K-derived methylation patterns, with the majority of hERVs significantly associated with demethylation (2,639 hERVs with generalized linear model [GLM] FDR-corrected $P \leq 0.05$; 2,205 with coefficient <0; 434 with coefficient >0; Figure 1C).

We next examined the association between hERV expression and immune features, age, and survival among tumor types. We first performed multivariable linear regression of hERV expression by cancer type with 46 immune gene signatures (IGS) previously described in the literature (28–33) (Figure 1D and Supplemental Figure 6). A small population of hERVs demonstrated near ubiquitous positive or negative association with all IGS, with the majority of hERVs showing a split association pattern. Included among IGS that demonstrated positive association with the majority of significant hERVs (GLM FDR-corrected $P < 0.05$) were those associated with immune cells known to have antitumor effector function, including effector and central memory T cells and NK cells. Additionally, a signature of anti–PD-1 (aPD1) responsiveness (IPRES_aPD1_responder) was positively associated with hERV expression in 79.2% (1,472 of 1,858) of significantly associated hERVs, while a signature for nonresponder tumor biopsies (IPRES_aPD1_nonresponder) was negatively associated with all hERV expression in 83.0% (1,679 of 2,024) of significantly associated hERVs (34). We next examined the association between hERV expression and age, controlling for tumor type, and observed that the majority of significantly associated hERVs demonstrated negative association between expression and patient age (GLM FDR-corrected $P < 0.05$; 150 with coefficient <0; 13 with coefficient >0; Supplemental Figure 7). To elucidate whether hERV expression associated with clinical outcome, we performed Cox's proportional hazard regression (CoxPH) for hERV expression across all cancer types. Association of survival with mean hERV expression identified 3 tumor types with prognostic mean hERV expression (KICH, COAD, and KIRC). In all 3 tumor types, mean

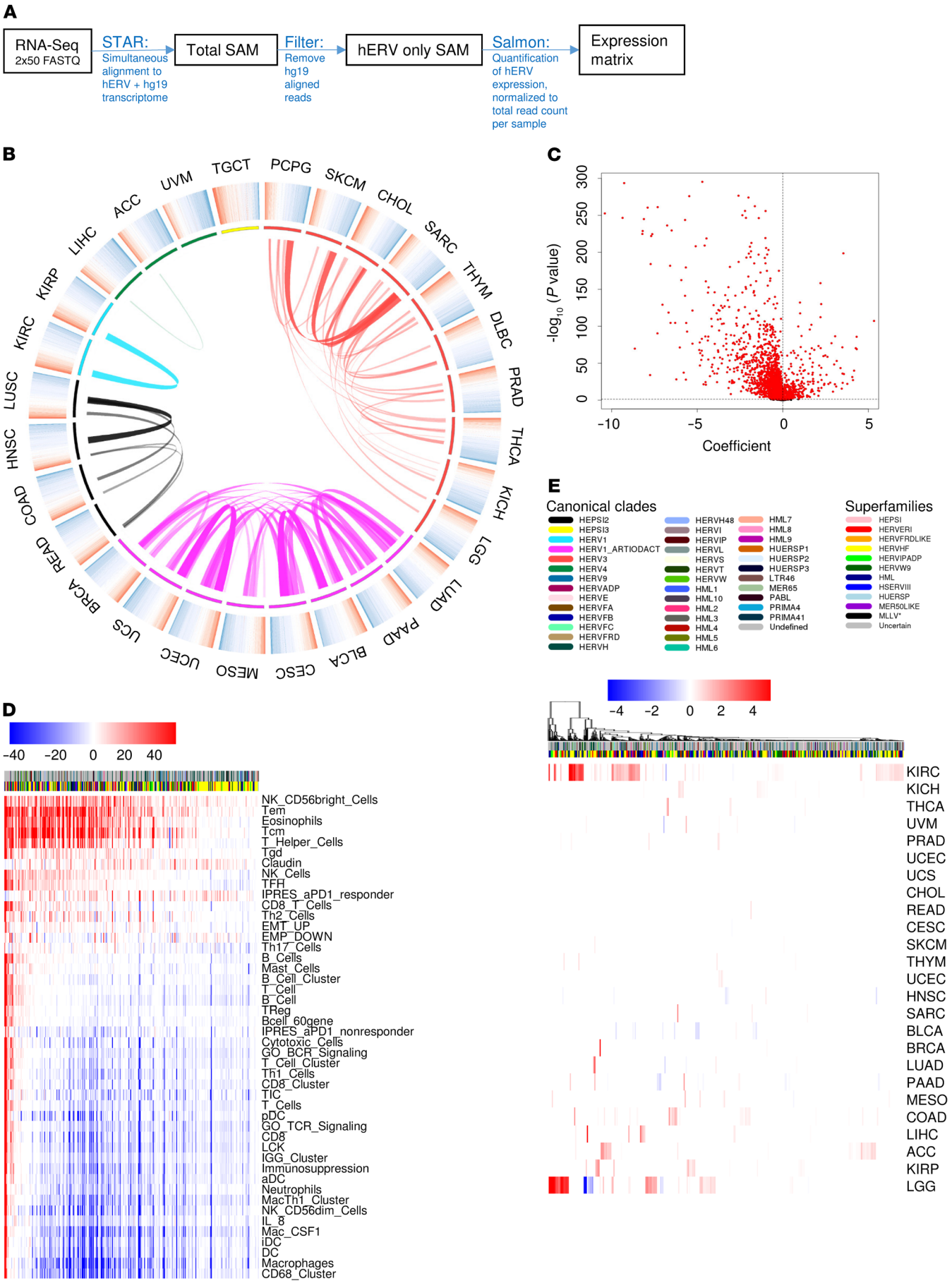
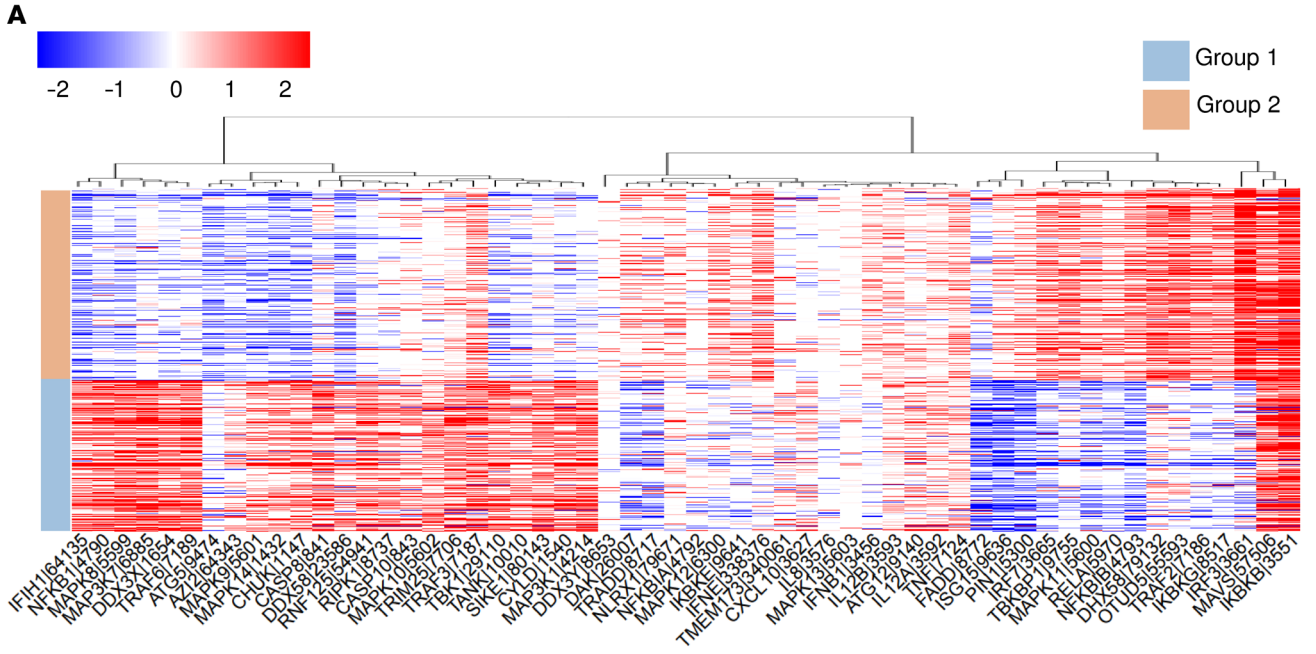

A
RNA-Seq
2x50 FASTQ
STAR:
Simultaneous alignment to hERV + hg19 transcriptome
Total SAM
Filter:
Remove hg19 aligned reads
hERV only SAM
Salmon:
Quantification of hERV expression, normalized to total read count per sample
Expression matrix
B
TGCT PCPG SKCM CHOL SARC THYM DLBC PRAD THCA KICH LGG LUAD PAAD BLCA CESC MESO UCEC UCS BRCA READ COAD HNSC LUSC KIRC KIRP LIHC ACC UVM
C
-log10 (P value)
0 50 100 150 200 250 300
-10 -5 0 5
Coefficient
D
-40 -20 0 20 40
NK_CD56bright_Cells
Tem
Eosinophils
Tcm
T_Helper_Cells
Tgd
Claudin
NK_Cells
TFH
IPRES_aPD1_responder
CD8_T_Cells
Th2_Cells
EMT_UP
EMP_DOWN
Th17_Cells
B_Cells
Mast_Cells
B_Cell_Cluster
T_Cell
B_Cell
TReg
Bcell_60gene
IPRES_aPD1_nonresponder
Cytotoxic_Cells
GO_BCR_Signaling
T_Cell_Cluster
Th1_Cells
CD8_Cluster
TIC
T_Cells
pDC
GO_TCR_Signaling
CD8
LCK
IGG_Cluster
Immunosuppression
aDC
Neutrophils
MacTh1_Cluster
NK_CD56dim_Cells
IL_8
Mac_CSF1
iDC
DC
Macrophages
CD68_Cluster
E
Canonical clades
HEPSI2
HEPSI3
HERV1
HERV1_ARTIODACT
HERV3
HERV4
HERV9
HERVADP
HERVE
HERVFA
HERVFB
HERVFC
HERVFRD
HERVH
HERVH48
HERVI
HERVIP
HERVL
HERVS
HERVT
HERVW
HML1
HML10
HML2
HML3
HML4
HML5
HML6
HML7
HML8
HML9
HUERSP1
HUERSP2
HUERSP3
LTR46
MER65
PABL
PRIMA4
PRIMA41
Undefined
Superfamilies
HEPSI
HERVERI
HERVFRDLIKE
HERVHF
HERVIPADP
HERVW9
HML
HSERVIII
HUERSP
MER50LIKE
MLLV*
Uncertain
-4 -2 0 2 4
KIRC
KICH
THCA
UVM
PRAD
UCEC
UCS
CHOL
READ
CESC
SKCM
THYM
UCEC
HNSC
SARC
BLCA
BRCA
LUAD
PAAD
MESO
COAD
LIHC
ACC
KIRP
LGG

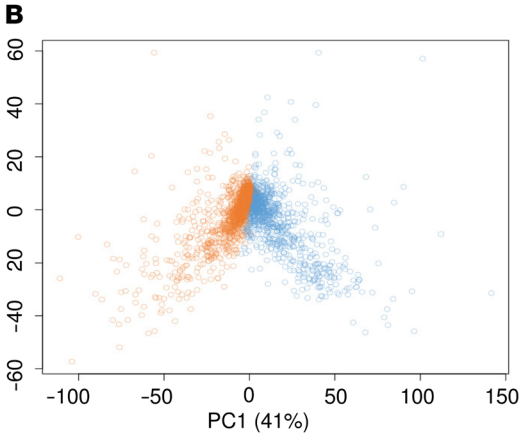
Figure 1. Human endogenous retrovirus expression and association in TCGA pan-cancer dataset. (**A**) Schematic of the hervQuant workflow. (**B**) hERV expression displayed by heatmaps in the outermost layer, ranked by mean expression across the pan-cancer dataset. Tumor groups shown in the middle ring, with colors representing clusters determined from a cut-tree (height = 140) of hierarchical clustering of Euclidean distance of mean hERV expression between each cancer type. Innermost lines represent hERV expression pairwise Euclidean distance ≤40 between tumor types. Opacity and width of inner lines increase with greater similarity. (**C**) Volcano plot of association (GLM) between read-normalized hERV expression and the mean of the methylation β coefficient, with GLM coefficient along the *x* axis and $-\log_{10}$ FDR-corrected *P* value along the *y* axis. (**D** and **E**) Association (GLM) between read-normalized hERV expression and (**D**) IGS expression and (**E**) survival among TCGA pan-cancer dataset. FDR- (**D**) or Bonferroni-corrected (**E**) *P* represented by intensity of color and direction of coefficient represented by color (red, positive; blue, negative). Color bar displays hERV superfamily and canonical clade classifications. (**D**) Rows and columns are ordered by number of significantly positive associations. (**E**) Survival analysis filtered by hERVs and tumor types with at least 1 significant comparison. See Supplemental Table 2 for number of samples per TCGA cancer cohort.

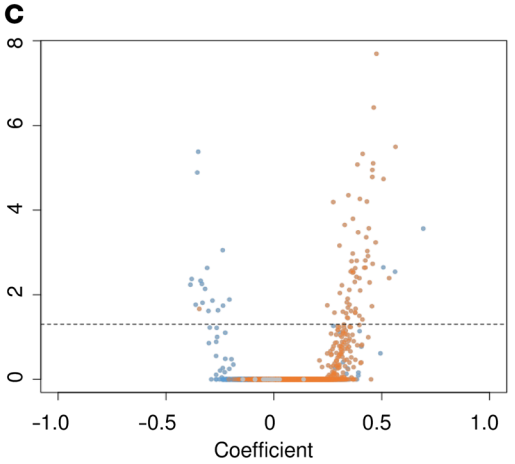
hERV expression was negatively prognostic (Supplemental Figure 8). Additionally, we examined Kaplan-Meier survival curves for each TCGA cancer type split by upper versus lower 50th percentile mean hERV expression, and observed 5 cancer types with significant separation of survival curves (Supplemental Figure 9; BLCA, COAD, KICH, KIRC, and PCPG; log-ranked $P < 0.05$). Among these 5 cancer types, KIRC was the most associated with survival. All cancer types except BLCA demonstrated shorter survival in patients with greater mean hERV expression. To perform a more detailed analysis, we associated survival with expression of each individual hERV (Figure 1E and Supplemental Figure 10). TCGA KIRC (ccRCC), a tumor type in which several hERVs have been shown to be actively translated (14, 35–37), constituted 25.1% of all significantly prognostic hERVs, with over 1.5× more significant hERVs than the next highest cancer, LGG (KIRC: 362; LGG: 230; Figure 1E). To elucidate the immune mechanisms behind this enrichment of prognostic hERVs in ccRCC, we focused on this cancer type for the remainder of our analyses.

hERV expression in ccRCC demonstrates evidence of immune stimulation through RIG-I–like signaling. Several groups have demonstrated that activation of select endogenous retroviral elements can trigger signaling through innate immune sensors, including double-stranded RNAs (dsRNA) that subsequently signal through cytosolic RIG-I–like receptors (26, 27). To elucidate a more comprehensive role for hERVs in the RIG-I–like pathway in ccRCC, we studied the association between hERV expression and genes in the RIG-I–like receptor signature (Molecular Signatures Database) (38), observing marked separation of genes into 2 groups by hierarchical clustering (Figure 2A). We defined 2 hERV groups (1 and 2; Supplemental Table 3) based on the ratio between each hERV's mean linear regression coefficients within each gene cluster (>1 or < 1) and validated their definitions using principal component analysis (Figure 2B). While both groups demonstrated significant positive association between hERV expression and genes that activate the RIG-I-like pathway, group 2 hERVs demonstrated a significant positive association with several key antagonist genes downstream of NF-κB signaling (most notably NFKBIB), along with a significant negative association to key agonistic genes in NF-κB signaling (e.g. TBK1, TANK, and AZI2). CoxPH of hERV expression within TCGA KIRC provided further evidence that these groups are biologically distinct, with the majority of group 1 and 2 hERVs providing association with longer and shorter overall survival, respectively (Figure 2C). In addition, group 2 and non-prognostic group 1 hERVs (CoxPH Bonferroni-corrected $P > 0.05$) demonstrated a significant positive association with the majority of IGS (93%, 57%, and 60%, respectively), while prognostic group 1 hERVs (Bonferroni-corrected $P \leq 0.05$; majority associated with longer overall survival) largely demonstrated a negative association with IGS (33%), including those for T cells, B cells, dendritic cells, macrophages, and NK cells (Figure 2D and Supplemental Figure 11). Despite these negative association patterns with IGS observed in prognostic group 1 hERVs, TCGA KIRC samples with greater expression of these hERVs had decreased ratios of Treg to $CD8^+$ IGS (Treg IGS divided by the mean of 3 $CD8^+$ IGS) compared with any other hERV group, suggesting the immune infiltrate associated with prognostic group 1 hERVs was less immunosuppressive than that of non-prognostic group 1 and group 2 hERVs (Supplemental Figure 12). Additionally, prognostic group 1 hERVs demonstrated positive association with signatures for Th17 T cells, which have been associated with a more favorable prognosis in ccRCC (39). Overall, this analysis provided the first evidence to our knowledge for biologically distinct hERV groups that differentially interact with innate immune sensing, with differential downstream prognostic and immunological effects and prognostic associations.

hERV expression in ccRCC demonstrates evidence of B cell activation. In addition to innate immune sensor signaling, hERVs can trigger antitumor immunity through tumor-specific expression of viral epitopes. In cancer patients, high antibody titers have been known to develop against hERV proteins with specificity of expression within the tumor, with little else known regarding the role of this B cell response (40). To determine whether hERVs show evidence of an adaptive immune response in ccRCC, we identified T/B cell clonotype repertoires in TCGA KIRC using MiXCR and filtered on T/B cell receptors (TCRs/BCRs; defined as shared CDR3 amino acid sequence) observed in ≥10% of patients (41). These filtering criteria resulted in no shared TCR clonotypes, suggesting potentially low sensitivity of detection for MiXCR-derived TCR data in RNA-Seq data. In contrast, 437 shared BCRs were identified, of which 397 were significantly associated with expression of ≥1 hERV (Figure 3A, left). Within this pool, 4 clones had significant positive association with the expression of 1,207 hERVs, suggesting a potential hERV epitope–driven B cell response (Figure 3A, right, and Supplemental Table 3). Differential superfamily distribution patterns were observed between BCR-associated and non-BCR-associated hERVs, suggesting certain superfamilies may have a greater propensity for triggering B cell activation (HERVERI, HML, HSERVIII, and HERVW9; FDR-corrected χ^2 test $P \leq 0.05$; Supplemental Figure 13). Furthermore, multiple sequence alignment (Clustal Omega) of proviral sequences from these BCR-associated hERVs identified large regions of high sequence identity (Supplemental Figure 14). Filtering on sequence identity of ≥25% of all BCR-associated hERVs with a sequence length ≥21 base pairs (the approximate minimal length necessary for immunoglobulin CDR3 region specificity) (42), we observed 8


A
-2 -1 0 1 2
Group 1
Group 2


B
PC1 (41%)


C
Coefficient

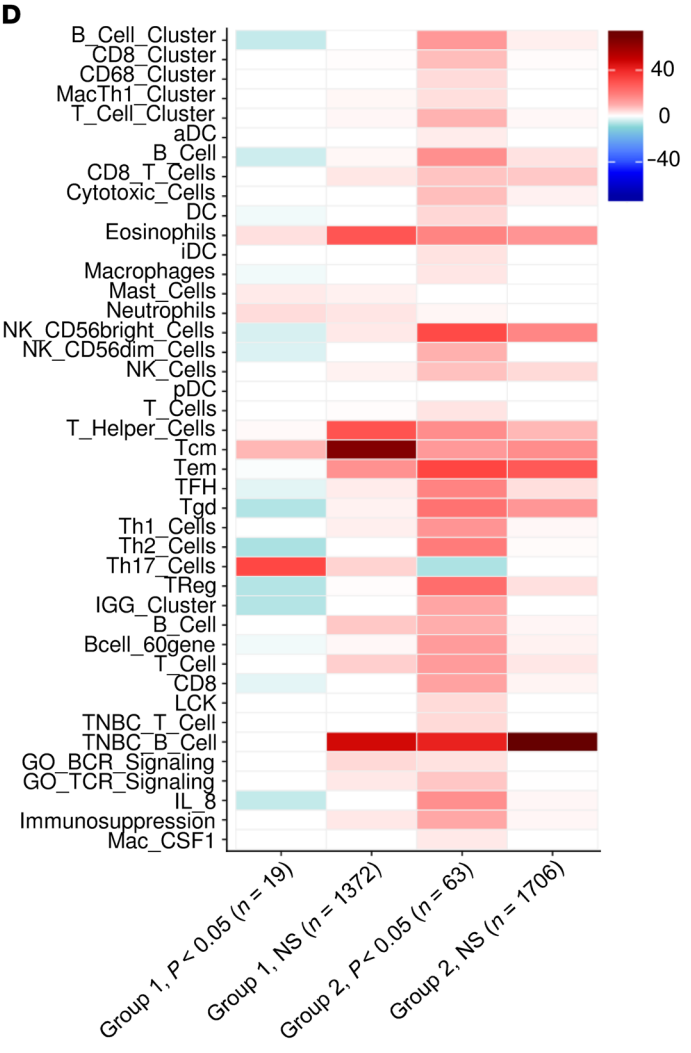
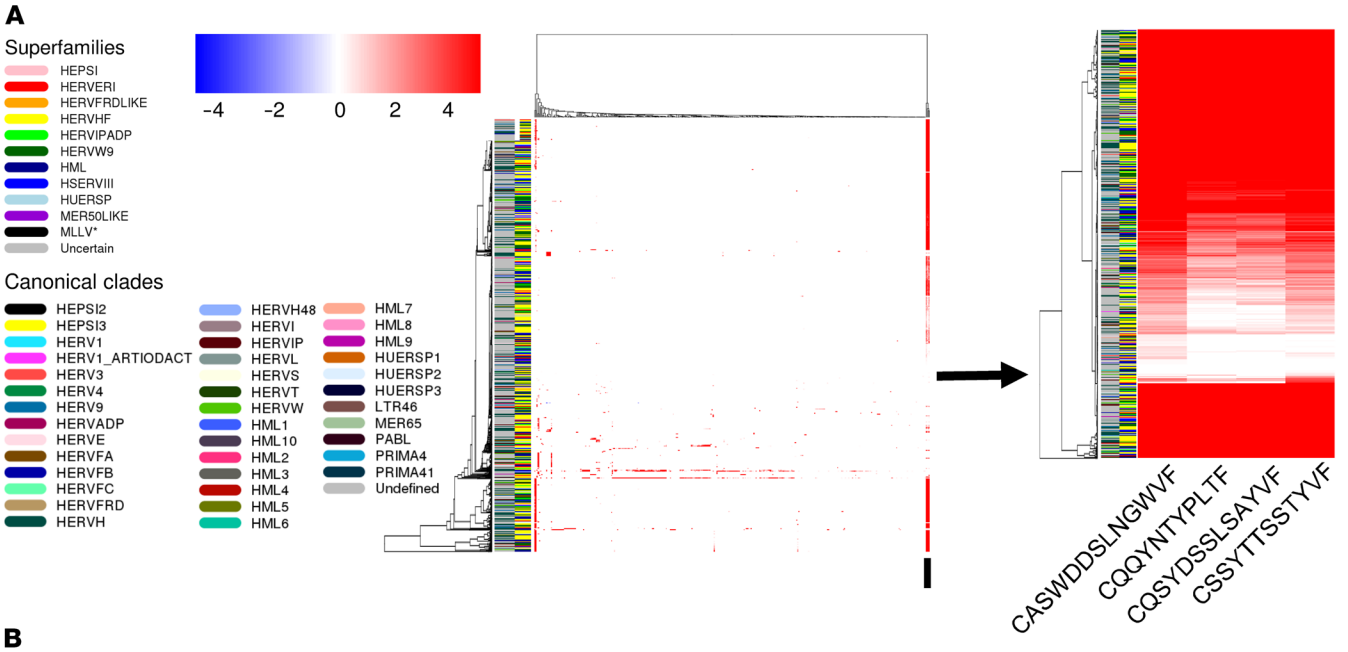

D
B_Cell_Cluster
CD8_Cluster
CD68_Cluster
MacTh1_Cluster
T_Cell_Cluster
aDC
B_Cell
CD8_T_Cells
Cytotoxic_Cells
DC
Eosinophils
iDC
Macrophages
Mast_Cells
Neutrophils
NK_CD56bright_Cells
NK_CD56dim_Cells
NK_Cells
pDC
T_Cells
T_Helper_Cells
Tcm
Tem
TFH
Tgd
Th1_Cells
Th2_Cells
Th17_Cells
TReg
IGG_Cluster
B_Cell
Bcell_60gene
T_Cell
CD8
LCK
TNBC_T_Cell
TNBC_B_Cell
GO_BCR_Signaling
GO_TCR_Signaling
IL_8
Immunosuppression
Mac_CSF1
Group 1, P < 0.05 (n = 19)
Group 1, NS (n = 1372)
Group 2, P < 0.05 (n = 63)
Group 2, NS (n = 1706)

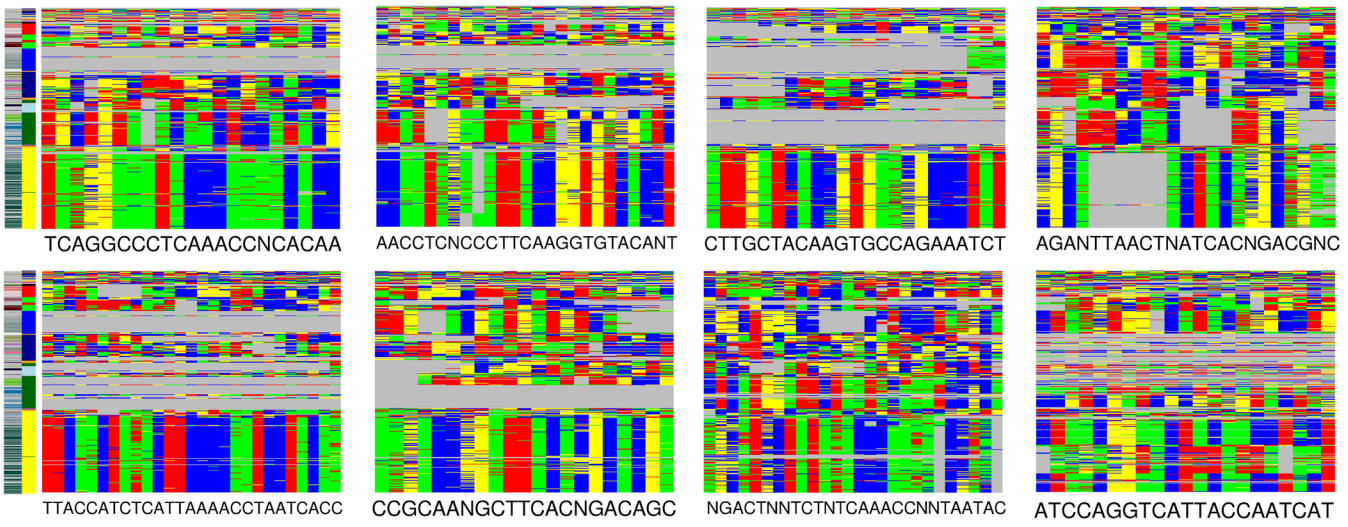
Figure 2. Mechanism of hERV-mediated RIG-I–like pathway signaling in ccRCC. (**A**) Heatmap of association (GLM) between hERV expression and RIG-I–like pathway–associated genes. FDR-corrected $-\log_{10}$(*P* value) represented by intensity of color, and direction of coefficient represented by color (red: positive, blue: negative). Group 1 (blue) and 2 (orange) hERVs are represented by color along the left-side color bar. (**B**) PC1 versus PC2 from PCA of association matrix in **A** between hERV expression and RIG-I–like pathway–associated genes from for group 1 and 2 hERVs. Percentage of variance for principal component 1 (PC1) and PC2 is shown in parentheses along each axis. (**C**) Volcano plot of CoxPH analysis of UQN hERV expression as a predictor of survival, with Bonferroni-corrected $-\log_{10}$(*P* value) displayed as a function of hazard ratio for each hERV. Dashed horizontal line represents FDR-corrected $P = 0.05$. (**B** and **C**) Groups 1 and 2, and other hERVs defined from **A** (group 1: blue; group 2: orange; neither: gray). (**D**) Heatmap of association (GLM) between expression of IGSs with group 1 and 2 hERV signatures (average expression), split by either significant or nonsignificant association with patient prognosis. FDR-corrected *P* values represented by intensity of color, and direction of coefficient represented by color (red, positive; blue, negative).

regions of conserved DNA similarity (Figure 3B). NIH Retrovirus Protein BLAST (https://www.ncbi.nlm.nih.gov/genome/viruses/retroviruses/) of these sequences showed similarity to known hERV *env* genes in 8 of 8 sequences, with additional similarity to other retroviral genes in 2 of 8 sequences. While suggestive of potentially targetable antigens within the hERV *env* region, CoxPH demonstrated significantly higher hazard ratios among BCR-associated compared with non-BCR-associated hERVs (Welch's *t* test $P = 2.4 \times 10^{-3}$; Figure 3C). Differential expression analysis (DESeq2) of BCR-associated hERVs demonstrated a balanced proportion of hERVs with both higher tumor–to–matched normal and matched normal–to–tumor expression (tumor: $n = 542$; matched: $n = 72$; Figure 3D), suggesting an overall lack of tumor specificity among BCR-associated hERVs.

hERV signatures of innate and adaptive immune activation provides prognostic value in ccRCC. Currently, clinical stage is the most robust prognostic variable for ccRCC. While molecular features such as M1–M4 molecular subtyping have been shown to be potentially prognostic, no molecular markers have been widely adapted for clinical decision making in ccRCC, making identification of a robust molecular marker for prognosis an appealing goal (43). Throughout this study, we identified pools of hERVs with evidence of both RIG-I–like–mediated innate immune activation and inhibition, as well as B cell–mediated adaptive immunity (Figure 4, A and B). To provide evidence that these classes can be used to generate a model of clinical outcome in ccRCC, we derived signatures corresponding to the mean expression of prognostic hERVs (CoxPH Bonferroni-corrected $P \leq 0.05$) within each class. According to log-rank test, Kaplan-Meier overall survival curves for patients within the upper versus lower 50th percentiles for each of the 3 signatures were significantly different (RIG-I–like upregulated [up]: $P = 4.5 \times 10^{-10}$; RIG-I–like downregulated [down]: $P = 6.3 \times 10^{-14}$; BCR-associated: $P = 1.1 \times 10^{-5}$; Figure 4C). Patients with both higher expression of RIG-I–like down and BCR-associated signatures had significantly shorter overall survival, while those with higher expression of the RIG-I–like up signature had longer overall survival. Recent analyses also provided metrics for disease-specific survival (DSS) and progression-free interval (PFI) in TCGA KIRC, additionally with an underpowered reporting of disease-free interval (DFI) (44). Of these metrics, DSS and PFI trended similarly to curves observed with overall survival, providing further evidence that these hERV signatures are specifically associated with disease burden (Supplemental Figure 15). We performed multivariable CoxPH modeling with clinical stage and with or without molecular subtype (M1–M4) and hERV signatures as predictors for patient outcome in TCGA KIRC. Comparing a full model against an all-but-one-feature model, all 3 signatures provided significant prognostic value in addition to stage and molecular subtype, with the RIG-I–like down signature contributing nearly as much prognostic power as traditional staging and each of the 3 signatures providing greater prognostic power than molecular subtyping (Figure 4D and Supplemental Table 4). To establish whether these hERV signatures were prognostic in other tumors, we performed univariable CoxPH for each signature within all TCGA cancer types (Figure 4E). Among these 3 signatures, BCR-associated hERVs were additionally prognostic in COAD and LGG, while RIG-I–like down hERVs were additionally prognostic in BLCA, COAD, KIRP, LGG, and LIHC, suggesting these additional cancer types may have hERV–immune microenvironment interactions similar to those in ccRCC. Included among these cancer types were KIRP and COAD, both of which were closely related to KIRC by hierarchical clustering of hERV expression patterns (Supplemental Figure 5), and LGG, which contained the second greatest number of prognostic hERVs after KIRC (Figure 1E).

hERVs demonstrate evidence of tumor-specific presentation of targetable viral epitopes. Previous studies have identified select tumor-specific hERV epitopes in ccRCC that trigger in vitro antitumor responses with limited in vivo efficacy (35–37). Studies regarding neoantigens have suggested that a large number of potential epitopes are required for screening in order to identify a few clinically relevant peptides with significant in vivo antitumor efficacy (45–48). We examined hERV expression patterns between tumors and matched normal tissue within TCGA KIRC and observed that normal samples clustered together (Supplemental Figure 16). The majority of hERVs were heavily upregulated in tumor compared with matched normal samples, leading us to hypothesize that there may be many more differentially expressed and targetable hERVs within tumor than previously described. In an attempt to expand the potentially targetable hERV epitope pool in ccRCC, we first ranked hERVs based on fold change in expression between tumor and matched normal samples (Supplemental Figure 17) (49). Notably, CT-RCC hERV-E (HERVERI/gammaretrovirus-like, designated as hERV 2256 in the reference database, also known as ERVE-4), one of the few hERVs demonstrated to be capable of eliciting a vaccine-inducible $CD8^+$ T cell response, ranked second highest in tumor versus normal fold change in expression (35–37). This same hERV was previously described by Rooney et al. (ERVE-4) and was found to be significantly upregulated in ccRCC and associated with a signature of cytotoxicity (15). To ensure that our analyses were consistent with these previously published findings, we performed linear regression between CT-RCC hERV-E and IGS expression including the Rooney signature for cytotoxicity (CYT), and observed a significant association between expression of this hERV and the majority of IGS in our set, including CYT (Supplemental Figure 18).


A
Superfamilies
HEPSI
HERVERI
HERVFRDLIKE
HERVHF
HERVIPADP
HERVW9
HML
HSERVIII
HUERSP
MER50LIKE
MLLV*
Uncertain
-4 -2 0 2 4
Canonical clades
HEPSI2
HEPSI3
HERV1
HERV1_ARTIODACT
HERV3
HERV4
HERV9
HERVADP
HERVE
HERVFA
HERVFB
HERVFC
HERVFRD
HERVH
HERVH48
HERVI
HERVIP
HERVL
HERVS
HERVT
HERVW
HML1
HML10
HML2
HML3
HML4
HML5
HML6
HML7
HML8
HML9
HUERSP1
HUERSP2
HUERSP3
LTR46
MER65
PABL
PRIMA4
PRIMA41
Undefined
CASWDDSLNGWVF
CQQYNTYPLTF
CQSYDSSLSAYVF
CSSYTTSSTYVF


B
TCAGGCCCTCAAACCNCACAA
AACCTCNCCCTTCAAGGTGTACANT
CTTGCTACAAGTGCCAGAAATCT
AGANTTAACTNATCACNGACGNC
TTACCATCTCATTAAAACCTAATCACC
CCGCAANGCTTCACNGACAGC
NGACTNNTCTNTCAAACCNNTAATAC
ATCCAGGTCATTACCAATCAT

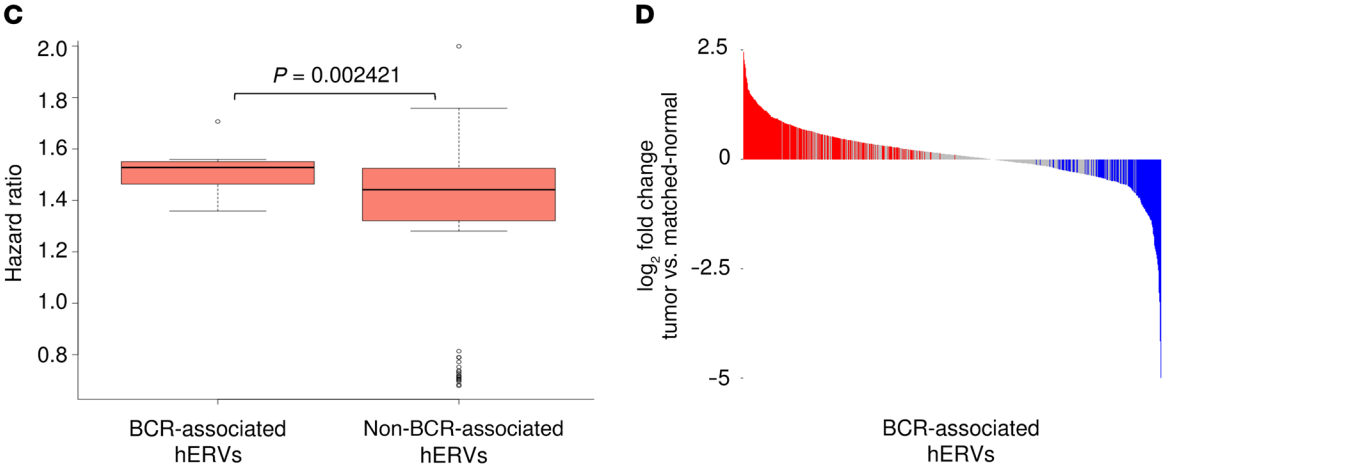

C
Hazard ratio
2.0
1.8
1.6
1.4
1.2
1.0
0.8
P = 0.002421
BCR-associated hERVs
Non-BCR-associated hERVs
D
log2 fold change tumor vs. matched-normal
2.5
0
-2.5
-5
BCR-associated hERVs

Figure 3. hERVs associated with expression of BCR clonotypes are negatively prognostic in ccRCC. (**A**) Heatmap of association (GLM) between hERV expression and expression of B cell clonotypes, displaying all TCRs and BCRs that demonstrate association (left, FDR-corrected $P \leq 0.05$) and a magnified view of the top 4 B cell clones with highest numbers of significantly associated hERVs (right, underscored by black box to the bottom left). FDR-corrected *P* values represented by intensity of color and direction of coefficient represented by color (red: positive, blue: negative). (**B**) Multiple sequence alignment of areas of DNA identity in ≥25% of hERVs (all hERVs significantly associated with the top 4 B cell clones) and ≥24 base pairs in length (minimum BCR epitope length). Base pair sequences displayed by color (A: blue; T: red; C: green; G: yellow; gap: gray) and sequence below. *y* axis order is conserved in all plots. (**A** and **B**) Color bars at left show superfamily and canonical clade classification. (**C**) Hazard ratios among all hERVs significantly associated to the top 4 B cell clones (left) or non-BCR-associated hERVs (right) within TCGA KIRC, with Welch's *t* test *P* value displayed. Data represent median (middle line), with boxes encompassing the 25th to 75th percentile, whiskers encompassing 1.5× the interquartile range from the box, and outliers shown by dots. (**D**) Waterfall plot displaying the $\log_2$ fold change in mean expression of hERVs associated with the top 4 B cell clones in the tumor compared with matched normal tissue. FDR-adjusted *P* value significance ($P \leq 0.05$) from DESeq2 analysis displayed in red (positive fold difference), blue (negative fold difference), and gray (nonsignificant).

Similar to the pattern observed in CT-RCC hERV-E, hERVs that were overexpressed within tumors were ubiquitously positively associated with IGS, while those that demonstrated overexpression within matched normal tissue demonstrated a mixed association pattern (FDR-corrected $P \leq 0.05$; Figure 5A), suggesting that preferential hERV expression in the tumor may facilitate immune activation. Interestingly, none of the top 10 hERVs by tumor versus normal expression were significantly associated with TCR/BCR clonotype expression or with survival. Given that (i) these hERVs were significantly associated with immune activation and (ii) there is evidence of functional epitopes and public hERV-specific T cells in at least one of these hERVs (CT-RCC hERV-E), the inability to computationally detect TCRs/BCRs significantly associated with these hERVs suggests we lacked the sensitivity necessary to identify these hERV-specific TCR/BCR clones. This lack of detectable public adaptive immune response is also characteristic of neoantigens, which despite failing to show association with TCR/BCR expression and survival in the absence of immunotherapy in ccRCC, have been recently demonstrated to provide vaccine-induced efficacy in melanoma (46, 50).

Tumor-specific transcription is necessary for epitope generation but is not sufficient without downstream translation. Since the majority of hERVs are translationally inactive, we ran hervQuant on a publicly available Ribo-Seq dataset comprising several regions from 2 ccRCC and matched normal kidney nephrectomy samples (4 regions per tumor; 2 regions per matched normal) (22). To filter for hERVs with the strongest evidence of differential expression by both Ribo-Seq and RNA-Seq, we ranked hERVs by the sum of RNA-Seq and Ribo-Seq fold change in expression in tumor versus normal samples (Supplemental Figure 19). Despite evidence of translation in the literature, CT-RCC hERV-E did not demonstrate coverage by Ribo-Seq in this ccRCC dataset, suggesting the relative insensitivity of Ribo-Seq– compared with RNA-Seq–based hERV identification. However, analysis of the GWIPS database (51) containing aggregate data from >30 Ribo-Seq datasets provided evidence for translation of CT-RCC hERV-E in several human lymphoblastic cell lines but minimal translation in all other sets, including normal human tissues, suggesting that CT-RCC hERV-E had the capacity for translation within tumor-like tissues (Supplemental Figure 20). hERV 4700 (HERVERI/gammaretrovirus-like), which demonstrated the highest tumor versus normal expression by RNA-Seq, was identified as the most differentially expressed hERV with greatest evidence of translation. Additionally, hERV 4700 was expressed at low levels in matched normal tissues from all other tumor subtypes (Supplemental Figure 21) and demonstrated additional evidence of translation among GWIPS tumor cell line samples (Supplemental Figure 22). Although Ribo-Seq coverage of hERV 4700 within ccRCC samples was relatively low, coverage patterns were similar to those observed by RNA-Seq (Figure 5B). Areas of coverage within the hERV 4700 proviral reference corresponded to viral *gag* (red), *pol* (blue), and *env* (green) genes. Protein-BLAST of these regions translated across each reading frame provided high sequence similarity with known reference hERV sequences across all 3 frames of *pol* and *env*, and frame 2 of *gag* (Figure 5C and Supplemental Figure 23). Using the longest sequence identified within each protein reading frame, we performed NetMHCPan4.0 epitope prediction, identifying 30 predicted HLA-A*02:01 binders (binding affinity ≤500 nM; Supplemental Table 5) (52). To ensure these predicted epitopes were hERV specific, we searched for overlap between amino acid sequences of each peptide with known human proteins in the GENCODE hg19 protein-coding transcript translated sequences, observing no overlap between epitopes and non-hERV proteins. Using an HLA-A*02:01 monomer UV exchange assay and HLA ELISA readout (53–58), we validated the binding of 30 of 30 predicted epitopes to HLA-A*02:01 with exchange efficiencies ranging from 16.1% to 73.1% (Figure 5D).

hERV epitopes associate with aPD1 response with evidence of epitope-specific T cells in ccRCC. To explore whether hERV 4700 expression is predictive for patient response to aPD1 therapy, we performed quantitative real-time PCR (RT-qPCR) quantification of hERV 4700 with 2 of each *gag*-, *pol*-, and *env*-specific primer/probe sets on ccRCC tumor biopsy RNA in aPD1-treated patients (responders: $n = 7$, nonresponders: $n = 6$; Figure 5E and Supplemental Tables 6–8). We observed greater mean RT-qPCR signal in aPD1 responders in all primer/probe sets (Mann-Whitney *U* test $P < 0.05$; Supplemental Table 9), as well as hervQuant-derived hERV 4700 expression from the same set with added samples (responders: $n = 10$, nonresponders: $n = 10$; Mann-Whitney *U* test $P = 0.0455$), suggesting that transcription of hERV 4700 is associated with greater responsiveness to immunotherapy. Additionally, multivariable linear regression (GLM) provided perfect fit of primer/probe sets as a predictor for response. To demonstrate the presence of an anti–hERV 4700 T cell immune response in ccRCC, we performed tetramer staining of an HLA-A*02:01 ccRCC tumor sample using the 30 MHC tetramers described above (Figure 6, A and B). Using a stepwise approach, we first screened the tumor using 5 pools of 6 tetramers, which demonstrated that pool 4 had the largest tetramer-positive $CD8^+$ T cell population (11.3% tetramer-positive). Running the 6 individual tetramers, we observed tetramers 2 and 3 to have the greatest staining, which corresponded to peptides derived from frame 2 of the *gag* (10.9% positive) and *pol* (13.5%) protein regions, respec-

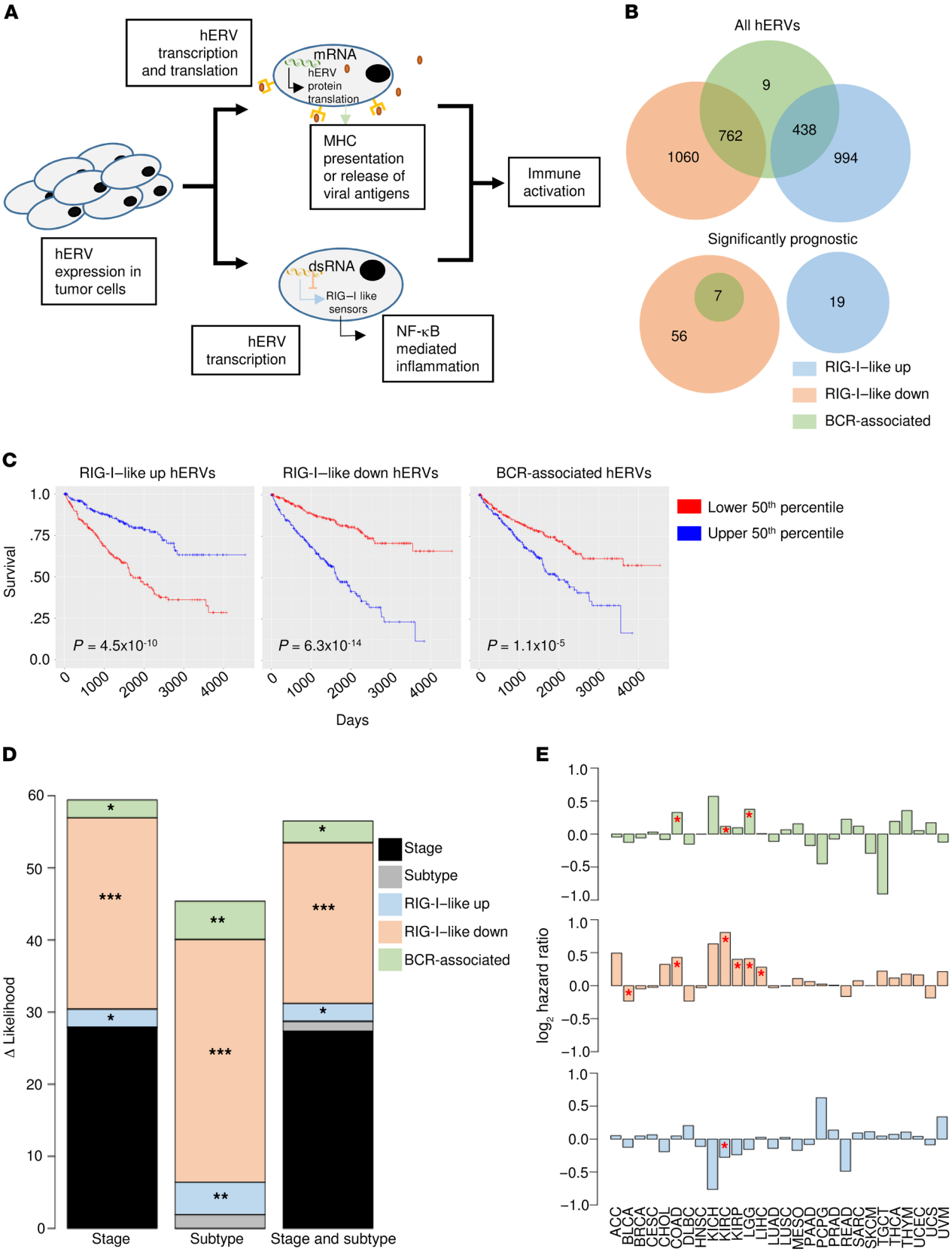

A
hERV transcription and translation
mRNA
hERV protein translation
MHC presentation or release of viral antigens
Immune activation
hERV expression in tumor cells
dsRNA
RIG–I like sensors
hERV transcription
NF-κB mediated inflammation
B
All hERVs
9
762
438
1060
994
Significantly prognostic
7
19
56
RIG-I–like up
RIG-I–like down
BCR-associated
C
RIG-I–like up hERVs
RIG-I–like down hERVs
BCR-associated hERVs
Survival
P = 4.5x10-10
P = 6.3x10-14
P = 1.1x10-5
Days
Lower 50th percentile
Upper 50th percentile
D
Δ Likelihood
Stage
Subtype
Stage and subtype
RIG-I–like up
RIG-I–like down
BCR-associated
E
log2 hazard ratio
ACC BLCA BRCA CESC CHOL COAD DLBC HNSC KICH KIRC KIRP LGG LIHC LUAD LUSC MESO PAAD PCPG PRAD READ SARC SKCM TGCT THCA THYM UCEC UCS UVM

Figure 4. Immune-related hERV signatures are prognostic for patient overall survival. (**A**) Schematic summary of hERV interactions with the immune system in the context of an anti-tumor immune response. (**B**) Venn diagram showing the number hERVs significantly associated (GLM, FDR-corrected $P < 0.05$) with genes corresponding to the upregulation (blue) or downregulation (orange) of the RIG-I–like pathway or positively associated (GLM, FDR corrected $P < 0.05$) with expression of B cell clones (green). (**C**) Kaplan-Meier survival curves for TCGA KIRC patients split by the upper (blue) and lower (red) 50th percentile of expression for each of the 3 hERV group signatures represented in **A**. (**D**) Change in multivariable CoxPH log-likelihood ratios in TCGA KIRC using clinical stage and/or M1–M4 molecular subtyping and the 3 classes of hERV groups represented in **B** as predictors for survival. Stacked bars show the change in likelihood ratio for each feature when removed from the full model, as well as the χ^2 test P value for each hERV group signature when removed from the full model (*$P \leq 0.05$, **$P \leq 0.01$, ***$P \leq 0.001$). (**E**) Univariable CoxPH coefficients for hERV signatures as a predictor for overall survival among each cancer type. FDR-corrected P value represented by red asterisks (*$P \leq 0.05$).

tively. We validated the presence of these T cell populations in 3 additional ccRCC tumors (*gag*: 10.9%–24.8%; *pol*: 13.5%–22.3%), as well as observing staining within the range of negative control tetramers in 4 healthy donor peripheral blood mononuclear cells (PBMC) samples (*gag*: 0.12%–1.51%, *pol*: 0.13%–0.76%; Figure 6C and Supplemental Figure 24). Overall, these data validate our epitope prediction method and provide evidence for the presence of hERV 4700–specific T cells within ccRCC.

Discussion

We report here a hierarchical analysis of hERV–immune microenvironment interactions within the TCGA pan-cancer dataset, integrated with Ribo-Seq data, RNA-Seq data from immunotherapy-treated patients, and functional biological assays, to provide insight into hERV immunobiology in cancer. Our broad survey of hERV expression and association patterns provided multiple lines of evidence that hERVs shape the tumor immune microenvironment in several cancer types. Conditioning on cancer type, we observed that gene signatures of immune responsiveness (aPD1-responsive signature, effector immune cells) were positively associated with hERV expression, suggesting that hERVs may either directly interact with antitumor immunity through immune activation or provide a biomarker for an active antitumor immune response. In agreement with this view, we observed that hERVs were significantly prognostic in multiple cancer types, with the greatest enrichment of prognostic hERVs observed in ccRCC. Interestingly, BLCA was the only cancer type in which greater average hERV expression resulted in significantly longer survival times. This finding suggests potentially different hERV-mediated tumor immunobiology in BLCA and should be further explored in future studies. For IGS and CoxPH analyses, hERV expression data were normalized either (i) to total RNA-Seq read count (reads per million; RPM) to determine the impact of absolute hERV expression or (ii) to upper quartile normalization (UQN) of hERV reads within each sample to determine the impact of relative hERV proportions (Supplemental Tables 10 and 11). IGS patterns of association were strongly conserved between hERV expression by UQN and read normalization. We observed variability in hERV association patterns with 3 $CD8^+$ T cell signatures derived from different publications (CD8_T_Cell, CD8_Cluster, CD8) (30, 32, 33), with CD8_T_Cell showing an association pattern different from the other 2 signatures. The CD8_T_Cell signature contained a set of 8 genes that accounted for its variation from the other 2 signatures —*HAUS3* (cytokinesis and mitosis), *SF1* (pre-mRNA splicing), *SFRS7* (pre-mRNA splicing), *ZNF91* (protein coding), *ZNF609* (protein coding), *THUMPD1* (gene expression/rRNA processing), *MYST3* (histone acetyltransferase), and *CDKN2A* (cell cycle regulator) — all of which are nonspecific to $CD8^+$ T cells in function (Supplemental Figure 25). Nevertheless, we included the CD8_T_Cell signature within all analyses (including Treg-to-$CD8^+$ ratio) because it remains a commonly used signature for $CD8^+$ T cells within the literature.

In contrast to IGS, CoxPH analysis with UQN hERV data contained a greater number of positively prognostic hERVs compared with read-normalized data, suggesting that the proportional expression of hERVs may also influence overall survival. We additionally observed that the majority of hERVs were associated with younger patient age. Since most tumor types show an association between older age and worse outcome, and the majority of significantly prognostic hERVs were associated with worse outcome, these results suggest that the association between hERVs and patient outcome was not simply due to an association with age.

Due to the diverse tumor-immune interactions observed among different cancer types, we narrowed down further the role of hERVs upon the tumor immune microenvironment to one cancer type. We focused on ccRCC to further study the role of hERVs in shaping the tumor immune microenvironment because (i) it contained the greatest number of prognostic hERVs and (ii) hERV proteins are known to be expressed and immunogenic in ccRCC (14, 35–37).

Within ccRCC, we considered the potential for hERVs to impact both arms of the immune system. The role of hERVs in triggering an innate immune response is underscored by several recent reports noting that epigenetic-modifying agents that promote greater DNA demethylation — decitabine (methyltransferase inhibitor) and abemaciclib (CDK4/6 inhibitor) (26, 27) — increased expression of retroviral elements and triggered subsequent antitumor responses through innate sensor signaling, including induction of RIG-I–like pathway detection of viral dsRNAs. While these previous reports demonstrated only the proinflammatory nature of selected hERV elements, we were surprised to find two strikingly distinct patterns of association between hERV expression in ccRCC and expression of genes associated with the RIG-I–like family. The implication of this clustering pattern (along with the significantly different patterns of association between these hERV groups with survival and IGS expression) is that hERVs may play both agonistic and antagonistic roles in innate sensor immunity. Potentially, group 2 hERVs (RIG-I–like down) may interfere with RIG-I–like signaling through a currently unknown mechanism, ultimately skewing the tumor immune microenvironment in favor of an immunosuppressive phenotype with greater Treg–to–$CD8^+$ T cell ratios and negatively impacting patient prognosis.

Next, we studied the role of hERVs in triggering an adaptive immune response through hERV-mediated immune activation of retroviral epitope-driven T and B cell responses. MiXCR analysis of TCGA KIRC failed to identify TCR clones that were shared across at least 10% of samples, suggesting that while hERV

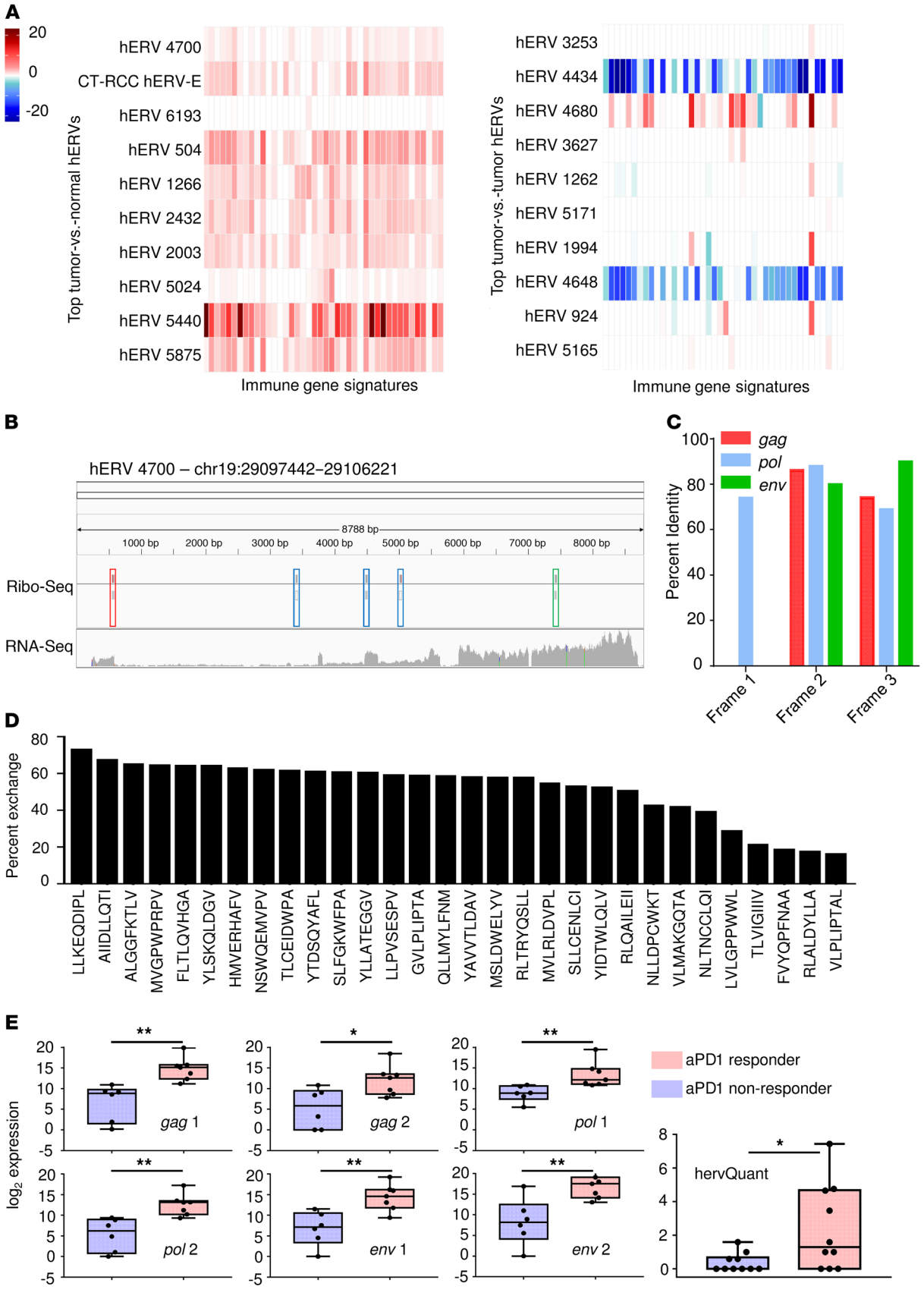

A
20
0
-20
hERV 4700
CT-RCC hERV-E
hERV 6193
hERV 504
hERV 1266
hERV 2432
hERV 2003
hERV 5024
hERV 5440
hERV 5875
Top tumor-vs.-normal hERVs
Immune gene signatures
hERV 3253
hERV 4434
hERV 4680
hERV 3627
hERV 1262
hERV 5171
hERV 1994
hERV 4648
hERV 924
hERV 5165
Top tumor-vs.-tumor hERVs
Immune gene signatures
B
hERV 4700 – chr19:29097442–29106221
8788 bp
1000 bp
2000 bp
3000 bp
4000 bp
5000 bp
6000 bp
7000 bp
8000 bp
Ribo-Seq
RNA-Seq
C
gag
pol
env
Percent Identity
Frame 1
Frame 2
Frame 3
D
Percent exchange
LLKEQDIPL
AIIDLLQTI
ALGGFKTLV
MVGPWPRPV
FLTLQVHGA
YLSKQLDGV
HMVERHAFV
NSWQEMVPV
TLCEIDWPA
YTDSQYAFL
SLFGKWFPA
YLLATEGGV
LLPVSESPV
GVLPLIPTA
QLLMYLFNM
YAVVTLDAV
MSLDWELYV
RLTRYQSLL
MVLRLDVPL
SLLCENLCI
YIDTWLQLV
RLQAILEII
NLLDPCWKT
VLMAKGQTA
NLTNCCLQI
LVLGPPWWL
TLVIGIIIV
FVYQPFNAA
RLALDYLLA
VLPLIPTAL
E
log2 expression
gag 1
gag 2
pol 1
pol 2
env 1
env 2
aPD1 responder
aPD1 non-responder
hervQuant

Figure 5. hERVs demonstrate evidence of targetable epitope expression in ccRCC. (**A**) Association (GLM) of the 10 most positively (left) and negatively (right) differentially expressed hERVs (TCGA KIRC tumor relative to matched normal tissue) with IGS expression. FDR-corrected *P* values represented by intensity of color and direction of coefficient represented by color (red: positive, blue: negative). (**B**) Read coverage from ccRCC Ribo-Seq data for hERV 4700, demonstrating read coverage of coding regions for *gag* (red), *pol* (blue), and *env* (green) genes. (**C**) Percent identity between all reading frames of translated amino acid sequences from the reference *gag* (red), *pol* (blue), and *env* (green) sequences for hERV 4700 with known hERV proteins in the NIH retroviral protein BLAST database. (**D**) Exchange efficiency for HLA-A*02:01 monomer UV exchange of predicted hERV 4700 epitopes. (**E**) Left: RT-qPCR (responders: $n = 7$; nonresponders: $n = 6$) $\log_2$ expression of hERV 4700 *gag*, *pol*, and, *env* sequences. Right: hervQuant-derived (responders: $n = 10$; nonresponders: $n = 10$) hERV 4700 expression in nivolumab-treated (aPD1-treated) ccRCC tumor biopsies. Statistical analysis performed using Mann-Whitney *U* test (*$P \leq 0.05$, **$P \leq 0.01$, NS: $P > 0.05$). Data presented as values (dots) and median (middle line), with boxes encompassing the 25th to 75th percentile and whiskers encompassing minimum to maximum values.

epitopes have the capacity to trigger a T cell–driven antitumor response (35–37), we lacked the sensitivity to computationally identify public hERV-specific TCR clones. In agreement with this, comparison of MiXCR-derived TCR expression with previously described TCRs derived from amplicon-based adaptive TCR repertoire profiling in 3 TCGA KIRC samples demonstrated low total TCR counts of MiXCR data with low frequencies of overlapping clones (Supplemental Figure 26). In contrast, we observed a large pool of shared BCRs. It is important to note that BCR repertoires are likely more completely sampled from RNA-Seq data than are TCR repertoires, as we observed increased BCR sequence reads, consistent with the greater transcription of immunoglobulin mRNA from cells of the B cell lineage compared with TCR mRNA transcription from activated T cells. Thus, our study had greater power to detect BCR than TCR repertoire associations. Multiple sequence alignment of BCR-associated hERVs demonstrated clustering of proviral sequences by superfamily, suggesting that a B cell response generated against shared hERV epitopes is likely to occur within one or several closely related superfamilies. The higher hazard ratios among BCR-associated hERVs may be related to the lack of tumor specificity for these hERVs. The majority of IGS in ccRCC, including those for B cells, have been shown to be associated with worse prognosis (59). While the mechanism for this finding is currently undetermined, a potential contributor to this pattern may be a B cell response in which hERVs are generated in the tumor with epitopes shared by hERVS upregulated within the surrounding normal tissues. Further investigation should be performed to study the importance of this potential anti-hERV B cell response in ccRCC.

Evidence for hERV-mediated activation of the innate and adaptive immune responses suggests that expression of these proviruses within tumors may contribute to immune editing of tumor cell populations. Highly immunogenic hERVs with the capacity to be recognized by endogenous T and B cell responses are likely cleared by the immune system or otherwise expressed under a heavily immunosuppressed microenvironment. There may also exist additional hERV epitopes that generate immune responses too weak to promote antitumor immunity. These two groups can both be potentially targeted for immune activation through the use of nonspecific (e.g., checkpoint blockade therapy, innate immune agonists) or epitope-specific (vaccination, adoptive T cell therapy) immunotherapies. Further time-course immune profiling studies should be performed to study the mechanisms of hERV-mediated immune surveillance in a developing tumor.

With evidence of hERV-mediated activation of both innate and adaptive immune responses, we sought to examine whether these responses could be used to develop a model for patient prognosis in ccRCC. Apart from molecular subtyping, no molecular markers have improved the prognostic capabilities of current clinical predictive systems in ccRCC, suggesting the potential for development of hERV-based signatures as a biomarker for survival. In attempt to identify such a prognostic biomarker, we created hERV signatures derived from our previous analysis of hERV interactions with the innate and adaptive immune response. Based on these signatures, we developed a model that provided significantly greater prognostic power than M1–M4 molecular subtyping and levels of prognostic information similar to those of traditional clinical staging. Additionally, while these hERV signatures were derived and optimized for ccRCC, we showed 2 signatures to provide prognosis in several other tumor models related to ccRCC by hERV expression patterns, level of prognostic hERVs, and tissue of origin, implying that additional hERV signatures for patient prognosis can be independently developed for other cancer types.

Last, we sought to develop a screening method for detection of hERVs actively undergoing translation. The implication of such a tool is the potential for development of immune response biomarkers and antitumor T cell vaccine therapies, similar to those developed in neoantigen-based vaccine studies. Our analysis of tumor-specific hERVs in ccRCC identified CT-RCC hERV-E as the second highest differentially expressed hERV by RNA-Seq expression. This particular hERV has been well described in the literature as a ccRCC tumor–specific provirus with evidence of hERV-specific T cell responses (35–37). Within our Ribo-Seq analysis, we were underpowered to detect evidence of CT-RCC hERV-E translation among 2 ccRCC samples. However, our analysis of the GWIPS database provided evidence for the translation of CT-RCC hERV-E in human tumor cells but not in normal blood, fibroblasts, or muscle tissue. This conforms to the view that CT-RCC hERV-E has the capacity for translation under tumor-specific conditions and suggests that deeper Ribo-Seq coverage in ccRCC may be needed to increase the sensitivity of our computational screening to broaden the set of potentially targetable hERV epitopes. Our analysis of CT-RCC hERV-E RNA-Seq expression in TCGA KIRC data supports the previous report by Rooney et al. identifying this hERV as being upregulated in ccRCC and associated with a gene expression index of cytotoxicity (15). We observed the same significant association with their cytotoxicity signature and additionally identified a large proportion of other IGS strongly associated with its expression. Among these, the most significantly associated was the Treg signature, suggesting that expression of CT-RCC hERV-E may be also associated with immunosuppression. This strong association with immunosuppressive signatures suggests CT-RCC hERV-E may be another potential marker of response for immunotherapies such as aPD1 checkpoint blockade therapy.

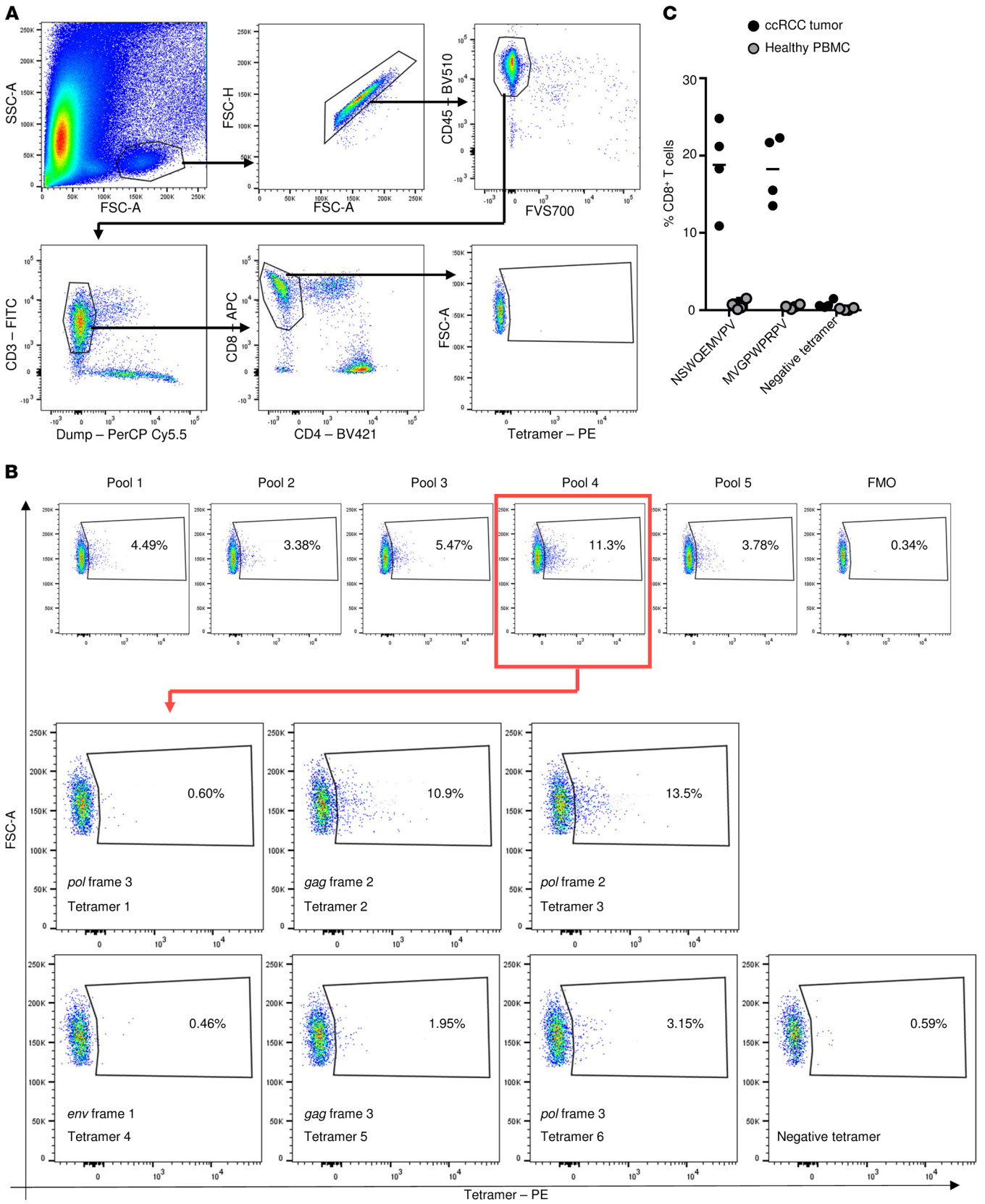

A
SSC-A
FSC-A
FSC-H
FSC-A
CD45 – BV510
FVS700
CD3 – FITC
Dump – PerCP Cy5.5
CD8 – APC
CD4 – BV421
FSC-A
Tetramer – PE
C
ccRCC tumor
Healthy PBMC
% CD8+ T cells
30
20
10
0
NSWQEMVPV
MVGPWPRPV
Negative tetramer
B
Pool 1
Pool 2
Pool 3
Pool 4
Pool 5
FMO
4.49%
3.38%
5.47%
11.3%
3.78%
0.34%
FSC-A
0.60%
pol frame 3
Tetramer 1
10.9%
gag frame 2
Tetramer 2
13.5%
pol frame 2
Tetramer 3
0.46%
env frame 1
Tetramer 4
1.95%
gag frame 3
Tetramer 5
3.15%
pol frame 3
Tetramer 6
0.59%
Negative tetramer
Tetramer – PE

Figure 6. hERV 4700 epitope–derived HLA-A*02:01 tetramers identify the presence of *gag*- and *pol*-specific T cells in ccRCC. (**A**) Flow cytometric representative gating strategy for identification of $CD8^+$ epitope-specific T cells in ccRCC tumor. (**B**) Epitope gating for 5 pools of 6 tetramers (top), as well as staining of individual tetramers from pool 4 (bottom) in ccRCC. (**C**) Percent tetramer-specific $CD8^+$ T cells for epitopes identified in **B** (tetramer 2: NSWQEMVPV; tetramer 3: MVGPWPRPV) in ccRCC tumors (n = 4) and healthy donor PBMC samples (n = 4). Dots represent values for each sample, with bars representing the mean across each group. Negative controls for gating definitions include tetramer fluorescence-minus-one (FMO) (**A**) and nonspecific HLA-A*02:01-negative tetramer (**B** and **C**). Data presented in Figure 6 represent results from 4 independent experiments.

RNA-Seq analysis of hERV 4700 demonstrated preferential expression within ccRCC, with modest expression in normal kidney and liver. This preferential expression underscores the potential for hERV 4700–targeted immunotherapies, with the caveat that a particularly robust anti–hERV 4700 immune response could potentially result in on-target/on-tissue and on-target/off-tissue toxicity. We provided additionally validation for the transcription of this hERV through RT-qPCR and hervQuant analysis of an aPD1-treated ccRCC dataset, and showed that expression of hERV 4700 is associated with responsiveness to immunotherapy.

Ribo-Seq screening provided evidence for translation of hERV 4700, supporting translation of epitopes that we further validated to bind MHC. Additionally, tetramer staining of predicted hERV 4700 epitopes in 4 ccRCC tumors demonstrated the presence of infiltrating T cells with receptors specific for *gag*- and *pol*-derived epitopes, supporting the idea that (i) hERV 4700 may act as a direct target in ccRCC, whereby aPD1 could trigger an antitumor response against hERV 4700–derived epitopes, and (ii) hERV 4700 expression may be a new biomarker of aPD1 responsiveness in ccRCC. These same T cell populations were scarce to absent in healthy donor PBMCs, confirming the specificity of these T cells in ccRCC tumors. Tetramer-specific T cell frequencies were particularly high among ccRCC tumors (NSWQEMPV, 10.9%–24.8%; MVFPWPRPV, 13.5%–22.3%), suggesting that as much as 40% of tumor-infiltrating $CD8^+$ T cells may be specific for these 2 hERV 4700 epitopes. We recognize that these frequencies are particularly high for a tumor-infiltrating population, and several caveats exist for our analyses. First is the potential for T cell cross-reactivity against these tetramers, as well as peptide impurities that recognize other infiltrating T cell populations. Additionally, tetramer-positive populations contained a large range of fluorescence intensities, suggesting these T cells do not necessarily comprise a single clone but likely several different clones with different TCR affinities. Future studies to characterize the TCR sequences and phenotypic characteristics of these tetramer-positive populations should be performed to further elucidate the role of these populations and determine the basis for these and other potential caveats.

In addition to hERV 4700, we observed 172 other hERVs that were differentially expressed between aPD1 responders and nonresponders by hervQuant profiling (Wilcoxon's test, $P < 0.05$), suggesting that a more comprehensive set of hERV expression signatures may exist for the development of an aPD1 response biomarker in ccRCC (Supplemental Figure 27). Of these hERVs, 6 demonstrated overlap with the RIG-I–like down signature, one with the BCR-associated signature and 34 with all prognostic hERVs, suggesting relatively low overlap between the set of predictive and prognostic hERVs. Overall, hervQuant is the first described method to our knowledge for comprehensive identification of potentially targetable hERV epitopes. Further validation should be performed to confirm the capacity of these potential hERV epitopes as therapeutic vaccine targets and to develop a robust hERV-based biomarker for immunotherapy response in ccRCC.

In summary, we describe a computational workflow, hervQuant, for robust quantification of individual hERVs using RNA-Seq data. The data gained through hervQuant provide insights into the pan-cancer landscape of hERV expression and immune modulation. Within ccRCC, we found a distinct group of hERVs that were inversely associated with RIG-I–like signaling genes, prognosis, and IGS expression. Additionally, we examined the interaction between hERV expression in ccRCC and activation of B cell clonotypes, and demonstrated the capacity of the above-mentioned hERV classes to provide a multivariable model of patient prognosis that significantly outperforms traditional clinical staging and molecular subtype prognosis models in ccRCC. We provide evidence for a new method of hERV epitope prediction based on differential hERV expression in the tumor, Ribo-Seq screening for translation, computational epitope prediction, in vitro validation for HLA binding, and in vivo detection of epitope-specific T cells in a ccRCC tumor. Importantly, we observed that hERV sequences identified through this approach were significantly associated with aPD1 responsiveness in ccRCC tumors, supporting continued research into hERVs as biomarkers and therapeutic targets for immunotherapy. With the recent increasing interest in the role of hERVs in modulating the tumor immune microenvironment, we believe the work presented here substantially expands our understanding of hERV biology and opens the way for future development of technologies to exploit hERV biology for new therapeutic tools.

Methods

Alignment and quantification of hERV expression from RNA-Seq data. hERV genomic coordinates were derived from a previously a published study by Vargiu et al. (3). Full-length hERV sequences were masked for low complexity reads (9 or more repeating single nt; 7 or more repeating double nt; 4 or more repeating nt patterns of 3; 3 or more repeating nt patterns of 4; 2 or more repeating patterns of 5; 2 or more repeating nt patterns of 5) and compiled alongside human hg19 transcriptome reads into a reference file for downstream alignment. RNA-Seq FASTQ files were aligned to the hERV reference using STAR v2.5.3 (multimaps ≤10, mismatch ≤7) (60). BAM output files were filtered for reads that mapped to hERV reference using SAMtools (v1.4) (61), then quantified using Salmon v0.8.2 (Quant mode, –l ISF) (62). Raw expression matrices were either normalized to hERV counts per million total FASTQ reads and $\log_2$ transformed, or normalized to the upper quartile hERV expression value among non-zero values within each sample and $\log_2$ transformed (Supplemental Tables 12–14). Only TCGA pan-cancer samples sequenced with Illumina HiSeq 2 × 50 bp were analyzed. See the supplemental material for optimization details and input parameters.

RNA-Seq expression, IGS analysis, and survival analysis. MapSplice-aligned, RSEM-quantified RNA-Seq expression matrices and survival data were downloaded from FireBrowse (http://firebrowse.

org/). Expression matrices were merged between all cancer types, upper quartile normalized within each sample, and $\log_2$ transformed. IGS were derived from previously described signatures (28–32), with expression calculated as the mean expression of each gene within the signature. TCGA LAML samples were omitted from analysis in order to prevent skewing of IGS patterns.

TCR/BCR alignment. MiXCR (v2.1.1) was used for identification of TCR and BCR sequences with TCGA KIRC (41). Following suggested run methods provided by MiXCR's documentation for RNA-Seq data (https://mixcr.readthedocs.io/en/latest/rnaseq.html), paired-end FASTQ files were run through alignment in RNA-Seq mode, 2 rounds of contig assembly, extension of incomplete CDR3s, assembly, and export. Data were subsequently converted into an expression matrix, dropping all clones (defined as conserved amino acid CDR3 sequence) with expression in fewer than 10% of all TCGA KIRC samples, and scaled to counts per billion total FASTQ reads.

*HLA-A*02:01 monomer UV exchange and β2-microglobulin ELISA*. Epitope prediction was performed with the NetMHCpan 4.0 Server interface, defining predicted HLA binders as those with binding affinity ≤500 nM (52). Predicted hERV epitopes were synthesized through New England Peptide array technology. Monomer exchange reaction was carried out using the BioLegend Flex-T HLA-A*02:01 monomer UV exchange protocol (57). Peptide exchange efficiency was performed using the BioLegend HLA class I ELISA protocol (58).

RT-qPCR validation of hERV 4700. Expression levels of hERV 4700 were assessed by RT-qPCR in a collection of ccRCC formalin-fixed, paraffin-embedded (FFPE) archival tissue from responders (n = 7 patients; 9 samples) and nonresponders (n = 6 patients; 6 samples). RT-qPCR was performed on all available samples, with no further selection process. Total RNA isolation was performed using the RNAeasy FFPE Kit (QIAGEN). DNAse treatment was performed during RNA isolation using RNase-free DNase I (QIAGEN). RNA quality and concentration were assessed using a NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies).

First-strand cDNA synthesis was performed using 250 ng total RNA, random hexamers, and the SuperScript IV Reverse Transcriptase Kit (Life Technologies). RT-qPCR was performed on a CFX96 Touch Real-Time PCR Detection System (Bio-Rad) using TaqMan Universal PCR Master Mix (Applied Biosystems). RT-qPCR primer and probe sequences are shown in Supplemental Table 7. All analyses were performed in triplicate, and relative RNA levels were determined using hypoxanthine phosphoribosyltransferase 1 (HPRT1) as an endogenous internal control (Applied Biosystems, catalog 4333768). A HeLa control RNA sample was included for inter-plate calibration. hERV 4700 expression levels were calculated using the ΔΔCt method. Expression levels for 2 sample pairs derived from the same patients were averaged for statistical analyses in Figure 5E.

Flow cytometric analysis. Tetramer and cell surface staining was performed as described previously (63). Briefly, viably frozen, histologically subtyped ccRCC tumor samples were thawed and stained for HLA-A2 (BD Biosciences; clone BB7.2, allophycocyanin [APC]). Separately, samples positive for HLA-A2 were treated with 50 nM dasatinib for 30 minutes at 37°C, then stained using approximately 10 μg/ml tetramer (phycoerythrin [PE]) or Beckman Coulter iTAg MHC class I human–negative tetramer control on ice for 30 minutes. Cells were then washed and incubated on ice with 5 μg/ml biotin-conjugated anti-PE antibody (BioLegend; PE001) for 20 minutes, followed by 2 washes, then further incubation with 5 μg/ml streptavidin, R-PE conjugate (SAPE) for 10 minutes on ice. Cells were then washed and stained for viability using BD fixable viability dye FVS700 according to the manufacturer's directions. Last, cells were Fc blocked using mouse immunoglobulin (Millipore-Sigma, catalog I5381) for 10 minutes, followed by surface staining for 20 minutes on ice with the following markers: anti-CD45 (BD Biosciences; clone HI30, BV510), anti-CD3 (BD Biosciences; clone UCHT1, FITC), anti-CD8 (Beckman Coulter; SFCI21THy2D3 [T8], APC), anti-CD4 (BD Biosciences; clone RPA-T4, BV421), anti-CD14 (BD Biosciences; clone MϕP9, PerCP Cy 5.5), anti-CD19 (BD Biosciences; clone HIB19, PerCP Cy5.5), and anti-CD56 (BD Biosciences; clone BI59, PerCP Cy5.5).

A minimum of 1,000,000 events were collected for each sample on a BD LSRFortessa flow cytometer. FlowJo flow cytometry software version 10 was used for analyses of all flow cytometric data. Tumors were derived from viably frozen nephrectomy samples from UNC Chapel Hill and Vanderbilt University hospital patients with clear cell histology. Healthy donor PBMCs were screened by and purchased from Gulf Coast Regional Blood Center, Houston, Texas, USA.

Data availability. TCGA analyses were performed on data collected and generated by the TCGA Research Network — expression matrices can be accessed at http://firebrowse.org/; TCGA raw data can be accessed in the database of Genotypes and Phenotypes (dbGaP, accession phs000178). Ribo-Seq analysis was performed on data collected by Loayza-Puch et al. and can be accessed in the NCBI's Gene Expression Omnibus database (GEO GSE59821) (22). hervQuant expression matrices for TCGA pan-cancer (UQN and RPM) and aPD1-treated ccRCC (raw reads) RNA-Seq datasets are available in Supplemental Tables 12–14. The GWIPS ribosomal profiling database is available at https://gwips.ucc.ie/. The hervQuant workflow reference and instructions are available for download at https://unclineberger.org/vincent/resources.

Statistics. GLM using the R "glm" package was used for all univariable regression, unless otherwise stated. Univariable and multivariable CoxPH was performed with the R "survival" package. Multiple sequence alignment was performed with Clustal Omega through the R "msa" package (64). Differential hERV expression was calculated using the DESeq2 R package (49). For all CoxPH analyses, P value correction was performed using Bonferroni's correction to maintain a conservative cutoff of significance. For all other analyses, 5% FDR multiple testing correction for P values was performed unless otherwise stated. Welch's t test was performed for statistical calculation in Figure 3C. Log rank test was performed for statistical calculation in Figure 4C, with no multiple testing correction. Multivariable CoxPH and χ^2 test were performed for statistical calculation in Figure 4D, with no multiple testing correction. Mann-Whitney U test was performed for statistical calculation in Figure 5E, with no multiple testing correction. $P < 0.05$ was considered significant for all statistical tests performed.

Study approval and sample acquisition. The present studies in humans were reviewed and approved by the Vanderbilt University Human Research Protections Program, and the University of North Carolina at Chapel Hill IRB and the Office of Human Research Ethics (CB 7097). Subjects provided written informed consent prior to their participation in the study. Biopsy samples were collected according to a protocol approved by the Vanderbilt University IRB (no. 160979), and the UNC IRB approved the biorepository protocol (LCCC 1212). Patients were identified through an IRB-approved protocol and identified using a pharmacy-based list. Line

of treatment for each patient varied. The response was first determined by chart review of clinicians' notes and then confirmed by the authors of this article based on RECISTS 1.1 imaging criteria.

Author contributions

BGV and SRS conceived the study. CCS, DSB, SRS, and BGV created and ran the hervQuant software. CCS, SL, and SRS performed statistical analyses for the manuscript. CCS, SRS, SJL, and BGV analyzed RNA-Seq and Ribo-Seq data and interpreted results. CCS and LMB performed flow cytometric studies. KEB, EMW, MIM, and WYK curated patient samples and information for flow cytometric and/or RT-qPCR studies. CCS, AADC, KEB, AP, SG, GB, and WKR performed the design, running, and analysis of RT-qPCR results. CCS and SRS prepared the manuscript, with participation from WKR, RS, JSS, JSP, and BGV.

Acknowledgments

We would like to acknowledge the UNC Tissue Procurement Facility for their assistance in collection of samples for this study. The project described was supported by the William Guy Forbeck Foundation Collaborative Research Award, the UNC University Cancer Research Fund and UNC Oncology Clinical Translational Research Training Program (5K12CA120780), and the NIH (5-P50-CA058223-22, 2-P30-CA016086-40, 1-U24-CA210988-01, and U54-CA198999). Funding for CCS was supported by the NIH (1F30CA225136-01). Funding for KEB was supported by the Merck-Cancer Research Institute Irvington Postdoctoral Fellowship. The results published here are in part based on data generated by the Cancer Genome Atlas managed by the NCI and National Human Genome Research Institute (NHGRI, dbGaP accession phs000178). The content is solely the responsibility of the authors and does not necessarily represent the official views of any funding agency.

Address correspondence to: Benjamin G. Vincent or Sara R. Selitsky or Jonathan S. Serody, Lineberger Comprehensive Cancer Center, University of North Carolina, CB# 7295, Chapel Hill, North Carolina 27599-7295, USA. Phone: 919.966.8412; Email: benjamin_vincent@med.unc.edu (B.G. Vincent). Phone: 919.445.0297; Email: selitsky@email.unc.edu (S.R. Selitsky). Phone: 919.962.8409; Email: jonathan_serody@med.unc.edu (J.S. Serody).

1. Löwer R, Löwer J, Kurth R. The viruses in all of us: characteristics and biological significance of human endogenous retrovirus sequences. *Proc Natl Acad Sci U S A*. 1996;93(11):5177–5184.
2. Bannert N, Kurth R. The evolutionary dynamics of human endogenous retroviral families. *Annu Rev Genomics Hum Genet*. 2006;7:149–173.
3. Vargiu L, et al. Classification and characterization of human endogenous retroviruses; mosaic forms are common. *Retrovirology*. 2016;13:7.
4. Katzourakis A, Rambaut A, Pybus OG. The evolutionary dynamics of endogenous retroviruses. *Trends Microbiol*. 2005;13(10):463–468.
5. Boller K, et al. Human endogenous retrovirus HERV-K113 is capable of producing intact viral particles. *J Gen Virol*. 2008;89(pt 2):567–572.
6. Faff O, Murray AB, Schmidt J, Leib-Mösch C, Erfle V, Hehlmann R. Retrovirus-like particles from the human T47D cell line are related to mouse mammary tumour virus and are of human endogenous origin. *J Gen Virol*. 1992;73 (pt 5):1087–1097.
7. Wang-Johanning F, et al. Expression of multiple human endogenous retrovirus surface envelope proteins in ovarian cancer. *Int J Cancer*. 2007;120(1):81–90.
8. Büscher K, Trefzer U, Hofmann M, Sterry W, Kurth R, Denner J. Expression of human endogenous retrovirus K in melanomas and melanoma cell lines. *Cancer Res*. 2005;65(10):4172–4180.
9. Wang-Johanning F, et al. Expression of human endogenous retrovirus k envelope transcripts in human breast cancer. *Clin Cancer Res*. 2001;7(6):1553–1560.
10. Contreras-Galindo R, et al. Human endogenous retrovirus K (HML-2) elements in the plasma of people with lymphoma and breast cancer. *J Virol*. 2008;82(19):9329–9336.
11. Wang-Johanning F, et al. Detecting the expression of human endogenous retrovirus E envelope transcripts in human prostate adenocarcinoma. *Cancer*. 2003;98(1):187–197.
12. Yoshida M, Miyoshi I, Hinuma Y. Isolation and characterization of retrovirus from cell lines of human adult T-cell leukemia and its implication in the disease. *Proc Natl Acad Sci U S A*. 1982;79(6):2031–2035.
13. Kalyanaraman VS, Sarngadharan MG, Robert-Guroff M, Miyoshi I, Golde D, Gallo RC. A new subtype of human T-cell leukemia virus (HTLV-II) associated with a T-cell variant of hairy cell leukemia. *Science*. 1982;218(4572):571–573.
14. Florl AR, Löwer R, Schmitz-Dräger BJ, Schulz WA. DNA methylation and expression of LINE-1 and HERV-K provirus sequences in urothelial and renal cell carcinomas. *Br J Cancer*. 1999;80(9):1312–1321.
15. Rooney MS, Shukla SA, Wu CJ, Getz G, Hacohen N. Molecular and genetic properties of tumors associated with local immune cytolytic activity. *Cell*. 2015;160(1-2):48–61.
16. Haase K, Mösch A, Frishman D. Differential expression analysis of human endogenous retroviruses based on ENCODE RNA-seq data. *BMC Med Genomics*. 2015;8:71.
17. Paces J, Pavlícek A, Zika R, Kapitonov VV, Jurka J, Paces V. HERVd: the Human Endogenous RetroViruses Database: update. *Nucleic Acids Res*. 2004;32(Database issue):D50.
18. Paces J, Pavlícek A, Paces V. HERVd: database of human endogenous retroviruses. *Nucleic Acids Res*. 2002;30(1):205–206.
19. Tongyoo P, Avihingsanon Y, Prom-On S, Mutirangura A, Mhuantong W, Hirankarn N. EnHERV: enrichment analysis of specific human endogenous retrovirus patterns and their neighboring genes. *PLoS One*. 2017;12(5):e0177119.
20. Levy A, Sela N, Ast G. TranspoGene and microTranspoGene: transposed elements influence on the transcriptome of seven vertebrates and invertebrates. *Nucleic Acids Res*. 2008;36(Database issue):D47–D52.
21. Smit A, Hubley R, Green P. RepeatMasker Open-4.0. 2013–2015. Institute for Systems Biology. http://repeatmasker. org. Accessed September 11, 2018.
22. Loayza-Puch F, et al. Tumour-specific proline vulnerability uncovered by differential ribosome codon reading. *Nature*. 2016;530(7591):490–494.
23. Lavie L, Kitova M, Maldener E, Meese E, Mayer J. CpG methylation directly regulates transcriptional activity of the human endogenous retrovirus family HERV-K(HML-2). *J Virol*. 2005;79(2):876–883.
24. Okada M, et al. Role of DNA methylation in transcription of human endogenous retrovirus in the pathogenesis of systemic lupus erythematosus. *J Rheumatol*. 2002;29(8):1678–1682.
25. Stengel S, Fiebig U, Kurth R, Denner J. Regulation of human endogenous retrovirus-K expression in melanomas by CpG methylation. *Genes Chromosomes Cancer*. 2010;49(5):401–411.
26. Chiappinelli KB, et al. Inhibiting DNA Methylation causes an interferon response in cancer via dsRNA including endogenous retroviruses. *Cell*. 2015;162(5):974–986.
27. Goel S, et al. CDK4/6 inhibition triggers anti-tumour immunity. *Nature*. 2017;548(7668):471–475.
28. Chan KS, et al. Identification, molecular characterization, clinical prognosis, and therapeutic targeting of human bladder tumor-initiating cells. *Proc Natl Acad Sci U S A*. 2009;106(33):14016–14021.
29. Prat A, et al. Phenotypic and molecular characterization of the claudin-low intrinsic subtype of breast cancer. *Breast Cancer Res*. 2010;12(5):R68.
30. Iglesia MD, et al. Prognostic B-cell signatures using mRNA-seq in patients with subtype-specific breast and ovarian cancer. *Clin Cancer Res*. 2014;20(14):3818–3829.
31. Kardos J, et al. Claudin-low bladder tumors are immune infiltrated and actively immune suppressed. *JCI Insight*. 2016;1(3):e85902.
32. Bindea G, et al. Spatiotemporal dynamics of intratumoral immune cells reveal the immune landscape in human cancer. *Immunity*. 2013;39(4):782–795.
33. Palmer C, Diehn M, Alizadeh AA, Brown PO. Cell-type specific gene expression profiles of leukocytes in human peripheral blood. *BMC Genomics*. 2006;7:115.
34. Hugo W, et al. Genomic and transcriptomic

features of response to anti-PD-1 therapy in metastatic melanoma. *Cell*. 2016;165(1):35–44.
35. Cherkasova E, et al. Detection of an immunogenic HERV-E envelope with selective expression in clear cell kidney cancer. *Cancer Res*. 2016;76(8):2177–2185.
36. Takahashi Y, et al. Regression of human kidney cancer following allogeneic stem cell transplantation is associated with recognition of an HERV-E antigen by T cells. *J Clin Invest*. 2008;118(3):1099–1109.
37. Cherkasova E, et al. Inactivation of the von Hippel-Lindau tumor suppressor leads to selective expression of a human endogenous retrovirus in kidney cancer. *Oncogene*. 2011;30(47):4697–4706.
38. Liberzon A, Subramanian A, Pinchback R, Thorvaldsdóttir H, Tamayo P, Mesirov JP. Molecular signatures database (MSigDB) 3.0. *Bioinformatics*. 2011;27(12):1739–1740.
39. Şenbabaoğlu Y, et al. Erratum to: Tumor immune microenvironment characterization in clear cell renal cell carcinoma identifies prognostic and immunotherapeutically relevant messenger RNA signatures. *Genome Biol*. 2017;18(1):46.
40. Sauter M, et al. Human endogenous retrovirus K10: expression of Gag protein and detection of antibodies in patients with seminomas. *J Virol*. 1995;69(1):414–421.
41. Bolotin DA, et al. MiXCR: software for comprehensive adaptive immunity profiling. *Nat Methods*. 2015;12(5):380–381.
42. Hancock DC, O'Reilly NJ. Synthetic peptides as antigens for antibody production. *Methods Mol Biol*. 2005;295:13–26.
43. Ljungberg B, et al. EAU guidelines on renal cell carcinoma: 2014 update. *Eur Urol*. 2015;67(5):913–924.
44. Liu J, et al. An integrated TCGA pan-cancer clinical data resource to drive high-quality survival outcome analytics. *Cell*. 2018;173(2):400–416.e11.
45. Schumacher TN, Schreiber RD. Neoantigens in cancer immunotherapy. *Science*. 2015;348(6230):69–74.
46. Ott PA, et al. An immunogenic personal neoantigen vaccine for patients with melanoma. *Nature*. 2017;547(7662):217–221.
47. Wang RF, Wang HY. Immune targets and neoantigens for cancer immunotherapy and precision medicine. *Cell Res*. 2017;27(1):11–37.
48. Kreiter S, et al. Mutant MHC class II epitopes drive therapeutic immune responses to cancer. *Nature*. 2015;520(7549):692–696.
49. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014;15(12):550.
50. Sahin U, et al. Personalized RNA mutanome vaccines mobilize poly-specific therapeutic immunity against cancer. *Nature*. 2017;547(7662):222–226.
51. Michel AM, et al. GWIPS-viz: development of a ribo-seq genome browser. *Nucleic Acids Res*. 2014;42(Database issue):D859–D864.
52. Jurtz V, Paul S, Andreatta M, Marcatili P, Peters B, Nielsen M. NetMHCpan-4.0: improved peptide-MHC class I interaction predictions integrating eluted ligand and peptide binding affinity data. *J Immunol*. 2017;199(9):3360–3368.
53. Altman JD, et al. Phenotypic analysis of antigen-specific T lymphocytes. *Science*. 1996;274(5284):94–96.
54. Rodenko B, et al. Generation of peptide-MHC class I complexes through UV-mediated ligand exchange. *Nat Protoc*. 2006;1(3):1120–1132.
55. Toebes M, et al. Design and use of conditional MHC class I ligands. *Nat Med*. 2006;12(2):246–251.
56. Bakker AH, et al. Conditional MHC class I ligands and peptide exchange technology for the human MHC gene products HLA-A1, -A3, -A11, and -B7. *Proc Natl Acad Sci U S A*. 2008;105(10):3825–3830.
57. Protocol for fluorescent Flex-T™ generation and antigen specific CD8+ T cell staining. https://www.biolegend.com/media_assets/support_protocol/Protocol%20for%20fluorescent%20tetramer%20generation%20and%20cell%20staining%2006202016.pdf. Revised June 20, 2016. Accessed September 11, 2018.
58. Protocol for HLA class I ELISA to evaluate peptide exchange. http://www.biolegend.com/media_assets/flex-t/Protocol_for_HLA_class_I_ELISA_05272016.pdf. Revised May 23, 2016. Accessed September 11, 2018.
59. Iglesia MD, Parker JS, Hoadley KA, Serody JS, Perou CM, Vincent BG. Genomic analysis of immune cell infiltrates across 11 tumor types. *J Natl Cancer Inst*. 2016;108(11):djw144.
60. Dobin A, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*. 2013;29(1):15–21.
61. Li H, et al. The sequence alignment/map format and SAMtools. *Bioinformatics*. 2009;25(16):2078–2079.
62. Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. Salmon provides fast and bias-aware quantification of transcript expression. *Nat Methods*. 2017;14(4):417–419.
63. Dolton G, et al. More tricks with tetramers: a practical guide to staining T cells with peptide-MHC multimers. *Immunology*. 2015;146(1):11–22.
64. Bodenhofer U, Bonatesta E, Horejš-Kainrath C, Hochreiter S. msa: an R package for multiple sequence alignment. *Bioinformatics*. 2015;31(24):3997–3999.