

# Highly Sensitive Virome Capture Sequencing Technique VirCapSeq-VERT Identifies Partial Noncoding Sequences but no Active Viral Infection in Cutaneous T-Cell Lymphoma

*Journal of Investigative Dermatology* (2018) ■, ■-■; doi:10.1016/j.jid.2018.01.024

## TO THE EDITOR

The potential role of viruses as oncogenic triggers in cutaneous T-cell lymphoma (CTCL) pathogenesis is a subject of ongoing investigation. CTCL occurs with an increased incidence in immunosuppressed patients (Nikolaou et al., 2015; Pomerantz et al., 2010; Wilkins et al., 2006). Spectratyping studies have shown depleted T-cell receptor diversity in CTCL patients similar to that seen in patients with advanced HIV, and Ingenuity Pathway Analysis (Qiagen, Hilden, Germany) shows increased expression of genes critical to host viral response (Yawalkar et al., 2003). However, no consistent association between a viral pathogen and CTCL has been established (Mirvish et al., 2013).

The development of high-throughput sequencing (HTS) has provided powerful new tools for pathogen discovery, yet HTS studies have not identified viral sequences in CTCL (Dereure et al., 2013; Dulmage et al., 2015; Lee et al., 2012). One explanation for failure is that viral burden in samples may be below the threshold for detection in complex backgrounds of host nucleic acid. To address this possibility, we used a positive selection method for HTS with enhanced sensitivity (Briese et al., 2015).

The Virome Capture Sequencing Platform for Vertebrate Viruses (VirCapSeq-VERT; Roche, Pleasanton, CA) is a positive-selection probe-based method that targets all 207 known vertebrate viruses, allowing for a 100- to 10,000-

fold increase in the number of viral reads over traditional HTS, a degree of sensitivity comparable to targeted real-time PCR. Over 90% of genome recovery can be achieved with only 100 viral copies in 50 ng of whole-blood nucleic acid. Superior to real-time PCR, however, VirCapSeq-VERT can detect viruses that differ from known sequences by 40% (Briese et al., 2015).

We used VirCapSeq-VERT to search for viral sequences in mononuclear cells derived from the peripheral blood of leukemic CTCL patients with Sézary syndrome. In total, 27 million paired-end reads were generated using Illumina MiSeq, with an average of 2.25 million paired-end reads per sample. An average of 11% of the reads were removed after filtration; 74% of the reads were removed after host subtraction. Overall, 49,349 sequences showed partial homology to 23 unique viral species including human endogenous retroviruses (HERVs) H and K, human herpesvirus (HHV) 4 and 5, HIV, human T-lymphotropic virus-1 (HTLV-1), Lassa virus, Luna virus, bovine viral diarrhea virus, and Ngari virus (Table 1).

Bovine viral diarrhea virus sequences are common in bovine serum products used in sample collection and thus were excluded from further analysis. HERV-H/env62 and HERV-H/env59 sequences were present in 2 of 6 and 3 of 6 samples, respectively, matching to a proviral copy. The HHV-4 and HHV-5 sequences were present in 1 of 6 samples tested; 36 reads matched to the partial HHV genome.

HIV-1 sequences were present in 1 of 6 samples, partially matching to a nonfunctional *pol* protein gene. Short sequences with partial homology to Lassa, Ngari, and Luna viruses were present in 4 of 6, 5 of 6, and 6 of 6 samples, respectively. Detailed analysis indicated that these sequences were 100% homologous to human genomic sequences (accession number KF478765.1 for Lassa, KJ716849.1 for Ngari, and AB972431.1 and KX121618.1 for Luna). Thus, the reference sequences likely represent misannotations. HERV-K and HTLV-1 sequences were present in 6 of 6 samples. The HERV-K sequences comprised partially coding proviral sequences, and the HTLV-1 sequences matched to defective proviral sequences, both showing high similarity to the host genome.

We conducted this study using VirCapSeq-VERT to search for viral sequences in patients with CTCL. VirCapSeq-VERT's ability to detect all known vertebrate viruses offers the most sensitive detection methods compared with conventional HTS and targeted reverse transcription PCR. No substantive coding sequences for viral pathogens or unknown viruses, or evidence for active infection, were obtained. All six Sézary syndrome samples expressed partial, noncoding sequences for HERV-K and HTLV-1, but these sequences had low read counts and contiguous sequence (contig) depths, and were thus insufficient to be considered as positive results. Nevertheless, their presence is of interest, because both viruses have been previously implicated in CTCL. HTLV-1's role in CTCL has been debated at length (Mirvish et al., 2013). HERVs are expressed in all human tissues but have been implicated in cancer pathogenesis, and increased expression

*Abbreviations:* contig, contiguous sequence; CTCL, cutaneous T-cell lymphoma; HERV, human endogenous retrovirus; HHV, human herpesvirus; HTLV-1, human T-lymphotropic virus-1; HTS, high-throughput sequencing; NCBI, National Center for Biotechnology Information; TNA, total nucleic acid; VirCapSeq-VERT, virome capture sequencing of vertebrate viruses

Accepted manuscript published online 7 February 2018; corrected proof published online XXX

© 2018 The Authors. Published by Elsevier, Inc. on behalf of the Society for Investigative Dermatology.

**Table 1. Potential viral species counts from MegaBLAST**

Viral Species	Sample ID						Maximum Sequence Length (nt) <sup>1</sup>	Maximum Contig Depth (Number of Reads) <sup>1</sup>	Significance
	1	2	3	4	5	6			
BVDV	8,264	10	8,844	—	14	—	1,837	108	Library preparation artifact
HERV-H/env59	8	—	—	—	9	10	145	2	Proviral sequences
HERV-H/env62	15	8	—	—	—	—	—	—	—
HERV-K	60	19	96	41	45	52	875	14	Proviral sequences
HHV-4	24	—	—	—	—	—	145	3	Low abundance/prevalence
HHV-5	12	—	—	—	—	—	—	—	—
HIV-1	10	—	—	—	—	—	84	4	Nonfunctional <i>pol</i> polymerase protein gene
HTLV-1	15	32	8	28	69	22	145	2	Proviral sequences
Lassa virus	11	8	—	—	26	20	145	1	Likely misannotation
Luna virus	3,912	3,804	3,673	111	372	3,533	432	26	Likely misannotation
Ngari virus	33	20	26	—	18	16	145	2	Likely misannotation

Abbreviations: BVDV, bovine viral diarrhea virus; contig, contiguous sequence; HERV, human endogenous retrovirus; HHV, human herpes virus; HTLV-1, human lymphotropic virus 1; ID, identification; nt, nucleotide.

<sup>1</sup>Maximum sequence length and contig depth are in sample 1.

of HERVs has been seen in CTCL patients (Fava et al., 2016; Maliniemi et al., 2013). Sequencing of healthy control patient samples with matched demographics was not available for this pilot study for direct comparison, but further investigation may elucidate the significance of these partial noncoding sequences and whether these viruses could play a genuine role in CTCL pathogenesis.

Six Sezary syndrome patients diagnosed according to the World Health Organization-European Organization for Research and Treatment of Cancer criteria were enrolled after receiving Columbia University Medical Center Institutional Review Board approval. Written informed patient consent was obtained from all patients (5 women, 1 man; age range = 58–82 years; 3 Caucasian, 2 Hispanic, and 1 African American).

Whole blood from each patient was collected. Peripheral blood mononuclear cells were isolated via density gradient centrifugation (Ficoll-Plaque; Millipore Sigma, St. Louis, MO), and plasma was isolated via centrifugation.

Total nucleic acid (TNA) was extracted from the plasma and peripheral blood mononuclear cells using NucliSENS easyMAG (BioMérieux, Marcy l'Etoile, France). TNAs were processed according to Center for Infection and Immunity standard operating procedures for VirCapSeq-VERT. Briefly, the VirCapSeq-VERT probe library was added to the TNAs. Positively selected and enriched sequences

underwent PCR amplification and sequencing using the Illumina (San Diego, CA) MiSeq platform. The 150-nucleotide-long paired end reads were generated (an average of more than 2 million reads per sample). HTS data were analyzed using the Center for Infection and Immunity bioinformatics viral discovery pipeline: the de-multiplexed fastq files were filtered for low-quality and low-complexity reads, adaptors were removed, reads were trimmed, and short reads were discarded. These pre-processed reads were subjected to computational subtraction against human reference databases from the National Center for Biotechnology Information (NCBI) to remove host background sequences. The databases used for host subtraction include human genomic, ribosomal, and mitochondrial sequences. The host-subtracted reads were then assembled de novo using MIRA (Bastien Chevreux, Lexington, MA) assembler, version 4.0. Contigs and singletons were annotated using homology search program BLAST+ from NCBI (Silver Spring, MD). The sequences were screened for highly similar sequences using megaBLAST against the GenBank nucleotide database. To ensure that mutant viral sequences were not discarded, sequences with low/no homology to the GenBank viral nucleotide database were reblasted using Blastx against the GenBank database. Contigs and reads were then mapped to reference genomes

identified from BLAST using Geneious, version 6.0.6 (Biomatters INC, Newark, NJ). The sequence data supporting the results of this article are available in NCBI under accession number PRJNA415045 with biosample accession numbers SAMN07812478–83.

#### ORCIDiS

Mary Elizabeth Anderson: <http://orcid.org/0000-0003-1364-1248>

Christina Chung Patrone: <http://orcid.org/0000-0001-8906-7460>

#### CONFLICT OF INTEREST

The authors state no conflict of interest.

#### ACKNOWLEDGMENTS

We thank Lisa Keller and Saky Yakas for their philanthropic support of cutaneous lymphoma research at Columbia University.

**Mary E. Anderson<sup>1,4</sup>, Dorottya Nagy-Szaka<sup>2,4</sup>, Komal Jain<sup>2</sup>, Christina C. Patrone<sup>1</sup>, Mark G. Frattini<sup>3</sup>, W. Ian Lipkin<sup>2</sup> and Larisa J. Geskin<sup>1,\*</sup>**

<sup>1</sup>Department of Dermatology, Columbia University Medical Center, New York, New York, USA; <sup>2</sup>Center for Infection and Immunity, Mailman School of Public Health, Columbia University Medical Center, New York, New York, USA; and <sup>3</sup>Department of Medicine, Columbia University Medical Center, New York, New York, USA

<sup>4</sup>These authors contributed equally to this work.

\*Corresponding author e-mail: [ljpg2145@cumc.columbia.edu](mailto:ljpg2145@cumc.columbia.edu)

#### REFERENCES

Briese T, Kapoor A, Mishra N, Jain K, Kumar A, Jabado OJ, et al. Virome capture sequencing enables sensitive viral diagnosis and comprehensive virome analysis. *MBio* 2015;6:e01491–15.

- Dereure O, Cheval J, Du Thanh A, Pariente K, Sauvage V, Claude Manuguerra J, et al. No evidence for viral sequences in mycosis fungoides and Sezary syndrome skin lesions: a high throughput sequencing approach. *J Invest Dermatol* 2013;133:853–5.
- Dulmage BO, Feng H, Mirvish E, Geskin L. Black cat in a dark room: the absence of a directly oncogenic virus does not eliminate the role of an infectious agent in cutaneous T-cell lymphoma pathogenesis. *Br J Dermatol* 2015;172:1449–51.
- Fava P, Bergallo M, Atrua C, Brizio M, Galliano I, Montanari P, et al. Human endogenous retrovirus expression in primary cutaneous T-cell lymphomas. *Dermatology* 2016;232:38–43.
- Lee CS, Ungewickell A, Bhaduri A, Qu K, Webster DE, Armstrong R, et al. Transcriptome sequencing in Sezary syndrome identifies Sezary cell and mycosis fungoides-associated lncRNAs and novel transcripts. *Blood* 2012;120:3288–97.
- Maliniemi P, Vincendeau M, Mayer J, Frank O, Hahtola S, Karenko L, et al. Expression of human endogenous retrovirus-w including syncytin-1 in cutaneous T-cell lymphoma. *PLoS One* 2013;8(10):e76281.
- Mirvish JJ, Pomerantz RG, Falo LD Jr, Geskin LJ. Role of infectious agents in cutaneous T-cell lymphoma: facts and controversies. *Clin Dermatol* 2013;31:423–31.
- Nikolaou V, Papadavid E, Economidi A, Marinos L, Moustou E, Karampidou K, et al. Mycosis fungoides in the era of antitumour necrosis factor-alpha treatments. *Br J Dermatol* 2015;173:590–3.
- Pomerantz RG, Campbell LS, Jukic DM, Geskin LJ. Posttransplant cutaneous T-cell lymphoma: case reports and review of the association of calcineurin inhibitor use with posttransplant lymphoproliferative disease risk. *Arch Dermatol* 2010;146:513–6.
- Wilkins K, Turner R, Dolev JC, LeBoit PE, Berger TG, Maurer TA. Cutaneous malignancy and human immunodeficiency virus disease. *J Am Acad Dermatol* 2006;54:189–206.
- Yawalkar N, Ferenczi K, Jones DA, Yamanaka K, Suh KY, Sadat S, et al. Profound loss of T cell receptor repertoire complexity in cutaneous T-cell lymphoma. *Blood* 2003;102:4059–66.