Science Advances
AAAS

Supplementary Materials for

Insertional activation of *STAT3* and *LCK* by HIV-1 proviruses in T cell lymphomas

John W. Mellors, Shuang Guo, Asma Naqvi, Leah D. Brandt, Ling Su, Zhonghe Sun, Kevin W. Joseph, Dimiter Demirov, Elias K. Halvas, Donna Butcher, Beth Scott, Aaron Hamilton, Marintha Heil, Baktiar Karim, Xiaolin Wu*, Stephen H. Hughes*

*Corresponding author. Email: forestwu@mail.nih.gov (X.W.); hughesst@mail.nih.gov (S.H.H.)

Published 13 October 2021, *Sci. Adv.* **7**, eabi8795 (2021)
DOI: 10.1126/sciadv.abi8795

The PDF file includes:

Other Supplementary Material for this manuscript includes the following:

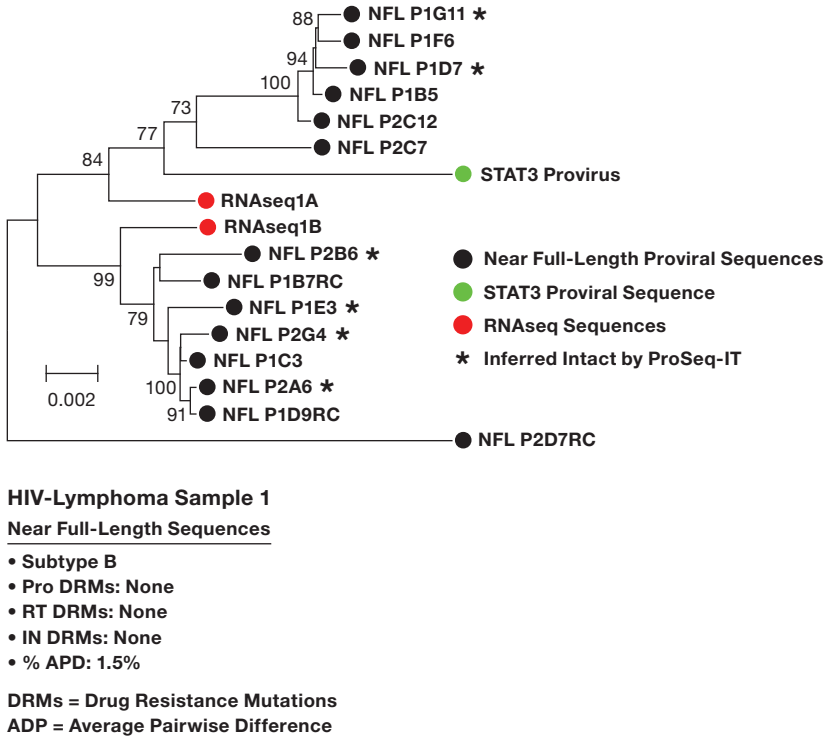

Figure S1. The proviruses in the T cell lymphoma sample 1A are diverse. Proviruses were amplified and sequenced as described in Materials and Methods. NFL and RNAseq tumor 1A and tumor 1B sequences were aligned and trimmed to the length of the STAT3 proviral sequence. A phylogenetic tree shows that the proviruses arising from superinfection of the lymphoma were diverse with an average pairwise difference of 1.5%. Of the 14 proviruses that were analyzed, 6 were intact (marked on the figure by black asterisks), and 8 were defective due to hypermutations, deletions, or other mutations as determined by the ProSeq-IT online software from the National Cancer Institute Proviral Sequence Database.

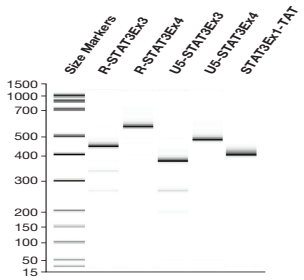

Figure S2. cDNAs that were derived from the LTR-STAT3 and STAT3-Tat fusion RNAs in lymphoma 1A were amplified using primers that were designed to selectively amplify only cDNA products and not genomic DNA (see text and Materials and Methods). The primers used in the amplifications are indicated above the lanes in the figure. The resulting PCR products were fractionated using an Agilent TapeStation, and the sizes of the PCR products were estimated using markers of defined sizes (lane 1). The sizes of the PCR products matched the sizes that were predicted based on the structure of the provirus that was integrated in the STAT3 in sample 1A.

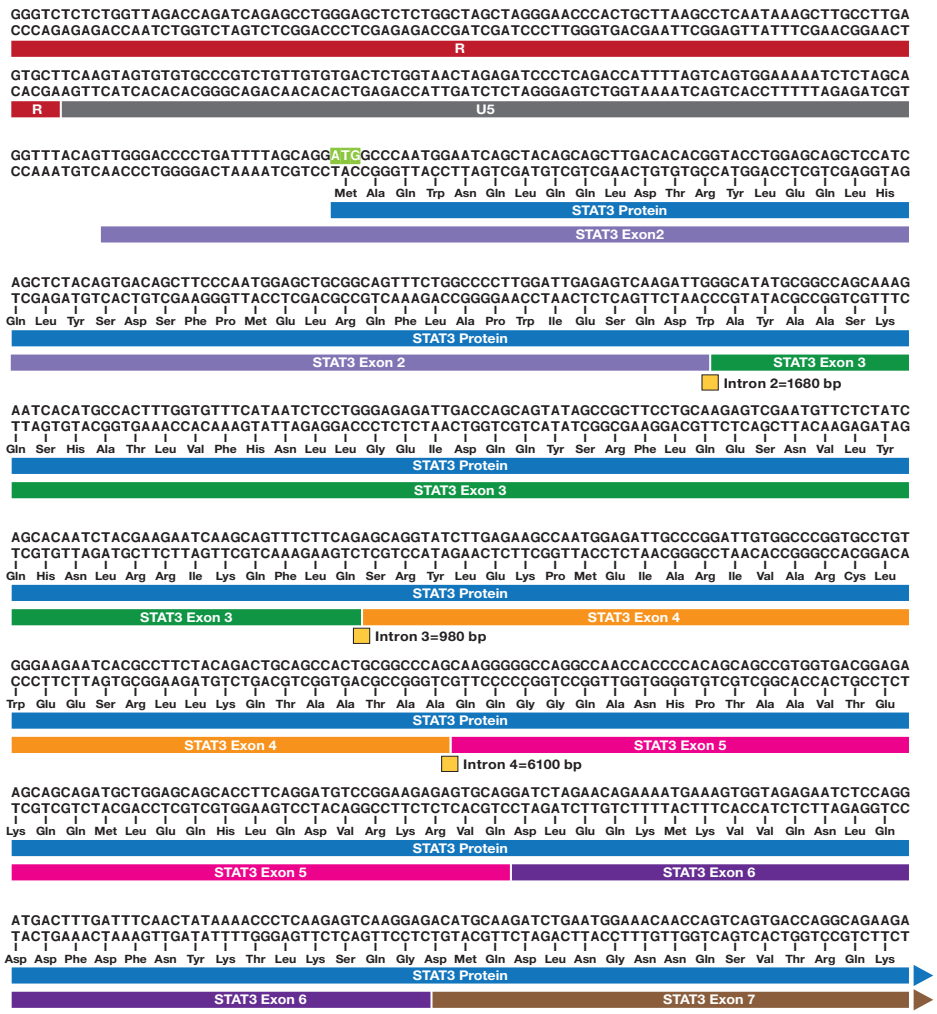

Figure S3. Sequence of the LTR-STAT3 fusion mRNA. In this diagram, the HIV portion of the mRNA is shown in red, and STAT3 is in several colors, to distinguish the different exons. The sequence of STAT3 protein is also shown.

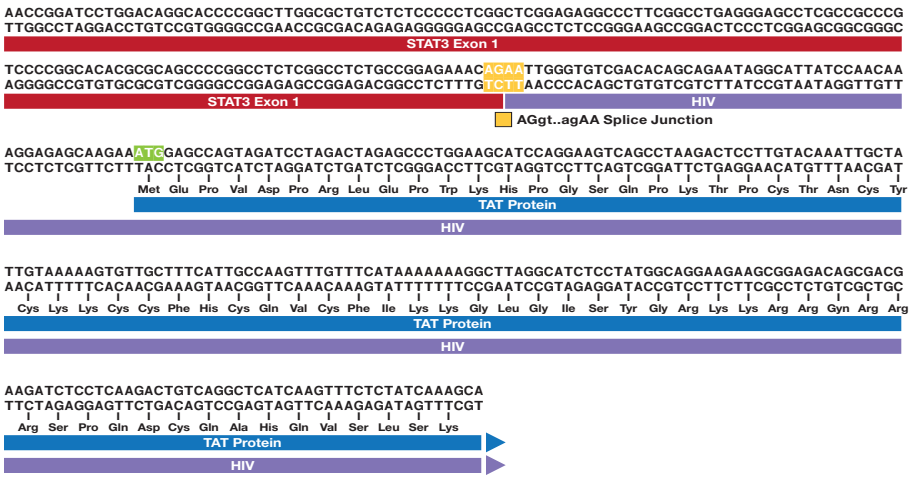

Figure S4. Sequence of the STAT3-Tat fusion mRNA. The STAT3 portion of the mRNA is shown in red, the provirus is in blue. The sequence of the splice junction is indicated. The sequence of the Tat protein is also shown.

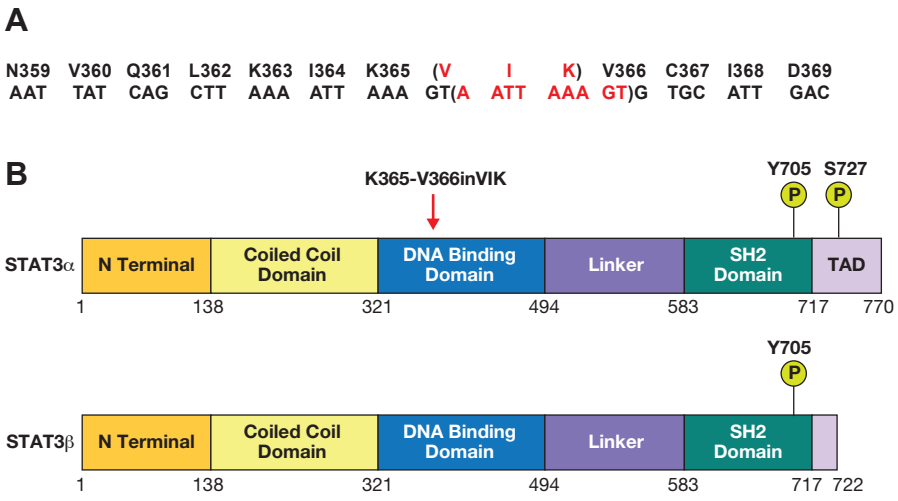

Figure S5. In lymphoma sample 1A, there is a 9 nucleotide insertion in the STAT3 gene in the allele that is activated by an HIV provirus. Panel A. The 9 nucleotide insertion leads to a three amino acid insertion in the portion of STAT3 that encodes the DNA binding domain. Panel B. The diagram shows the two forms of the STAT3 protein, and their domains. The position of the VIK insertion is marked with an arrow, as are the sites where the proteins are phosphorylated.

Table S1 - PLEASE SEE SEPARATE FILE ATTACHED

Table S2

Sample	Percentage of the total 5'LTR reads
Sample 1A 2-LTR junctions	0.10
Sample 1A linear ends	16.12
Sample 1B 2-LTR junctions	0.11
Sample 1B linear ends	2.67

Table S3 - PLEASE SEE SEPARATE FILE ATTACHED

Table S4 - PLEASE SEE SEPARATE FILE ATTACHED

Table S5

ddPCR Primer IDs	ddPCR Primer Sequences
STAT3_40500566_F	CCATCACCTGTACCCATACATT
STAT3_40500566_R	CCAGAGCATCTTTATCCCTAGTC
HIV_U5-PBS-R_lyph	TCGCTTTCGCGTCCCTGTTC
HIV-env PRB Set 1	/56-FAM/CTTGGGAGC/ZEN/AGCAGGAAGCACTAT/3IABkFQ/
STAT3-5LTRjunction REV	TCTCTGCTGTCCCTGTAATAAAC
STAT3-5LTRjunction PRB	/56-FAM/TCCCTGATT/ZEN/GGCAGAATTACACACCA/3IABkFQ/
STAT3-5LTRjunction FWD	CTGGGACTTGTGGTGAACAT
HIVpsi FWD Set 2	GGACTCGGCTTGCTGAAG
HIVpsi REV Set 2	GCACCCATCTCTCTCCTTCTA
HIVpsi PRB Set 2	/5HEX/TTTGGCGTA/ZEN/CTCACCAGTCGCC/3IABkFQ/
HIV-env FWD Set 1	CTAGGAGCTTTGTTCCTTGGG
HIV-env REV Set 1	TGTACCGTCAGCGTTATTG
T1-STAT3-TCR FWD Set 1	TGCCAGGCCCTCACATA
T1-STAT3-TCR REV Set 1	GAACCGAAGGTGTAGCCATT
T1-STAT3-TCR PRB Set 1	/56-FAM/CAGTACCTC/ZEN/TGTGCCAGCAGTGAC/3IABkFQ/
T12-STAT3 FWD Set 3	CTATCTGTCTGGTAACTAGAGATCC
T12-STAT3 REV Set 3	CACACGCAAACATACTAGAGTTC
T12-STAT3 PRB Set 3	/56-FAM/AGGTGCTTC/ZEN/TTCGTGCCTTCTCTG/3IABkFQ/
T12-LCK FWD Set 4	CTATCTGTCTGGTAACTAGAGATCC
T12-LCK REV Set 4	GACACAATCATGGCTCACTG
T12-LCK PRB Set 4	/56-FAM/AGGATCACC/ZEN/TGAGCTATGGGAGGT/3IABkFQ/

Supplementary Data File S1. Sequence of the provirus integrated in STAT3 in tumors 1A and 1B. **PLEASE SEE SEPARATE FILE ATTACHED**