

# It Takes Two (Genomes) to Cancer: Paired Viral and Host Transcriptome Analysis Provides New Insights about EBV Carcinogenicity

Sam M. Mbulaiteye<sup>1</sup> and Ludmila Prokunina-Olsson<sup>2</sup>



The discovery of Epstein-Barr virus (EBV) in 1964 gave birth to the field of viral oncology. Despite significant scientific and clinical developments in research on several other viruses discovered and linked to cancer risk much later, our understanding of EBV as a carcinogen and a possible target for therapeutic interventions remains limited. In this issue of *Cancer Research*, Chakravorty and col-

leagues present results of massive reanalysis of public RNA-sequencing data for 291 control and 1,051 tumor samples representing 15 cancer types. Their paired analysis of the viral and host transcriptome sheds light on mechanisms of EBV carcinogenicity and provides new leads for translational applications.

See related article by Chakravorty et al., p. 6010

The seminal discovery of Epstein-Barr virus (EBV) in a Burkitt lymphoma tumor in 1964 ignited global interest in virus-associated cancers. Over the past five decades, seven viruses have been linked to 17 cancer types (1), accounting for 1,385,000 cancer cases worldwide (about 15% of the global cancer burden) in 2012 (2). This knowledge paved the way for the development of vaccines and treatments that now promise to prevent 1.2 million cancers caused by viruses, hepatitis B (HBV), human papillomavirus (HPV) types 16 and 18, and hepatitis C (HCV), which were all discovered and linked to cancer risk after EBV discovery—in 1967, 1983, and 1989, respectively.

Several criteria have been developed to evaluate potential cancer-causing viruses (3). First, viral genomes should be detectable in tumors. Although EBV was detected in tumors, it was in an episomal, not integrated into the host DNA form. Second, viral genomes should encode proteins with transforming activity. EBV met this criterion by encoding the transcription factor Zta (BZLF1) and latent membrane protein 1 (LMP1; ref. 1), although these genes were expressed at low levels or not at all in most EBV-associated tumors. Third, viral cell tropism should match tumor tissue/cell of origin. EBV primarily infects B cells, but is detected in various tumor types, including epithelial (nasopharyngeal carcinoma and gastric cancer), B cells (Burkitt lymphoma, Hodgkin lymphoma, and diffuse large B-cell lymphoma), natural killer T cells (T/NK-cell lymphoma), and smooth muscle (1). Fourth, the epidemiology of the cancer-causing virus and corresponding cancer should match. Up to 90% of the world's population is

infected with EBV, yet the classical EBV-associated malignancies showed distinct and nonoverlapping distribution patterns. Specifically, endemic Burkitt lymphoma occurs in Africa and affects children, while nasopharyngeal carcinoma is most common in East Asia and affects adults. These molecular and epidemiologic contradictions cast a shadow on EBV's role as a cancer-causing virus and slowed the progress in understanding molecular mechanisms of EBV carcinogenicity and development of clinical interventions (4). Today, EBV-associated cancers still constitute a major unaddressed public health problem, estimated at 224,000 individual cancers or 15% of all infection-associated cancer burden (1), urging consolidated efforts to understand and control this infection and prevent and treat the associated cancers (4).

Rapid advances in next-generation sequencing (NGS) tools have accelerated many research areas, including cancers and host-pathogen interactions. Perhaps, the most exciting findings are emerging at the intersection of these fields, providing new clues on genomics and biology of cancers caused by viruses. Mining RNA-sequencing data generated by The Cancer Genome Atlas for traces of RNA viruses and transcribed DNA viruses (such as EBV) has already generated many exciting findings. For example, genomic analyses of EBV-associated tumors identified an EBV molecular signature that provided stratification of gastric adenocarcinomas (5) and pediatric Burkitt lymphoma (6), and analyses in Burkitt lymphoma (7) and NPC (8) suggested the existence of high-risk, tumor-specific EBV variants.

In this issue of *Cancer Research*, Chakravorty and colleagues (9) capitalized on multiple studies that generated public RNA-sequencing data for 291 control and 1,051 tumor samples representing 15 cancer types associated with EBV infection. The authors performed massive reanalysis of the paired virus and host transcriptome with the goal to define a dynamic interaction between the two genomes (viral and host), identify molecular markers of this interaction, characterize corresponding biological mechanisms, and nominate viral and host biomarkers that might be useful for diagnosis, prognosis, or treatment of EBV-associated cancers.

First, the authors defined the EBV reactivation signature as a load of EBV RNA detectable in the total tumor transcriptome generated by RNA sequencing. This signature correctly called all EBV-positive

<sup>1</sup>Infections and Immunoepidemiology Branch, Division of Cancer Epidemiology and Genetics, NCI, NIH, Bethesda, Maryland. <sup>2</sup>Laboratory of Translational Genomics, Division of Cancer Epidemiology and Genetics, NCI, NIH, Bethesda, Maryland.

**Corresponding Author:** Ludmila Prokunina-Olsson, Division of Cancer Epidemiology and Genetics, NCI, NIH, Bethesda, MD 20892-4605. Phone: 240-760-6531; Fax: 240-541-4442; E-mail: prokuninal@mail.nih.gov

Cancer Res 2019;79:5917-9

doi: 10.1158/0008-5472.CAN-19-2996

©2019 American Association for Cancer Research.

and EBV-negative lymphoblastoid cell lines (LCL) and detected EBV infection in 100% of nasopharyngeal carcinoma, 80% of endemic Burkitt lymphoma, and 71% of T/NK-cell lymphoma, while detecting 0% of tumors of eight other types where the infection was not expected. The authors then compared viral RNA load in EBV-positive tumors with EBV-transformed LCLs derived from healthy donors and, unexpectedly, found that most EBV genes expressed in LCLs were repressed in EBV-positive tumors. Because latent and lytic phases of EBV life cycle are controlled by genes that prevent or induce EBV reactivation, the authors combed through LCL EBV expression data in search for genes negatively correlated with viral RNA load, that is, virostatic genes (inhibitors of viral lytic replication). The search identified *EBNA1* as a gene with the highest negative correlation value, as well as *EBNA3C*, *EBNA2*, *EBNA-LP*, and genes for most latent membrane proteins. Consistently, virostatic genes had 4-fold more mutations than other EBV genes. *In vitro* studies confirmed that mutations in representative virostatic genes (such as Q322X and G342X in LMP1) were associated with loss of viral repression and high viral RNA load. Taken together, these results suggest that EBV carcinogenesis might be modulated by genetic variation in virostatic genes that control EBV functional properties and viral load.

Because EBV can be considered a mutagen, the authors tried but failed to identify a mutational signature in the host genome that would distinguish EBV-positive versus EBV-negative tumors. The focus on mutations in specific cancer driver genes was more successful and showed that high viral RNA load was associated with more somatic mutations in *DDX3X* or *MYC* genes, although this association was significant only in endemic Burkitt lymphoma tumors. These results are descriptive and need experimental validation, but they nonetheless provide a strong suggestion that high viral RNA load may be a marker for mutations in some host cancer driver genes. Analysis of host transcriptome dichotomized all EBV-positive tumors, regardless of their tissue of origin, into two groups based on activated or inhibited IFN signature. The biology of this unexpected finding should be further investigated, especially because the activated IFN signature also included known negative regulators of the immune response, *PD-L1* and *IDO1*. These findings suggest that EBV-associated cancers might be treated with immune checkpoint inhibitors but with differential clinical responses that could be predicted on the basis of IFN signatures of the tumors.

This work also breaks new ground on the role of EBV integration and destabilization of the host genome in tumor development. The hit-and-run scenario has been proposed as a mechanism relevant to EBV-associated cancers with undetectable EBV. The authors reported viral integration events in 60 of 112 EBV<sup>+</sup> cancers (with a median of three integration events per sample). Although this is significantly higher than a previously reported rate, it may still be an underestimate because of technical limitations of analysis based on RNA-sequencing data that can detect events only in genes expressed at a reasonably high level. Viral integration was specifically enriched within regions of open

chromatin characterized as super-enhancers, possibly regulating highly expressed viral response genes and genes inducing apoptosis in response to DNA damage. However, the RNA-sequencing approach may be sensitive to reverse causality, which makes it difficult to separate the cause and effect, such as high expression caused by viral integration versus more likely detection of integration within genes that are highly expressed. The authors explored integration events based on whole-genome DNA sequencing in one of the samples, but more extensive analysis of EBV-associated tumors by whole-genome DNA sequencing is necessary for a comprehensive evaluation of both transcribed and nontranscribed viral integration events.

The study provided many answers but also left a number of questions to be explored by follow-up studies. For example, whether high-risk EBV variants are shared or specific to certain EBV-associated cancers? How EBV, a master epigenetic manipulator, modulates methylation of the host DNA and affects the host transcriptome? What is the relative contribution of the virus in episomal versus host DNA-integrated form in tumor initiation and progression, in line with a hit-and-run hypothesis?

We believe this study will have a strong positive impact on the multidisciplinary field of viral oncology. First, researchers who are interested in viral biomarkers and mechanism discovery are likely to find the methodical approach presented here a useful template to analyze multiple public high-dimensional datasets, which are increasing in number, depth, and complexity, to explore viral-host genome interactions across multiple cancers and viruses. The article sets valuable precedence of providing extensive supplementary datasets that can be mined for further discoveries. In addition, this article should invite complementary analyses of DNA sequencing and methylome datasets together with the RNA-sequencing data to provide a more comprehensive picture of EBV variation, RNA expression, protein function, and their effect on the host. These efforts will be boosted by the ongoing efforts of the International Cancer Genome Consortium/The Cancer Genome Atlas that are generating whole-genome sequencing data for many primary tumors (10). These multidimensional datasets will provide abundant opportunities for mining for traces of EBV and other known or unknown pathogens and explore their relationship with the host, resulting in cancer. Access to these datasets and their proper analysis will help to generate biologically plausible models that can be studied by other genetic, genomic, and epidemiologic approaches, and, eventually, improve public health.

#### Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

#### Acknowledgments

This work was supported by the Intramural Research Program of the Division of Cancer Epidemiology and Genetics, NCI, NIH.

Received September 24, 2019; accepted September 26, 2019; published first December 2, 2019.

#### References

- Shannon-Lowe C, Rickinson A. The global landscape of EBV-associated tumors. *Front Oncol* 2019;9:713.
- Plummer M, de Martel C, Vignat J, Ferlay J, Bray F, Franceschi S. Global burden of cancers attributable to infections in 2012: a synthetic analysis. *Lancet Glob Health* 2016;4:e609–16.
- Evans AS, Mueller NE. Viruses and cancer. *Causal associations. Ann Epidemiol* 1990;1:71–92.
- Cohen JL, Fauci AS, Varmus H, Nabel GJ. Epstein-Barr virus: an important vaccine target for cancer prevention. *Sci Translat Med* 2011;3:107fs7.

5. The Cancer Genome Atlas Research Network. Comprehensive molecular characterization of gastric adenocarcinoma. *Nature* 2014;513:202–9.
6. Grande BM, Gerhard DS, Jiang A, Griner NB, Abramson JS, Alexander TB, et al. Genome-wide discovery of somatic coding and noncoding mutations in pediatric endemic and sporadic Burkitt lymphoma. *Blood* 2019;133:1313–24.
7. Liao HM, Liu H, Lei H, Li B, Chin PJ, Tsai S, et al. Frequency of EBV LMP-1 promoter and coding variations in Burkitt lymphoma samples in Africa and South America and peripheral blood in Uganda. *Cancers* 2018;10. doi: 10.3390/cancers10060177.
8. Xu M, Yao Y, Chen H, Zhang S, Cao SM, Zhang Z, et al. Genome sequencing analysis identifies Epstein-Barr virus subtypes associated with high risk of nasopharyngeal carcinoma. *Nat Genet* 2019;51:1131–6.
9. Chakravorty S, Yan B, Wang C, Wang L, Quaid JT, Lin CF, et al. Integrated pan-cancer map of EBV-associated neoplasms reveals functional host-virus interactions. *Cancer Res* 2019;79:6010–23.
10. Campbell PGG, Stuart J, Korbel J, Stein LD. Pan-cancer analysis of whole genomes. *BioRxiv* 2017. doi: <https://doi.org/10.1101/162784>.

# Cancer Research

The Journal of Cancer Research (1916–1930) | The American Journal of Cancer (1931–1940)

## It Takes Two (Genomes) to Cancer: Paired Viral and Host Transcriptome Analysis Provides New Insights about EBV Carcinogenicity

Sam M. Mbulaiteye and Ludmila Prokunina-Olsson

*Cancer Res* 2019;79:5917-5919.

**Updated version** Access the most recent version of this article at:  
<http://cancerres.aacrjournals.org/content/79/23/5917>

**Cited articles** This article cites 8 articles, 2 of which you can access for free at:  
<http://cancerres.aacrjournals.org/content/79/23/5917.full#ref-list-1>

**E-mail alerts** [Sign up to receive free email-alerts](#) related to this article or journal.

**Reprints and Subscriptions** To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at [pubs@aacr.org](mailto:pubs@aacr.org).

**Permissions** To request permission to re-use all or part of this article, use this link <http://cancerres.aacrjournals.org/content/79/23/5917>. Click on "Request Permissions" which will take you to the Copyright Clearance Center's (CCC) Rightslink site.