

Endogenous Retrovirus Transcript Levels Are Associated with Immunogenic Signatures in Multiple Metastatic Cancer Types



James T. Topham¹, Emma Titmuss², Erin D. Pleasance², Laura M. Williamson², Joanna M. Karasinska¹, Luka Culibrk², Michael K.C. Lee³, Shehara Mendis³, Robert E. Denroche⁴, Gun-Ho Jang⁴, Steve E. Kalloger^{1,5}, Hui-Li Wong³, Richard A. Moore², Andrew J. Mungall², Grainne M. O'Kane⁶, Jennifer J. Knox⁶, Steven Gallinger^{4,6}, Jonathan M. Loree³, Dixie L. Mager^{7,8}, Janessa Laskin^{2,3}, Marco A. Marra^{2,8}, Steven J.M. Jones^{2,8}, David F. Schaeffer^{1,5,9}, and Daniel J. Renouf^{1,3,10}

ABSTRACT

Next-generation sequencing of solid tumors has revealed variable signatures of immunogenicity across tumors, but underlying molecular characteristics driving such variation are not fully understood. Although expression of endogenous retrovirus (ERV)-containing transcripts can provide a source of tumor-specific neoantigen in some cancer models, associations between ERV levels and immunogenicity across different types of metastatic cancer are not well established. We performed bioinformatics analysis of genomic, transcriptomic, and clinical data across an integrated cohort of 199 patients with metastatic breast, colorectal, and pancreatic ductal

adenocarcinoma tumors. Within each cancer type, we identified a subgroup of viral mimicry tumors in which increased ERV levels were coupled with transcriptional signatures of autonomous antiviral response and immunogenicity. In addition, viral mimicry colorectal and pancreatic tumors showed increased expression of DNA demethylation gene *TET2*. Taken together, these data demonstrate the existence of an ERV-associated viral mimicry phenotype across three distinct metastatic cancer types, while indicating links between ERV abundance, epigenetic dysregulation, and immunogenicity.

Introduction

Negative modulation of the immunogenic potential of tumor cells is recognized as a key step in cancer development and progression (1). Immunotherapeutic regimes such as immune checkpoint blockade aim to harness the endogenous antitumor response machinery to eliminate tumor cells with high specificity (2). Previous studies have demonstrated that the efficacy of many cancer immunotherapies is limited to a subset of patients (3). Tumor mutation burden (TMB) has been shown to be a strong predictor of immunotherapy response in cancer types with high mutation rates such as melanoma (4, 5) and non-small cell

lung cancer (6). However, TMB is not predictive of immunotherapy response in many cancer types with lower mutation rates such as pancreatic, breast, and mismatch repair intact colorectal cancers (7), driving the need for additional biomarkers of immunotherapy sensitivity that are consistent across a greater range of cancer types.

Endogenous retroviruses (ERV) are long-terminal repeat (LTR)-containing genomic elements that are the remnants of ancestral infections and retrotranspositions (8). ERV-related sequences and solitary LTRs encompass approximately 8% of the human genome (9) and are generally epigenetically repressed in adult human tissues (10, 11). However, some ERV genomic loci can be transcribed, either autonomously (i.e., from their own LTR promoter) or, more frequently, as parts of other transcripts, for example, in gene 3' untranslated regions (12) or in long noncoding RNAs (lncRNA; ref. 13). In this study, we will hereafter use the term "ERV levels" to refer to the abundance of ERV-containing transcripts, as many such transcripts may not represent autonomous ERV transcription.

In vitro studies have demonstrated an association between increased ERV levels and activation of innate antiviral pathways, resembling a response against viral double-stranded RNA (dsRNA) consistent with that of virus-infected cells (11, 14). Increased ERV levels have been correlated with immunotherapy response in urothelial (15) and renal cell (16) cancers, thereby positioning ERVs as an alternative source of tumor-specific neoantigen. However, the relevance of ERVs to tumor biology and immunogenicity has not been comprehensively explored in other cancer types.

The Personalized Oncogenomics (POG; NCT02155621) program of British Columbia Cancer aims to identify the impact of incorporating whole-genome and transcriptome sequencing into treatment decision-making for patients with advanced cancer (17). Data generated by the POG program provide a unique resource for retrospectively studying ERV levels in metastatic cancers, as noncoding RNAs can be detected in whole-transcriptome sequencing data.

¹Pancreas Centre BC, Vancouver, British Columbia, Canada. ²Canada's Michael Smith Genome Sciences Centre, Vancouver, British Columbia, Canada. ³Division of Medical Oncology, BC Cancer, Vancouver, British Columbia, Canada. ⁴Ontario Institute for Cancer Research, Toronto, Ontario, Canada. ⁵Department of Pathology and Laboratory Medicine, University of British Columbia, Vancouver, British Columbia, Canada. ⁶University Health Network, University of Toronto, Toronto, Ontario, Canada. ⁷Terry Fox Laboratory, BC Cancer, Vancouver, British Columbia, Canada. ⁸Department of Medical Genetics, University of British Columbia, Vancouver, British Columbia, Canada. ⁹Division of Anatomic Pathology, Vancouver General Hospital, Vancouver, British Columbia, Canada. ¹⁰Department of Medicine, University of British Columbia, Vancouver, British Columbia, Canada.

Note: Supplementary data for this article are available at Molecular Cancer Therapeutics Online (<http://mct.aacrjournals.org/>).

D.F. Schaeffer and D.J. Renouf contributed equally to this article.

Corresponding Author: Daniel J. Renouf, BC Cancer, 600 West 10th Ave., Vancouver, British Columbia V5Z 4E6, Canada. Phone: 604-877-6000, ext. 672445; Fax: 604-877-0585; E-mail: drenouf@bccancer.bc.ca

Mol Cancer Ther 2020;19:1889-97

doi: 10.1158/1535-7163.MCT-20-0094

©2020 American Association for Cancer Research.

Here, we perform a comprehensive analysis of ERV elements in a collection of 199 patients with metastatic breast, colorectal, and pancreatic ductal adenocarcinoma (PDAC) tumors. Bioinformatics analysis of 393,507 ERV loci revealed distinct patterns of ERV abundance in each of the three cancer types. Within each cancer type, we found a significant association between ERV levels and signatures of antiviral response and adaptive immunogenicity. *TET2* encodes the primary DNA demethylase in mammalian cells, and has been shown to mediate increased ERV levels in cancer cell lines treated with DNA demethylating agents (18). For colorectal and PDAC tumors showing elevated ERV levels concomitant with signatures of antiviral response, we observed significant increases in *TET2* mRNA. Overall, our data provide a novel exploration of ERV elements in RNA sequencing (RNA-seq) data of advanced cancers, and support associations between ERV levels, epigenetic dysregulation and immunogenicity in advanced cancers.

Materials and Methods

Patient enrollment

Patients were enrolled in the POG (NCT02155621) or PanGen (NCT02869802) clinical trials at British Columbia Cancer (British Columbia, Canada). The study was approved by the University of British Columbia Research Ethics Committee (REB, catalog nos. H12-00137, H14-00681) and was conducted in accordance with international ethical guidelines. Written informed consent was obtained from each patient prior to genomic profiling. Patients consented to potential publication of findings. Raw sequence data and downstream analytics were maintained within a secure computing environment.

Whole transcriptome library preparation and sequencing

Tumor RNA was extracted, polyadenylated RNA purified, and RNA-seq libraries constructed as described previously (19). Sequencing was performed on Illumina HiSeq and HiSeqX instruments using v3 or v4 chemistry to a target depth of 150–200 million reads.

Computational prediction of tumor content

A novel statistical modeling algorithm was used to determine tumor purity from whole-genome sequencing data. Briefly, the germline genomes were divided into segments with equal coverage at a threshold. Next, somatic read counts within these bins were enumerated and clusters were identified by univariate density approximation following kernel density estimation. Each cluster corresponded to a distinct copy-number state within the genome. We generated models that described alternate copy-number fits for each set of clusters. We then fitted a linear model to each set of selected clusters and computed the penalty for each model. The penalty is an empirically derived function of the SE of fit, proportion of data explained by fit, and the relative proportion of predicted versus observed copy-number states. Next, allelic frequencies of heterozygous germline SNPs within the selected clusters were used to determine the absolute copy number of each cluster. Assuming that all reads mapping to homozygously deleted regions originate from normal tissue contamination, the number of reads mapping to homozygous regions was proportional to the normal contamination. This value was used to compute tumor content. When tumor content was not able to be predicted computationally, the pathology-measured value was used.

RNA-seq analysis

Paired-end RNA-seq reads were aligned to human reference genome hg19 (GRCh37-lite) using Burrows-Wheeler Aligner

v.0.7.6 (20) and JAGuaR v.2.2 (21). Raw read counts were assigned to Ensembl 75 genes using Subread v.1.4.6 (22) with parameters `-T 3 -s 1 -C -t "gene,"` and default parameters otherwise. Gene expression values were normalized for gene size and sequencing depth calculated as reads per thousand base pair per million mapped reads (RPKM).

Batch correction

RNA-seq libraries could be stratified into two technical batches based on the library preparation protocol that was used, with BRCA and COAD cohorts each containing two batches. For BRCA and COAD samples, we corrected normalized gene expression values (RPKM) for these batches using a location scaling approach (23), in which z-scores were calculated for each gene within each batch. Principal component analysis (PCA) of the top 25% most variable genes demonstrated successful correction of batches using this approach (see Supplementary Fig. S1, Additional File 1).

ERV analysis

A list of 5,467,457 repeat elements and their hg19 coordinates were downloaded from RepeatMasker Open v.4.0.5 (24) on March 12, 2018. These were filtered for ERVs (702,533) on chromosomes {1–22, X} ($n = 685,058$). Raw expression values were calculated for each ERV by counting the number reads that mapped unambiguously (mates mapped within 10 kb) to each region. A threshold insert size of 10 kb was chosen to account for the previously reported distribution of intron sizes in lncRNAs (25). ERVs were filtered for those located at least 10 kb away from their nearest UCSC gene ($n = 393,507$). ERV levels were normalized for sample-specific sequencing depth and ERV loci coordinates. ERV size was based on RepeatMasker genome coordinates for all ERVs, including those located within a lncRNA.

A list of 106,958 lncRNAs and their hg19 coordinates were obtained from LNCipedia (26; high confidence set v.5.2). ERVs were considered to be contained within a lncRNA if their coordinates overlapped with a lncRNA transcript region by at least 1 bp.

Each tumor type was considered a separate RNA-seq batch. Colorectal and breast cancer samples were generated using one of two separate RNA-seq library protocols, further segregating these tumor types into an additional two batches each. To standardize ERV loads for each of these five batches, ERV loads were converted to z-scores within each batch. High ERV samples were defined as having an ERV load greater than the mean ERV load for their respective batch ($z\text{-score} > 0$). For antiviral response scores, normalized gene expression values for genes belonging to the “GO_response_to_dsRNA” gene set (27) were averaged (mean) for each sample. High antiviral response samples were defined as having an average antiviral response expression greater than the mean (mean of means).

Differential expression and enrichment analysis

Within each cancer type, differential expression analysis (DEA) was performed between VMP and all other samples using DESeq2 v.1.14.1 (28), with the design formula “ \sim batch + VMP” for BRCA and COAD, and “ \sim VMP” for PDAC. Gene set enrichment analysis (GSEA) was performed on upregulated and downregulated ($P_{\text{adjusted}} < 0.05$) genes in each cancer type using hypergeometric tests followed by Benjamini-Hochberg (BH) test correction in R, and included all gene ontology (GO) gene sets ($n = 6,169$) provided by mSigDB (27).

Survival analysis

Kaplan–Meier plots were generated using R packages “survival” v.2.4.2 (29) and “survminer” v.0.4.2 (30). *P* values were calculated on

the basis of a log-rank test. Cox proportional hazards tests were conducted in R.

Statistical analysis

All statistics were performed using R v.3.3.2 (31). All Wilcoxon tests were two-tailed. Spearman correlations were used to test for association between two continuous variables. Wilcoxon rank-sum tests were used when comparing continuous values between two discrete groups, and Wilcoxon signed-rank tests were used when testing whether the distribution of a single continuous variable was symmetric about zero. Fisher exact test was used when performing contingency analyses. Whenever multiple comparisons were performed, *P* values were corrected using BH multiple hypothesis correction. All significance tests were assessed using an alpha = 0.05.

Ethics approval and consent to participate

The study was approved by the University of British Columbia Research Ethics Committee (REB, catalogue numbers H12-00137 and H14-00681) and was conducted in accordance with international ethical guidelines. Written informed consent was obtained from each patient prior to genomic profiling. Patient identity was anonymized and an identification code assigned to the case for communicating clinically relevant information to physicians. Patients consented to potential publication of findings. Raw sequence data and downstream analytics were maintained within a secure computing environment.

Availability of data and materials

RNA-seq datasets including metadata with library construction and sequencing approaches have been deposited at the European Genome-phenome Archive (EGA; <http://www.ebi.ac.uk/ega/>) under study accession number EGAS00001001159. Accession numbers for each individual sample are listed in Supplementary Table S5, Additional File 5.

Results

ERV elements detected by RNA-seq show tumor-specific patterns

To investigate ERV levels across distinct metastatic tumor types, we leveraged whole-transcriptome RNA-seq data generated by the POG program (17). Our analysis included breast (BRCA), colorectal (COAD), and pancreatic ductal adenocarcinoma (PDAC) samples (*n* = 101, 55 and 33, respectively; Fig. 1A). Metastatic or locally advanced lesions from a variety of biopsy sites (see Supplementary Table S1, Additional File 1) were sequenced between 2014 and 2018 and covered a wide range of tumor content [BRCA: 21.4% (min), 63.6% (median), 95.2% (max); COAD: 16.3% (min), 50.0% (median), 86.3% (max); PDAC: 14.5% (min), 45.5% (median), 79.8% (max)]. ERV coordinates were downloaded from the RepeatMasker database (24). To minimize the extent to which ERV read counts were biased by expression of nearby protein-coding genes, we filtered for ERVs located at least 10 kb away from their nearest gene, leaving 393,507 ERVs for analysis.

At the time of this study, it remains unclear to what extent ERV element abundance in RNA-seq data is a result of transcriptional hitch-hiking within other transcripts. Upregulation of ERV-containing lncRNA transcripts, either promoted by LTRs or other promoters, have been identified in cancer (13, 32), indicating that a proportion of cancer-specific ERV elements reside within lncRNAs. To investigate the proportion of ERV elements that fall within lncRNA transcript boundaries, we calculated the distance between each ERV

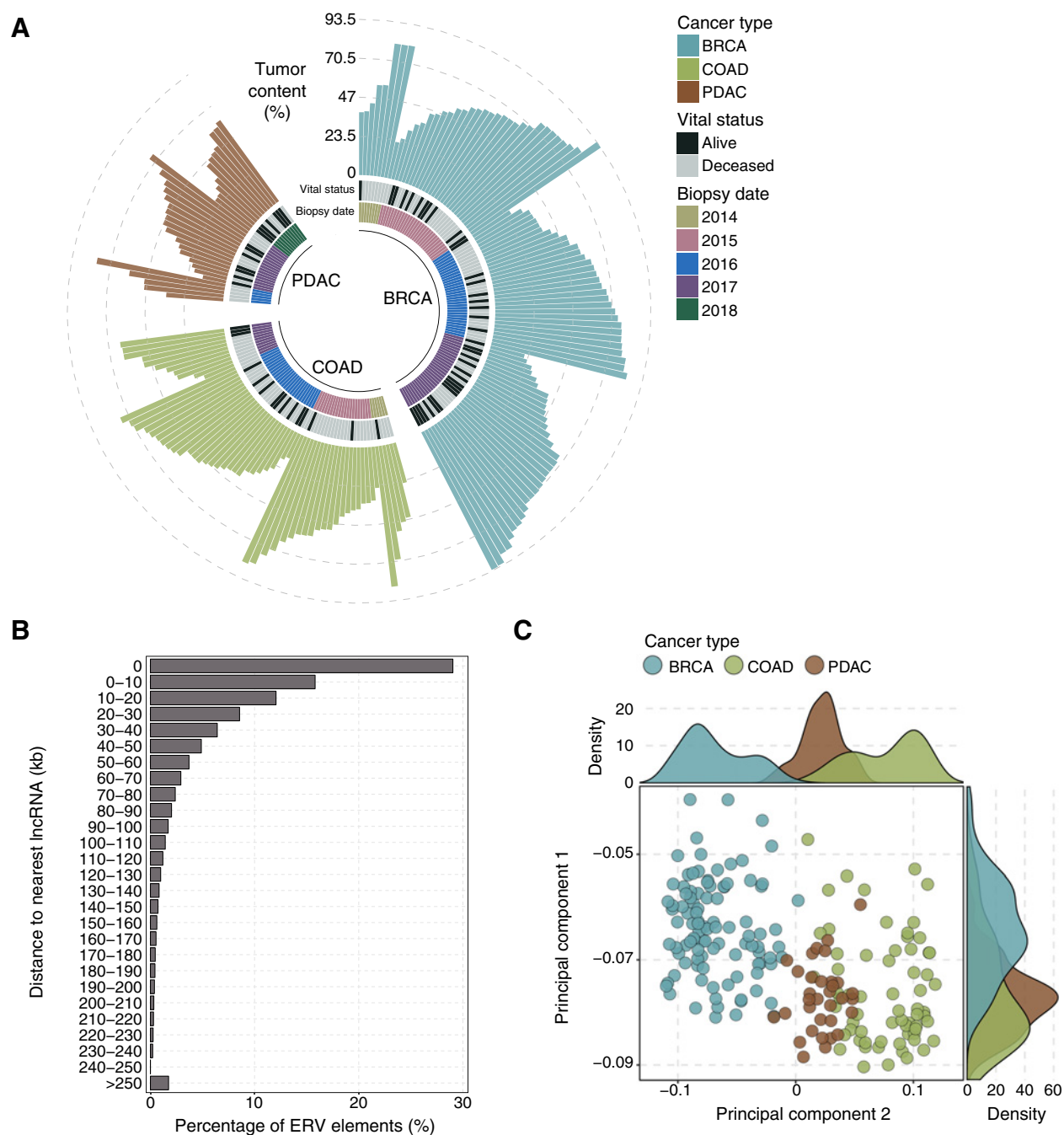
element and its nearest lncRNA (26). 114,047/393,507 (29.0%) of ERV elements overlapped with a lncRNA transcript region (Fig. 1B), indicating that a proportion of ERV sequences detected in RNA-seq data are likely a result of transcription of the lncRNA they reside in, rather than being autonomously transcribed with an ERV-specific promoter. It is also likely a proportion of transcribed ERVs are acting to promote transcription of nearby lncRNA (32, 33). Whether the immunogenic potential of an ERV depends on its method of transcription remains an open question, and we hereafter consider all ERV sequences equally in our analyses.

To explore levels of specific ERV elements across cancer types, we performed PCA on ERV elements that were most variable across all samples (top 5%; 19,676 ERVs). The first and second principal components captured 53.0% and 18.5% of the variance in the dataset, respectively, and variation between samples aligned to tumor type, particularly across the second principal component (Fig. 1C). This observation is compatible with the notion that the abundance of specific ERV elements is tissue-specific (34).

High-ERV tumors show transcriptional patterns indicative of autonomous antiviral response and increased immunogenicity

Increased ERV levels have been associated with increases in the expression of antiviral response genes in cancer cell lines (11, 18, 35, 36). These findings led to the proposed existence of a viral mimicry phenotype (VMP) of cancer cells, in which increased ERV levels lead to the accumulation of viral dsRNA in the cytoplasm and activation of antiviral response machinery in a manner consistent with a virally infected cell (37). We therefore hypothesized that ERV levels are positively correlated with the expression of genes involved in dsRNA response pathways in advanced cancers. To test this, we defined the ERV load of each sample by the sum of all normalized ERV read counts, and adjusted ERV loads for differences in RNA-seq protocols within each cancer type by converting to z-scores (23). We defined the antiviral expression score of each sample as the mean expression level across the set of 72 genes contained in the GO “response to dsRNA” gene set (27). Samples were then stratified into four groups based on their relative ERV and antiviral scores: ERV-low, high ERV-only, high antiviral-only, and VMP (Fig. 2A). ERV levels were positively correlated with antiviral scores in COAD ($\rho = 0.50$, $P = 2.7 \times 10^{-5}$), PDAC ($\rho = 0.58$, $P = 0.0005$), and BRCA ($\rho = 0.28$, $P = 0.005$) samples (Fig. 2B). We noted differences in tumor content only in the high antiviral-only group of the BRCA cohort, which showed significantly lower tumor content values compared with the rest of the BRCA cohort (BH-adjusted Wilcoxon rank sum $P = 0.047$; see Supplementary Fig. S2, Additional File 1). Furthermore, a direct comparison between tumor content and ERV levels yielded no significant correlation between the two (Supplementary Fig. S3, Additional File 1). ERV levels and group frequencies were similar across the various tumor biopsy sites (see Supplementary Fig. S4, Additional File 1).

While ERV levels have been associated with increased survival for patients with cancer receiving immunotherapy (15, 16), ERV levels have not shown prognostic value for patients receiving standard therapy (16). The POG cohort of patients with advanced cancers encompasses a variety of treatment regimens, the majority of which are standard-of-care therapies. We found no significant (log-rank $P < 0.05$) survival differences across the ERV/antiviral expression groups in any of the metastatic cancer cohorts (Fig. 2C). Cox proportional hazards tests, within each cancer type, also supported no significant differences (likelihood ratio test $P > 0.05$) between any of the four ERV/antiviral expression groups.

**Figure 1.**

Individual ERV elements show tumor type-specific patterns of abundance. **A**, Circular plot summarizing the BRCA, COAD, and PDAC cohorts, ordered by date of biopsy and tumor content, demonstrating heterogeneity in these features. **B**, Bar plot depicting the distribution of distances between each ERV element and its nearest lncRNA transcript region. Nearly 30% of all ERV loci are located within lncRNA transcript regions. **C**, Scatter plot showing PCA of the top 25% most variable ERVs ($n = 19,676$) across the three cancer types. Relative abundance of individual ERV elements are tissue-type specific.

Increased ERV levels have been associated with signatures of heightened immunogenicity in primary tumors (16), and the ability for ERV levels to predict response to immunotherapy has been attributed to their potential proimmunogenic capability (37). To investigate the association between ERV levels and immunogenicity in metastatic tumors, we calculated Spearman correlations between

ERV load and expression values for individual genes ($n = 550$) that were associated with infiltrative immune cell types and related pathways, as provided by ImSig (38). In each cancer type, Wilcoxon signed-rank tests were performed on Spearman rho values for each cell type/pathway, and P values were subjected to BH multiple test correction. All cancer types demonstrated significantly high positive correlations

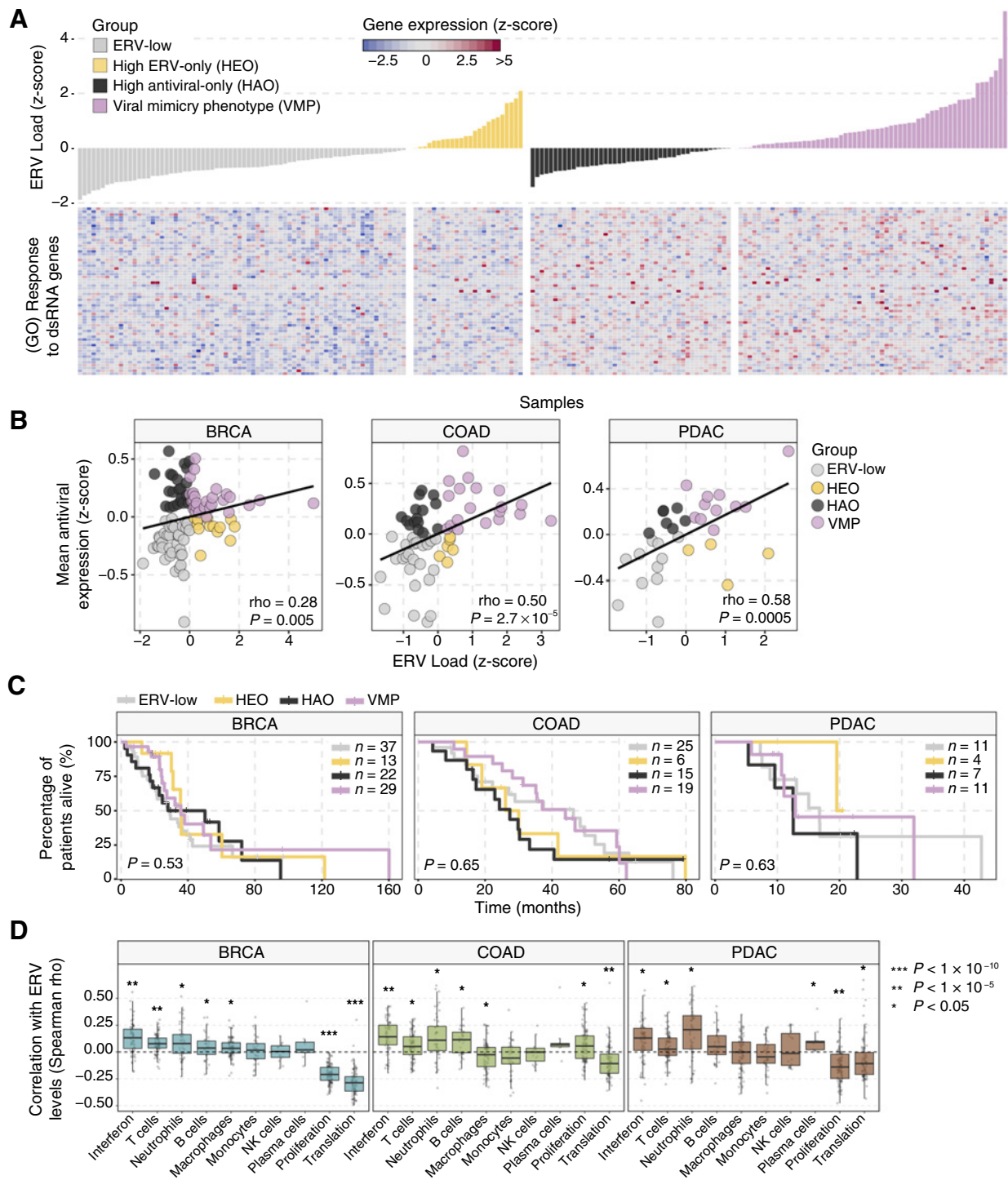


Figure 2.

ERV-high tumors show signatures of autonomous antiviral response and immunogenicity. **A**, Heatmap of expression values for genes involved in dsRNA response (z-scores; bottom) aligned with total ERV load (z-scores; top) across all tumor types. Samples are stratified into four groups based on relative ERV and antiviral scores. **B**, Scatter plot demonstrating significant positive correlation (Spearman $\rho > 0$, $P < 0.05$) between ERV load and mean antiviral expression. **C**, Kaplan-Meier curves showing overall survival patterns across the four ERV/antiviral groups in each cancer type. ERV/antiviral groups are not associated with overall survival. **D**, Boxplots depicting the distribution of correlation values (Spearman ρ) between total ERV load and genes involved in various immunogenic and related pathways in each cancer type. Both adaptive (T-cell) and innate (neutrophil) pathways are positively correlated with ERV load in all three cancer types.

Downloaded from <http://aacrjournals.org/mct/article-pdf/19/9/1899/1899880/1899.pdf> by guest on 05 October 2024

between ERV levels and genes associated with T cells (BRCA, $P = 9.4 \times 10^{-10}$; COAD, $P = 3.0 \times 10^{-4}$; PDAC, $P = 6.9 \times 10^{-3}$) and neutrophils (BRCA, $P = 9.5 \times 10^{-4}$; COAD, $P = 2.2 \times 10^{-4}$; PDAC, $P = 4.6 \times 10^{-5}$), indicating an association between ERV load and both adaptive and innate immune pathways, respectively (Fig. 2D). Furthermore, we observed positive correlations between ERV load and genes involved in IFN signaling (BRCA, $P = 1.3 \times 10^{-7}$; COAD, $P = 1.1 \times 10^{-7}$; PDAC, $P = 2.3 \times 10^{-4}$), and negative correlations for genes associated with translation (BRCA, $P = 9.3 \times 10^{-24}$; COAD, $P = 1.4 \times 10^{-7}$; PDAC, $P = 6.1 \times 10^{-5}$). As protein translation is known to be attenuated in virus-infected cells as a mechanism by which host cells limit translation of viral RNA (39, 40), these results further strengthen the association between ERV levels and viral mimicry in metastatic tumors. We also noted a positive correlation between ERV load and genes involved in proliferation in COAD (mean $\rho = 0.06$, $P = 1.0 \times 10^{-3}$), while proliferation genes tended to be negatively correlated with ERV load in BRCA (mean $\rho = -0.21$, $P = 2.5 \times 10^{-16}$) and PDAC (mean $\rho = -0.13$, $P = 1.9 \times 10^{-10}$), indicating that some factors associated with increased ERV load are tumor-type specific.

VMP tumors show increased transcription of *TET2* in colorectal and PDAC tumors

To comprehensively explore transcriptional pathways associated with VMP tumors, we performed DEA between VMP tumors and all other tumors for each cancer type. BRCA, COAD, and PDAC VMP tumors showed upregulation [\log_2 fold change (L2FC) > 0, $P_{\text{adjusted}} < 0.05$] of 112; 1,264; and 447 genes and downregulation (L2FC < 0, $P_{\text{adjusted}} < 0.05$) of 312; 1,203; and 393 genes, respectively (Fig. 3A; see Supplementary Table S2, Additional File 2). We noted significant upregulation of cell-cycle regulator *TP63* in VMP tumors exclusively in COAD (L2FC = 1.4, $P_{\text{adjusted}} = 3.2 \times 10^{-8}$), and GSEA of upregulated genes within each cancer type revealed significant enrichment (hypergeometric BH-adjusted $P < 0.05$) of cell cycle-related pathways in the COAD cohort, while no gene sets were significantly enriched among BRCA and PDAC (see Supplementary Table S3, Additional File 3 and Supplementary Table S4, Additional File 4). These data highlight cell-cycle regulation as a tissue type-specific correlate of viral mimicry in metastatic colorectal cancers.

ERV-containing transcripts often exhibit a transcriptionally silenced state in adult tissues through various epigenetic factors such as DNA methylation (10). DNA demethylating agents such as 5-aza-2-deoxycytosine and vitamin C have been shown to increase ERV levels and consequent IFN response signaling in cancer cell lines (11, 18, 35), and ERV levels were found to be correlated with chromatin regulation genes in primary tumor samples (16). To investigate the relationship between epigenetic dysregulation and viral mimicry in metastatic cancers, we filtered DEA results for genes significantly dysregulated ($P_{\text{adjusted}} < 0.05$) in at least two of the three cancer types and for members of GO gene sets “chromatin organization” and “chromatin modification.” The resulting 20 genes are shown in Fig. 3B and include *TET2*, which encodes the primary DNA demethylating enzyme in mammalian adult tissues (41). While *TET2* has been shown to mediate the effects of DNA demethylating agents on ERV levels in cancer cell lines (18), the relationship between *TET2* and ERV levels in patient tumor samples has remained an open question. Our observation of significant increases in the expression of *TET2* in VMP samples in COAD (L2FC = 0.39, $P_{\text{adjusted}} = 0.007$) and PDAC (L2FC = 0.39, $P_{\text{adjusted}} = 0.01$; Fig. 3C) is compatible with the notion that DNA demethylation plays a role in the transcriptional activation of ERV-containing transcripts in these tumor types.

Discussion

These results add to a growing body of evidence indicating the existence of a viral mimicry subtype of tumors, in which elevated ERV levels are coupled with signatures of antiviral response pathways and immune activation. Moreover, we expand upon previous studies of ERV expression in cancer by performing a comprehensive analysis of ERV loci, leveraging RNA-seq data from patients with advanced breast, colorectal, and pancreatic cancers.

Patterns of loci-specific ERV levels were unique to each cancer type, indicating that a more generalized measurement of overall ERV levels, rather than focusing on specific ERV elements, may prove more robust across a variety of tissue types. We also noted that many entries in the RepeatMasker ERV database were fragments of the same ERV element, thereby further supporting the strategy of quantifying ERV levels through summation of all ERV loci. A limitation to this approach is the reliance on whole-transcriptome RNA-seq. Antibodies capable of quantifying dsRNA may serve as an alternative approach that is less limited by cost, computational resources, and sample quantity (11, 42).

Attenuation of protein translation represents a host response to viral infection (39, 40), and genes involved in translation were negatively associated with ERV levels in breast, colorectal, and PDAC samples. Despite translation being a limiting factor in the proliferative capacity of human cells (43), only breast and PDAC samples showed decreases in proliferation-associated genes in high-ERV samples, while expression of such genes was increased in high-ERV colorectal samples. We also noted a significant increase in *TP63* expression in VMP colorectal samples, which could account for the increased cell cycling/proliferation (44). These results generate the hypothesis that there exists a relationship between cell-cycle activation and ERV load in metastatic colorectal cancer. While *TP63* has been shown to be regulated by an adjacent LTR12 element in testicular cancer cells (45), we did not find any correlation between *TP63* expression and any nearby ERV/LTR elements in our data.

Previous studies of ERV levels in cancer have contributed to the model by which activation of ERV-containing transcripts is driven by changes in epigenetic regulation, primarily in the form of DNA hypomethylation (37). *TET2* encodes the primary DNA demethylation enzyme in mammalian cells (46), and we observed increased mRNA levels of *TET2* in colorectal and pancreatic viral mimicry tumors. This result extends previous studies of *TET2*-mediated increases in ERV levels in cancer cell lines (18) and provides novel evidence supporting this notion in metastatic tumor samples. The extent to which epigenetic modifiers contribute to activation of ERV-containing transcripts in breast cancer samples remains unclear.

We observed no significant differences in overall survival when stratifying patients by the four ERV tumor groups. The lack of survival benefit conferred by ERV-high tumors is consistent with previous studies (16) and aligns with the notion that the prognostic significance of ERV-associated immunogenicity is limited to patients receiving immunotherapy treatment (16). While the lack of patients receiving immunotherapy in our study cohorts restricts our ability to make inferences on the predictive strength of ERV quantification, our data are compatible with the notion that an ERV-associated immunogenic signature is present in a subset of patients with advanced colorectal, breast, and pancreatic cancer. Whether such patients are particularly sensitive to immunotherapy remains to be determined by future studies.

Our study is limited as RNA-seq data for normal tissues is not available for any of the samples analyzed. While several studies have

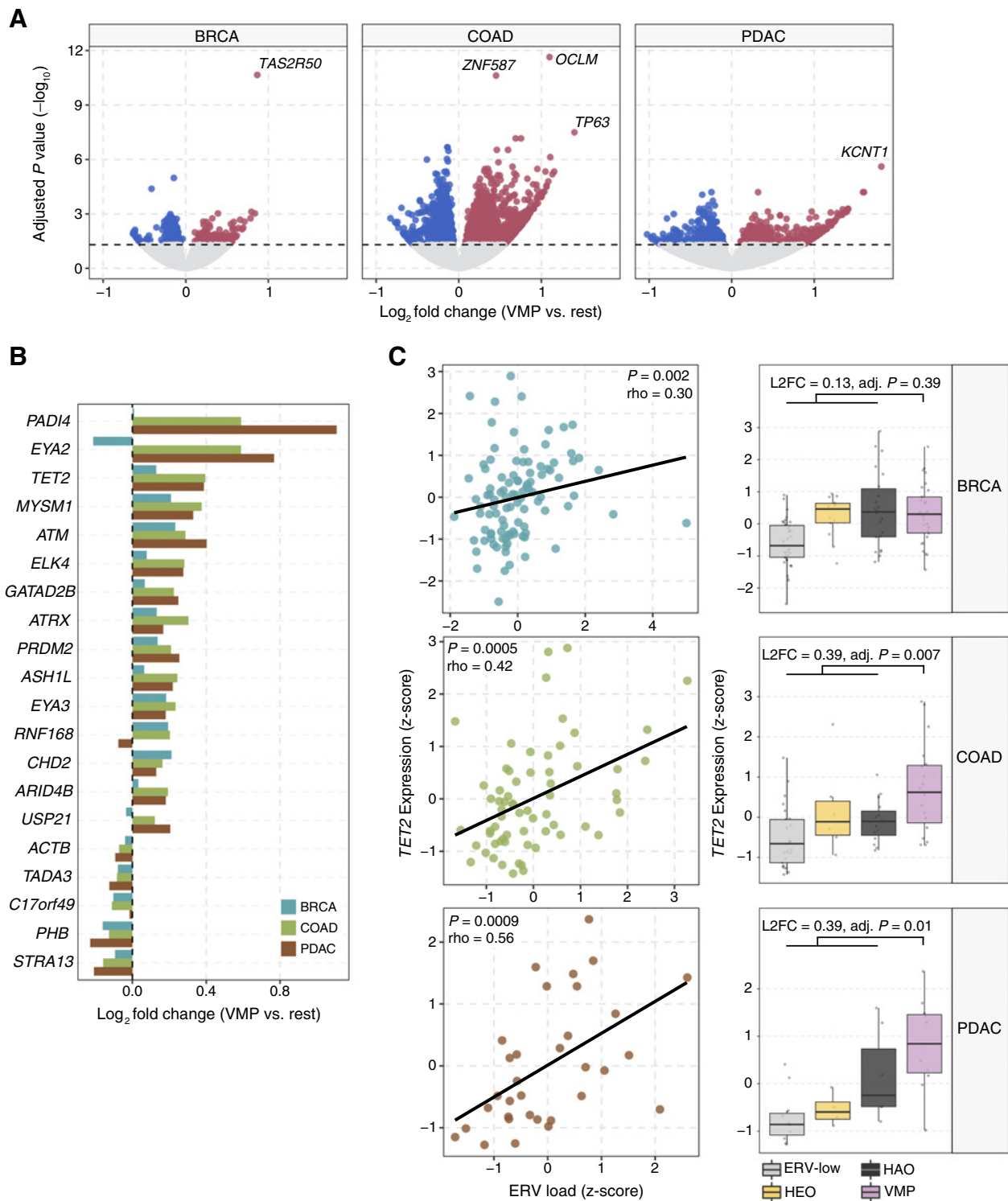


Figure 3.

VMP tumors show upregulation of *TET2* in COAD and PDAC. **A**, Volcano plots depicting results of DEA between VMP and all other samples in each cancer type. COAD samples show unique upregulation of *TP63*. **B**, Bar plot showing fold change differences between VMP and all other samples for each cancer type for 20 epigenetic regulation genes. **C**, Box and scatter plots demonstrating the correlation between ERV load and *TET2* expression in each cancer type (left), together with *TET2* expression values across the ERV/antiviral groups (right).

identified tumor-specific upregulation of specific ERV loci (47, 48), overall ERV levels may be comparable between normal and tumor cells (49). Here, we demonstrate that differences in the relative aggregate abundance of ERVs across tumor samples are biologically relevant. While we and others (18, 35) have provided results in support of the hypothesis that ERV transcripts arise from epigenetic dysregulation in tumors, future studies incorporating tumor-adjacent tissue samples will provide additional insight into whether elevated ERV transcription is indeed a tumor-specific phenomenon. Another limitation of our study is that our analysis is based on polyadenylated RNA-seq data, which has been shown to capture a subset of all ERV transcripts that would normally be captured by total RNA-seq approaches (15). However, our analysis was able to identify heterogeneity in ERV transcript levels that was correlated with multiple factors in support of a true ERV signal, namely, innate and adaptive immune signatures and *TET2* gene expression. We therefore consider our analysis to be evidence of the potential biological relevance provided by the subset of ERVs captured by polyadenylated RNA-seq. Finally, while our work builds on the hypothesis of previous studies in which ERV expression originates from tumor cells, we found ERV levels to be neither positively nor negatively correlated with tumor content. Tumor biopsy samples represent a mixture of cell types, and attempts to deduce the magnitude of ERV signal originating from various cell types lies out of the scope of this study. Future studies of tumor and nontumor cell-enriched samples, sequenced with the same RNA-seq protocol, will provide important insight into this matter.

In summary, our results are compatible with the existence of a viral mimicry subtype of metastatic tumors, and reveal similarities and differences in this subtype across three distinct tissue types. Importantly, the association between *TET2* expression and viral mimicry tumors provides further rationale toward clinically investigating the ability for DNA demethylating agents to drive viral mimicry in patients with colorectal and pancreatic cancer. Overall, our investigation of ERV transcripts in advanced cancers extends upon recent studies describing ERV-mediated immunogenicity in cancer cell lines and primary disease samples, and highlights overall ERV levels as a potential biomarker for immunotherapy response in patients with metastatic breast, colorectal, and pancreatic cancers.

Disclosure of Potential Conflicts of Interest

J. Laskin is an unpaid consultant/advisory board member for Roche, AstraZeneca, and Pfizer. D.J. Renouf receives speakers' bureau honoraria from Celgene, Roche, Bayer, Taiho, Servier, and Ipsen. No potential conflicts of interest were disclosed by the other authors.

References

- Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. *Cell* 2011;144:646–74.
- Martin-Liberal J, Ochoa de Olza M, Hierro C, Gros A, Rodon J, Taberner J. The expanding role of immunotherapy. *Cancer Treat Rev* 2017;54:74–86.
- Sambi M, Bagheri L, Szewczuk MR. Current challenges in cancer immunotherapy: multimodal approaches to improve efficacy and patient response rates. *J Oncol* 2019;2019:4508794.
- Snyder A, Makarov V, Merghoub T, Yuan J, Zaretsky JM, Desrichard A, et al. Genetic basis for clinical response to CTLA-4 blockade in melanoma. *N Engl J Med* 2014;371:2189–99.
- Hugo W, Zaretsky JM, Sun L, Song C, Moreno BH, Hu-Lieskovan S, et al. Genomic and transcriptomic features of response to anti-PD-1 therapy in metastatic melanoma. *Cell* 2016;165:35–44.

Authors' Contributions

J.T. Topham: Conceptualization, methodology, data curation, formal analysis, writing-original draft, writing-review and editing, project administration. **E. Titmuss:** Conceptualization, methodology, data curation, formal analysis, writing-original draft, writing-review and editing. **E.D. Pleasance:** Data curation, formal analysis. **L.M. Williamson:** Data curation. **J.M. Karasinska:** Data curation, writing-original draft, writing-review and editing. **L. Culibrk:** Methodology, writing-original draft, writing-review and editing. **M.K.C. Lee:** Data curation, formal analysis, writing-original draft, writing-review and editing. **S. Mendis:** Formal analysis, writing-original draft, writing-review and editing. **R.E. Denroche:** Formal analysis, project administration. **G.-H. Jang:** Data curation. **S.E. Kalloger:** Formal analysis, writing-original draft, writing-review and editing. **H.-L. Wong:** Data curation, writing-original draft, writing-review and editing. **R.A. Moore:** Data curation, project administration. **A.J. Mungall:** Data curation, writing-original draft, writing-review and editing. **G.M. O'Kane:** Data curation, writing-original draft, writing-review and editing. **J.J. Knox:** Data curation, writing-original draft, writing-review and editing, project administration. **S. Gallinger:** Data curation, writing-original draft, writing-review and editing, project administration. **J.M. Loree:** Formal analysis, writing-original draft, writing-review and editing. **D.L. Mager:** Formal analysis. **J. Laskin:** Supervision, data curation, writing-original draft, writing-review and editing, project administration. **M.A. Marra:** Conceptualization, data curation, formal analysis, writing-original draft, writing-review and editing. **S.J.M. Jones:** Data curation, formal analysis, writing-original draft, writing-review and editing. **D.F. Schaeffer:** Supervision. **D.J. Renouf:** Conceptualization, supervision, methodology, data curation, formal analysis, writing-original draft, writing-review and editing.

Acknowledgments

We gratefully acknowledge the participation of our patients, their families, the POG team, and the generous support of the Terry Fox Research Institute, BC Cancer Foundation, Pancreatic Cancer Canada. We thank the Library Construction, Biospecimen and Project Management teams at Canada's Michael Smith Genome Sciences Centre for expert technical assistance. We thank Artem Babaian for his expert advice on ERV databases and quantification, and interpretation of the results.

This research was supported through philanthropic donations received through the BC Cancer Foundation, as well as funding provided by the Terry Fox Research Institute, Ontario Institute for Cancer Research (PanCuRx Translational Research Initiative), Pancreatic Cancer Canada, and Genome British Columbia (project B20POG). We acknowledge contributions toward equipment and infrastructure from Genome Canada and Genome BC (projects 202SEQ, 212SEQ, 12002), Canada Foundation for Innovation (projects 20070, 30198, 30981, 33408), and the BC Knowledge Development Fund. M.A. Marra acknowledges infrastructure investments from the Canada Foundation for Innovation and the support of the Canada Research Chairs and CIHR Foundation (FDN-143288) programs. D.J. Renouf is a recipient of the MSFHR Health Professional-Investigator Award, and D.F. Schaeffer is a recipient of the VCHRI Investigator Award.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received February 6, 2020; revised April 28, 2020; accepted June 4, 2020; published first June 9, 2020.

12. van de Lagemat LN, Landry J-R, Mager DL, Medstrand P. Transposable elements in mammals promote regulatory variation and diversification of genes with specialized functions. *Trends Genet* 2003;19:530–6.
13. Gibb EA, Warren RL, Wilson GW, Brown SD, Robertson GA, Morin GB, et al. Activation of an endogenous retrovirus-associated long non-coding RNA in human adenocarcinoma. *Genome Med* 2015;7:22.
14. Strick R, Strissel PL, Baylin SB, Chiappinelli KB. Unraveling the molecular pathways of DNA-methylation inhibitors: human endogenous retroviruses induce the innate immune response in tumors. *Oncoimmunology* 2016;5:e1122160.
15. Solovyyov A, Vabret N, Arora KS, Snyder A, Funt SA, Bajorin DF, et al. Global cancer transcriptome quantifies repeat element polarization between immunotherapy responsive and T cell suppressive classes. *Cell Rep* 2018;23:512–21.
16. Panda A, de Cubas AA, Stein M, Riedlinger G, Kra J, Mayer T, et al. Endogenous retrovirus expression is associated with response to immune checkpoint blockade in clear cell renal cell carcinoma. *JCI Insight* 2018;3:e121522.
17. Laskin J, Jones S, Aparicio S, Chia S, Ch'ng C, Deyell R, et al. Lessons learned from the application of whole-genome analysis to the treatment of patients with advanced cancers. *Cold Spring Harb Mol Case Stud* 2015;1:a000570.
18. Liu M, Ohtani H, Zhou W, Ørskov AD, Charlet J, Zhang YW, et al. Vitamin C increases viral mimicry induced by 5-aza-2'-deoxycytidine. *Proc Natl Acad Sci U S A* 2016;113:10238–44.
19. Jones MR, Lim H, Shen Y, Pleasance E, Ch'ng C, Reisle C, et al. Successful targeting of the NRG1 pathway indicates novel treatment strategy for metastatic cancer. *Ann Oncol* 2017;28:3092–7.
20. Li H, Durbin R. Fast and accurate short read alignment with burrows-wheeler transform. *Bioinformatics* 2009;25:1754–60.
21. Butterfield YS, Kreitzman M, Thiessen N, Corbett RD, Li Y, Pang J, et al. JAGuaR: junction alignments to genome for RNA-seq reads. *PLoS One* 2014;9:e102398.
22. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* 2014;30:923–30.
23. Lazar C, Meganck S, Taminau J, Steenhoff D, Coletta A, Molter C, et al. Batch effect removal methods for microarray gene expression data integration: a survey. *Brief Bioinform* 2013;14:469–90.
24. Smit AFA. RepeatMasker Open-4.0 [Internet]; 2013. Available from: <http://www.repeatmasker.org>.
25. Derrien T, Johnson R, Bussotti G, Tanzer A, Djebali S, Tilgner H, et al. The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome Res* 2012;22:1775–89.
26. Volders PJ, Helsens K, Wang X, Menten B, Martens L, Gevaert K, et al. LNCipedia: a database for annotated human lncRNA transcript sequences and structures. *Nucleic Acids Res* 2013;41:D246–51.
27. Liberzon A, Subramanian A, Pinchback R, Thorvaldsdóttir H, Tamayo P, Mesirov JP. Molecular Signatures Database (MSigDB) 3.0. *Bioinformatics* 2011;27:1739–40.
28. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 2014;15:550.
29. Therneau TM. A Package for Survival Analysis in R [Internet]; 2015. Available from: <https://CRAN.R-project.org/package=survival>.
30. Kassambara A. Drawing survival curves using “ggplot2” [Internet]; 2018. Available from: <https://cran.r-project.org/web/packages/survminer/index.html>.
31. R Development Core Team. R: a language and environment for statistical computing. Version 2.0.1. *R Found Stat Comput* 2004; Vienna, Austria.
32. Babaian A, Mager DL. Endogenous retroviral promoter exaptation in human cancer. *Mob DNA* 2016;7:24.
33. Faulkner GJ, Kimura Y, Daub CO, Wani S, Plessy C, Irvine KM, et al. The regulated retrotransposon transcriptome of mammalian cells. *Nat Genet* 2009;41:563–71.
34. Lee YK, Chew A, Phan H, Greenhalgh DG, Cho K. Genome-wide expression profiles of endogenous retroviruses in lymphoid tissues and their biological properties. *Virology* 2008;373:263–73.
35. Chiappinelli KB, Strissel PL, Desrichard A, Li H, Henke C, Akman B, et al. Inhibiting DNA methylation causes an interferon response in cancer via dsRNA including endogenous retroviruses. *Cell* 2015;162:974–86.
36. Liu M, Thomas SL, DeWitt AK, Zhou W, Madaj ZB, Ohtani H, et al. Dual inhibition of DNA and histone methyltransferases increases viral mimicry in ovarian cancer cells. *Cancer Res* 2018;78:5754–66.
37. Attermann AS, Bjerregaard AM, Saini SK, Grønbaek K, Hadrup SR. Human endogenous retroviruses and their implication for immunotherapeutics of cancer. *Ann Oncol* 2018;29:2183–91.
38. Nirmal AJ, Regan T, Shih BB, Hume DA, Sims AH, Freeman TC. Immune cell gene signatures for profiling the microenvironment of solid tumors. *Cancer Immunol Res* 2018;6:1388–400.
39. Walsh D, Mathews MB, Mohr I. Tinkering with translation: protein synthesis in virus-infected cells. *Cold Spring Harb Perspect Biol* 2013;5:a012351.
40. Li MMH, MacDonald MR, Rice CM. To translate, or not to translate: viral and host mRNA regulation by interferon-stimulated genes. *Trends Cell Biol* 2015;25:320–9.
41. Rasmussen KD, Helin K. Role of TET enzymes in DNA methylation, development, and cancer. *Genes Dev* 2016;30:733–50.
42. Weber F, Wagner V, Rasmussen SB, Hartmann R, Paludan SR. Double-stranded RNA is produced by positive-strand RNA viruses and DNA viruses but not in detectable amounts by negative-strand RNA viruses. *J Virol* 2006;80:5059–64.
43. Polymenis M, Aramayo R. Translate to divide: control of the cell cycle by protein synthesis. *Microb Cell* 2015;2:94–104.
44. Lau CPY, Ng PKS, Li MS, Tsui SKW, Huang L, Kumta SM. p63 regulates cell proliferation and cell cycle progression-associated genes in stromal cells of giant cell tumor of the bone. *Int J Oncol* 2013;42:437–43.
45. Krönung SK, Beyer U, Chiamonte ML, Dolfini D, Mantovani R, Dobbelstein M. LTR12 promoter activation in a broad range of human tumor cells by HDAC inhibition. *Oncotarget* 2016;7:33484–97.
46. Melamed P, Yosefzon Y, David C, Tsukerman A, Pnueli L. Tet enzymes, variants, and differential effects on function. *Front Cell Dev Biol* 2018;6:22.
47. Wang-Johanning F, Li M, Esteva FJ, Hess KR, Yin B, Rycaj K, et al. Human endogenous retrovirus type K antibodies and mRNA as serum biomarkers of early-stage breast cancer. *Int J Cancer* 2014;134:587–95.
48. Rooney MS, Shukla SA, Wu CJ, Getz G, Hacohen N. Molecular and genetic properties of tumors associated with local immune cytolytic activity. *Cell* 2015;160:48–61.
49. Tokuyama M, Kong Y, Song E, Jayewickreme T, Kang I, Iwasaki A. ERVmap analysis reveals genome-wide transcription of human endogenous retroviruses. *Proc Natl Acad Sci U S A* 2018;115:12565–72.