

RESEARCH ARTICLE

Open Access



# Structural variation of centromeric endogenous retroviruses in human populations and their impact on cutaneous T-cell lymphoma, Sézary syndrome, and HIV infection

Mark H. Kaplan<sup>1</sup>, Mark Kaminski<sup>2</sup>, Judith M. Estes<sup>2</sup>, Scott D. Gitlin<sup>2,12</sup>, Joseph Zahn<sup>3</sup>, James T. Elder<sup>3,12</sup>, Trilokraj Tejasvi<sup>3,12</sup>, Elizabeth Gensterblum<sup>4</sup>, Amr H. Sawalha<sup>4</sup>, Joseph Patrick McGowan<sup>5</sup>, Michael H. Dosik<sup>6</sup>, Haner Direskeneli<sup>7</sup>, Guher Saruhan Direskeneli<sup>8</sup>, Sally N. Adebamowo<sup>9,10</sup>, Clement A. Adebamowo<sup>9,10,11</sup>, Mohammad Sajadi<sup>11</sup> and Rafael Contreras-Galindo<sup>1,13\*</sup>

## Abstract

**Background:** Human Endogenous Retroviruses type K HML-2 (HK2) are integrated into 117 or more areas of human chromosomal arms while two newly discovered HK2 proviruses, K111 and K222, spread extensively in pericentromeric regions, are the first retroviruses discovered in these areas of our genome.

**Methods:** We use PCR and sequencing analysis to characterize pericentromeric K111 proviruses in DNA from individuals of diverse ethnicities and patients with different diseases.

**Results:** We found that the 5' LTR-*gag* region of K111 proviruses is missing in certain individuals, creating *pericentromeric instability*. K111 deletion (–/– K111) is seen in about 15% of Caucasian, Asian, and Middle Eastern populations; it is missing in 2.36% of African individuals, suggesting that the –/– K111 genotype originated out of Africa. As we identified the –/–K111 genotype in Cutaneous T-cell lymphoma (CTCL) cell lines, we studied whether the –/–K111 genotype is associated with CTCL. We found a significant increase in the frequency of detection of the –/–K111 genotype in Caucasian patients with severe CTCL and/or Sézary syndrome ( $n = 35$ , 37.14%), compared to healthy controls ( $n = 160$ , 15.6%) [ $p = 0.011$ ]. The –/–K111 genotype was also found to vary in HIV-1 infection. Although Caucasian healthy individuals have a similar frequency of detection of the –/– K111 genotype, Caucasian HIV Long-Term Non-Progressors (LTNPs) and/or elite controllers, have significantly higher detection of the –/–K111 genotype (30.55%;  $n = 36$ ) than patients who rapidly progress to AIDS (8.5%;  $n = 47$ ) [ $p = 0.0097$ ].

**Conclusion:** Our data indicate that *pericentromeric instability* is associated with more severe CTCL and/or Sézary syndrome in Caucasians, and appears to allow T-cells to survive lysis by HIV infection. These findings also provide new understanding of human evolution, as the –/–K111 genotype appears to have arisen out of Africa and is distributed unevenly throughout the world, possibly affecting the severity of HIV in different geographic areas.

**Keywords:** HERV-K, K111, Pericentromeric instability, Centromeres, CTCL, AIDS, HIV

\* Correspondence: [rafaelc@med.umich.edu](mailto:rafaelc@med.umich.edu); [rcontrer@umn.edu](mailto:rcontrer@umn.edu)

<sup>1</sup>Division of Infectious Diseases, Department of Internal Medicine, University of Michigan, Ann Arbor, MI 48109, USA

<sup>13</sup>Hormel Institute, University of Minnesota, Austin, MN 55912, USA

Full list of author information is available at the end of the article



## Background

The Human Endogenous Retrovirus HERV-K HML-2 (HK2), first discovered in teratocarcinoma cell lines [1] and sequenced in 1986 [2], is one of the most recent retrovirus groups to have entered the genome of the primate lineage. They arose through a number of single exogenous infections of germ line cells over the last 35 million years of primate evolution [3–14]. HK2 proviruses, which use a Lys tRNA as a primer for reverse transcription, were discovered in human cells due to their similarity to the Mouse Mammary Tumor Virus (MMTV). These proviruses have 5' and 3' long terminal repeats (LTRs) that were identical at the time of integration but accumulated mutations over time. Some HK2 entered the genome after the Homo-Pan divergence about 6–8 million years ago, and therefore are only found in either humans or chimpanzees [13]. With time, most of these integrated proviruses became silenced by introduction of stop codons, frame shifts, and indels mutations. One major type of inactivation originated by recombination between the 5' and 3' LTRs of full-length proviruses, creating hundreds of solo LTRs. Utilizing the human genome assembly (GRCh37/hg19), early estimates suggested that there are approximately 91 full-length HK2 proviruses and 944 solo LTRs [15]. At least 11 of these HK2 are polymorphically inserted into the human genome [5, 10, 12, 13, 15–18]. Mining of data generated by Next Generation sequencing from the Cancer Genome Atlas Project and the WGS500 project uncovered 17 new HK2 proviruses. Some of these HK2 proviruses are present polymorphically in only 2 of 358 individuals studied and some in over 95% of individuals [19]. More recent mining of data from the 1000 Genome project and the Human Genome Diversity Project [20] revealed the existence of 36 more HK2 proviruses, which were found to be present in <0.05 to 75% of the different populations studied. Five of these sites represented new solo LTRs. One new HK2 provirus, Xq21.33, has intact ORFs for all viral genes and is polymorphically present in Pygmy and Nigerian populations [20]. Only two other nearly intact proviruses, notably K113 [21, 22] and K115 [23] theoretically could also be able to produce infectious viral particles, but studies have shown these viruses are not capable of replication.

Recently, we discovered three new type 1 HK2 endogenous retroviruses, which we named K111, K222, and a recombinant virus K111/K222 [24–26]. Unlike the other 117 known proviruses, these newly discovered proviruses are found in the centromeric and pericentromeric regions of the human genome, restricted to 15 specific chromosomes. They are somewhat similar to K105 [5] and K112 sequences previously reported [27]. K111 proviruses are mostly present in the centromeric region of chromosomes 21 and 22 [25]. The progenitor

of K111 appears to have integrated before the Homo-Pan divergence and expanded in copy number during the evolution of hominids but not the chimpanzee. K111 is present in the chimpanzee as one copy in the telomere of chromosome 7, but it is found in about 5 copies in the Denisovan and Neanderthal genomes [28–31] and in several hundred copies in modern humans [25]. Shortly after we discovered K111, we found another centromeric endogenous provirus that we named K222, which is present mostly in the pericentromeric region of chromosomes 13, 14, and 15 [26]. K222 infected the germ line approximately 25 million years ago and is present as a single copy in the genomes of high order primates [26]. K111 and K222 share some homologous regions, and some recombinant sequences of these two proviruses exist in the centromeres as well. It is not yet known how many K111, K222, and recombinant K111/K222 proviruses actually exist, but based on quantitative assays and deep sequencing we have estimated that there may be up to several hundreds to thousands of such proviruses present in the germ line of modern man [25].

In contrast to the previously known 117 proviruses, which spread mostly by infection, K111 and K222 proviruses appear to have spread in human centromeres via homologous recombination [24–26]. K222 lacks the 5' LTR and *gag* gene, and is inserted into a pericentromeric repeat, termed pCER [26]. K111 proviruses have a 5' and 3' LTR and mostly inactive *gag*, *pol* and *env* genes. K111, which is a Type 1 HK2, however, has an intact ORF for the accessory gene *np9* [32–36]. This oncogene arose after the split of Hominids from old world primates and is present in gibbons [33], gorillas, orangutans, chimpanzees and humans. It makes a 74 amino acid protein called Np9, which binds to the ubiquitin ligase MDM2 and inhibits its activity towards p53 [33]. It also acts as a molecular switch for co-activating beta-catenin, ERK, Akt and Notch1, and promotes the growth of human leukemia stem/progenitor cells [34]. Np9 also interacts with the promyelocytic leukemia zinc finger protein (PLZF), a tumor suppressor and transcriptional repressor [35], as well as with Ligand of numb protein X (LNX), ultimately increasing the transactivation activity of Notch and likely affecting tumorigenesis [36]. Interestingly, there are different forms of *np9* transcripts produced by K111s that come from their different sequence variants [32]. While the function of the canonical Np9 has begun to be elucidated, nothing is known about the function of Np9s expressed by K111.

In our original studies of K111, we found that a patient with Sézary syndrome (Sz), a severe form of cutaneous T-cell lymphoma (CTCL), was missing K111 proviruses from her genome. We also discovered that the CTCL cell line Hut 78, and a derivative of that cell line, H9

[37], that was developed to grow HIV at high concentrations, was also missing K111, a genotype we refer as  $-/-$ -K111 in this study. A  $-/-$ -K111 genotype would also indicate that hundreds or thousands of K111 proviruses are missing from the pericentromeres creating *pericentromeric instability*, a concept defining the high frequency of deletions in the pericentromeric area of our genome. As a result, we decided to study whether patients with CTCL have a  $-/-$ -K111 genotype. In addition, we studied the prevalence of the  $-/-$ -K111 genotype in human populations and the association of the  $-/-$ -K111 genotype with disease entities of humans. We also further characterized K111 sequences in chromosomes 21 and 22 to better define the heterogeneity of K111s in human centromeres.

## Methods

### Study samples

DNA samples from Human/Rodent somatic hybrid cell lines (each one containing a single human chromosome) and their parental rodent cells were obtained from the NIGMS Human/Rodent Somatic Cell Hybrid Mini Mapping Panel # 2 DNA, Coriell Cell Repositories.

DNA samples from people of different genders and ethnic origins were obtained from the HRC2 Human Random Control DNA panel 2 (Sigma Aldrich, 96 Caucasian people) and from Human variation panels HD03 (Indo Pakistani), HD05 (Middle Eastern), HD07 (Japanese), HD12 (Africans South of the Sahara), HD20 (Russian Krasnodar), HD21 (Italian), HD22 (Ashkenazi Jewish), HD32 (Chinese), and samples from Mb pygmy (NA10492, NA10493, NA10494, NA10495, NA10496) from the Coriell Cell Repositories; these DNA samples were previously used in another study [26].

DNA from a cohort of HIV patients was obtained from the peripheral blood lymphocytes (PBLs) of patients cared for at the University of Maryland, Baltimore, MD; North Shore University Hospital, Manhasset, NY; and the University of Michigan, Ann Arbor. Other samples from HIV-infected individuals were obtained from the PBLs and or EBV transformed cells from patients enrolled in the MACS, a natural history study of men who reported having sex with men. The study participants were over age 24 to age 79.

Samples of whole blood and/or PBLs extracted from whole blood were also obtained from patients with breast cancer, different lymphomas, cutaneous T-cell lymphoma (CTCL), chronic lymphocytic leukemia (CLL), and normal controls at the University of Michigan Health System and the VA Ann Arbor Healthcare System, Ann Arbor, MI, and/or at North Shore Hematology/Oncology Associates, East Setauket, NY. All patients had pathologically confirmed cancers. The patients in the study were over age 30 to age 82.

DNA from either PBLs and/or PBLs transformed by EBV from Caucasian patients with known psoriasis was obtained from a cohort of patients cared for at the University of Michigan Dermatology Department, ages 22 to 73 years. DNA extracted from peripheral blood mononuclear cells (PBMCs) from selected patients with known lupus erythematosus and with Takayasu arteritis was obtained from patients in studies at the Division of Rheumatology, University of Michigan, and Mamara and Istanbul University in Istanbul, Turkey, ages over 30 years. DNA from PBLs of patients from Nigeria, ages 35 to 72 years, was obtained from a study of patients with breast cancer in Nigeria conducted at the Institute of Human Virology, University of Maryland, Baltimore.

All patients with CTCL had biopsy-proven disease, with biopsies reviewed at the University of Michigan. Two patients with Sézary syndrome (Sz) were cared for elsewhere, one at the University of Michigan and the Dana Farber Hospital, Boston, MA and a second one (our index case) at Yale University, New Haven, CT and North Shore University Hospital Manhasset, NY. Patients considered to have Sz had evidence of circulating atypical T-cells (Sézary cells) early in disease, which were usually "CD4+, CD45RO+ with frequent loss of T-cell surface antigens CD2, CD5, and/or CD7 [38]. Most circulating Sézary cells are CD4+, CD7-, and CD26-. There was usually a CD4+/CD8+ ratio of >6, often with greater than 1000 aberrant cells present in circulation. Epidermotropism, Pautrier's abscess, and haloed lymphocytes were not common. Classic cutaneous T-cell lymphoma (CTCL) was defined pathologically as consisting of a proliferation of mature CD4+ CD45RO+ memory T-cells early in disease. In almost all cases, there was a confirmatory diagnostic test for T-cell clonality with alpha/beta or gamma/delta T-cell receptor (TCR) gene rearrangements [39–43]. Large cell transformation (LCT) was determined when patients with pre-existing CTCL showed evidence of a morphologic change in small to medium sized atypical T-cell lymphocytes to a large cell variant, usually in >25% of the lymphocyte population. These cells are typically CD30+, but can be also CD30- with usually no more than 75% of these cells having large cell morphology. These cells often show folliculotropism [44, 45].

We obtained samples from HIV-infected patients with different rates of disease progression in the MACS cohort and from other institutions. For the purposes of this study, we will use the term LTNP to include: 1. *Elite Controllers (EC)*, who had a viral load of less than 50 HIV RNA copies/ml on 2 or more occasions within 1.5 years in the absence of therapy (one detectable measurement of less than 1000 copies/ml was allowed between the two suppressed values). 2. *Viremic Controllers (VC)*, who had

a viral load of less than 2000 HIV RNA copies/ml on two or more tests within 1.5 years in the absence of therapy (no viral load spikes were allowed). 3. Other *Long-Term Non-Progressors (LTNPs)*, patients who remained AIDS-free without therapy for 15 or more years. A second category of patients are called Intermediate Progressors, and are people who developed AIDS 5 to 12 years following seroconversion. The third category, Rapid Progressors, defines patients who developed AIDS within 3 years of seroconversion in the MACs cohort or from transfusion 3 years earlier at our combined institutions. It also includes patients we call *First AIDS* who appeared with an opportunistic infection, a severe neurological disease, a lymphoma or Kaposi sarcoma, and/or a CD4 count of less than 200 cells/mL of blood when they were first seen. In some patients, treatment with anti-retroviral drugs prevented determining the natural history of the disease. These samples comprise our fourth category, which we call “Unclassified”.

#### Cell lines

The following cell lines were utilized in this study: Hut 78, and derivatives of Hut 78, notably H9 and H9/HTLVIII. These cell lines were obtained from the AIDS Research and Reference Reagent Program. Cells were maintained in RPMI medium and supplemented with 10% fetal bovine serum. Human/Rodent somatic hybrid cell lines were obtained from a human chromosomal DNA mapping panel (NIGMS Human/RodentSomatic Cell Hybrid Mini Mapping Panel # 2 DNA Coriell Cell Repositories).

#### DNA extraction

DNA was extracted from PBLs and/or cell lines using the DNeasy blood and tissue kit™ (Qiagen, <http://www.qiagen.com>) or from whole blood or cell lines using the Gentra Puregene Blood Kit™ (Qiagen).

#### Detection of K111 insertions

Centromeric K111s were amplified by PCR with the primers P1, which binds to K111 flanking centromeric repeat CER:D22Z3, and primer P4, which binds to the K111 *gag* gene to produce a ~ 1.6 Kb amplification product [25]. The PCR was carried out using the Expand Long Range dNTPack PCR kit (Roche Applied Science, Indianapolis, IN). PCR was performed at a final volume of 50 µl using an initial step of 92 °C for 2 min, followed by 35 PCR cycles consisting of denaturation at 92 °C for 15 s, annealing at 55 °C for 30 s, and extension at 68 °C for 5 min. Amplification products were confirmed by sequencing [25].

#### Walking the 5' genome of K111

PCR amplification was carried out using the primer P1 that binds to the K111 flanking centromeric repeat CER:D22Z3 and reverse primers located on the *gag*, *pro*, and *pol* of HK2, called P4, K111 986R, K111 1584R, K111 2499R, and K111 3460R. The amplification conditions are the same as those used above.

#### Long-range amplification and sequencing of K111

Production of K111 long sequences of chromosome 21 and 22 for Pac Bio sequencing were carried out using the primer set K111 5359F and P2R. Long sequences for CTCL and HIV samples and cell lines HUT 78 and H9 were produced using a primer set P1 and K111 2499R that binds to the integration site of K111 and to the *gag* region of K111, or a set of primers K111 6353F and P2R that bind to the *env* gene of K111 before the splice donor site of *np9* [30] and the 3' integration site, respectively. Amplifications were carried out with four different bar codes attached to the primer K111 2499R or to K111 6353F, respectively. Reactions were carried out in 50 µL using 250–500 ng of total DNA per reaction with the following conditions: 92 °C for 2 min, and 40 cycles consisting of 92 °C for 15 s, 55 °C for 15 s, and 68 °C for 5 min using the Expand Long Range dNTPack PCR kit (Roche Applied Science, Indianapolis, IN).

The bar-coded PCR products were size-selected and purified using the Blue Pippin gel analyzer. The chromosomal products were purified using similar conditions. Adapters were ligated onto the PCR products and the libraries were sequenced using the P6 polymerase kit and C4 sequencing reagents on a PacBio RS II SMRT DNA Sequencing System. The sequences were processed using the PacBio SMRT portal. The sequences were de-barcoded and processed using the RSReadsOfInsert software filtering for sequences that pass five times around each SMRT cell to obtain quality values ~ 99.9%. Sequences were de-duplicated and the edited sequences were aligned to the K111 reference genome (Acc No. GU476554.2) using the RS Resequencing and Quiver v1 parameters. In order to determine sequence variants of proviruses among human populations we used the minor variant calling platform in the SMRT portal for reads aligned to K111 consensus sequence that have passed five times around each SMRT cell. The sequence alignments were visualized in the IGV viewer.

#### Real-time qPCR specific for centromeric K111 and K222

The copy number of K111 found in the cellular DNA was measured by qPCR using the primer K111 1584F and a modified primer K111 1701R that has locked nucleic acid (LNA) in the CAA sequence underlined in the primer sequence below, which is highly specific for K111 to produce a 117 bp product that is readily detected

using Sybrgreen. This LNA primer detects specific mutations in K111 *gag* but not other HK2s [25]. The qPCR was performed using the goTaq<sup>®</sup> qPCR master mix in a reaction volume of total 20  $\mu$ l. The following PCR conditions were used: 10 min at 95 °C, followed by 50 cycles consisting of 30 s of denaturation at 95 °C, and 15 s of annealing/hybridization at 60 °C. The PCR reactions were run over 2 h. The relative K111 copy number was estimated using serial dilutions of DNA from one patient with the highest K111 copy number, starting at a concentration of 1000 ng per reaction. It was felt that total human DNA from a +/+K111 CTCL patient would be a more robust standard than K111 clone DNA. This DNA was frozen in small aliquots and used once in each run after defrosting. The specific detection of K111 was verified by sequencing analysis of the PCR product. The copy number of K222 in cellular DNA was measured by qPCR using a probe that specifically discriminates the K222 pCER-*pro* boundary as previously described [25].

### Statistical analysis

Statistically significant differences in the mean copy number (K111 and K222 copies/genome calculated by qPCR in each group) were determined using the Student's t-test. Frequencies of detection of K111 5' LTR among the groups were compared by chi-square ( $X^2$ ) analyses. Two-tailed *p* values were considered significant at *p* < 0.05.

### List of primers

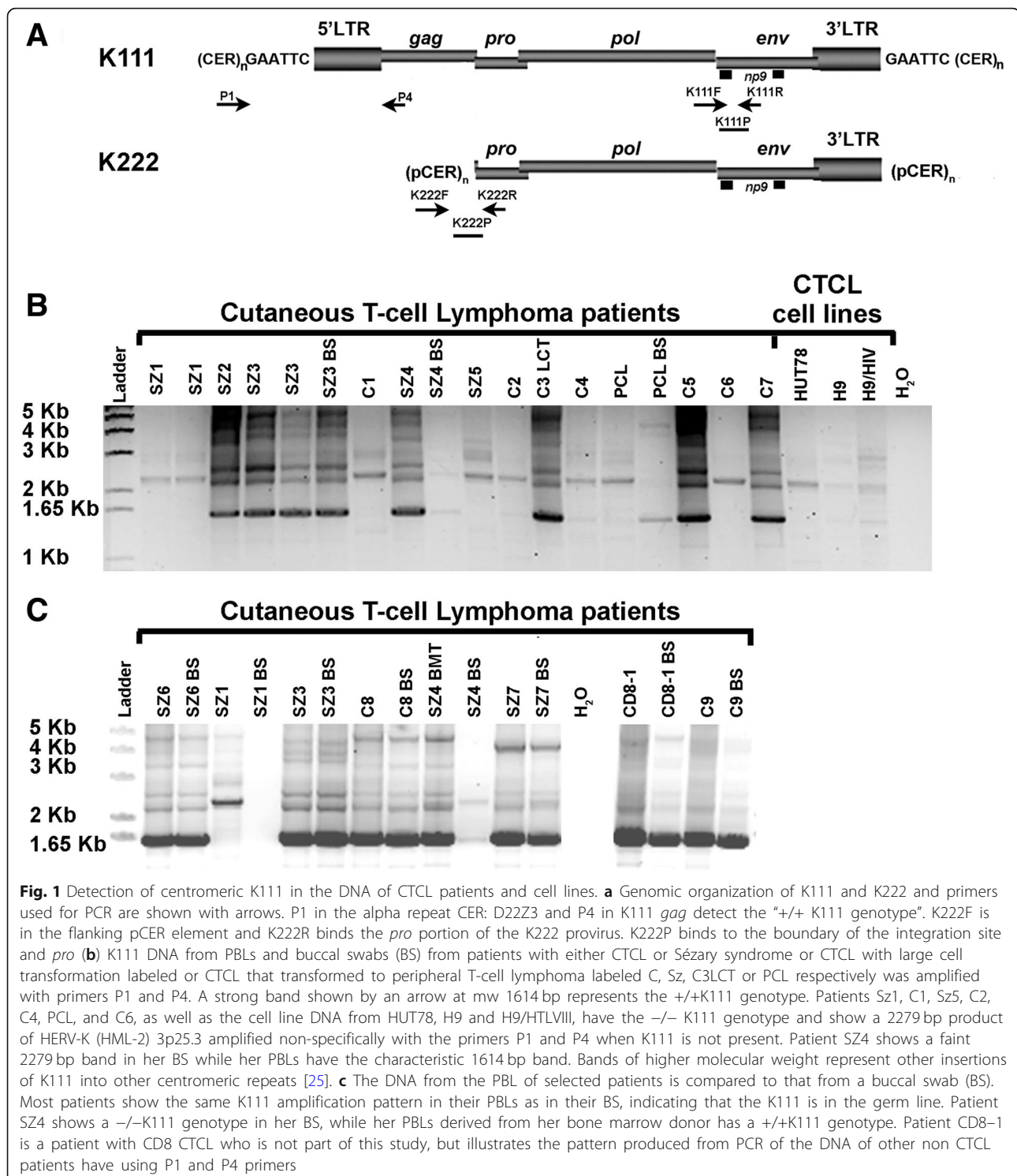
P1F: 5'-ACA TTC AGA CCA TGG TAG CCG TGT -3'  
 P2R: 5'-ACA GTG CTG TGT GGG TCT GAA TGA -3'  
 P4R (K111 986R): 5'-GTA CCT TCA CCC TAG AGA AAA GCC T -3'  
 GAPDHF: 5'-TGC ACC ACC AAC TGC TTA GCA CCC-3'  
 GAPDHR: 5'-CTT GAT GAC ATC ATA TTT GGC AGG-3'  
 K111 1584F: 5'-TCC TTA AGG TCA TAG TGG AGT TGT TGG TAT AC-3'  
 K111 1701R LNA: 5'-CAT AAG CAT AGC TTT ATG CAA AC-3'  
 K111 1965R: 5'-TCC AGG TGC CAT CGG TTG CAT-3'  
 K111 2499R: 5'-TTG AGC AAC ATC TTG GAG CCT TGC-3'  
 K111 3460R: 5'-ACT TGC CCA ATA TGC AGC CTT TCC-3'  
 K111 6353F: 5'-GAC ATG GGA AAT AGG GAA GGT AAT A-3'

K111 5359F 5'-TCG GCT CAA AGA GCA GAG ATG GTT-3'  
 K111F: 5'-AAG AGC ACC AGG ATG CTT AAT GCC-3'  
 K111R: 5'-AGT GAC ATC CCG CTT ACC ATG TGA-3'  
 K111P: 5'-FAM-TGC CGG TCC TAA CAG TAG ACT CAC-BHQ1-3'  
 K222F: 5'-CAG CGT TCT GGA ATC CTA TGT-3'  
 K222R: 5'-TGT ATT GTG GTA ACT GGG TAT ATG T-3'  
 K222P: 5'-FAM- ACC CAC ATG GCA GTG TTC TGG ATT-BHQ1-3'  
 Bar codes used for Pac Bio Sequencing;  
**B** ACATCG, **E** CACTGT, **G** GATCTG, **J** AAGCTA

## Results

### Detection of a -/-K111 [ $\Delta$ LTR-*gag*] genotype in CTCL patients

As we discovered the -/-K111 genotype in CTCL cells, we collected DNA from patients with CTCL and Sz (an aggressive form of CTCL) to determine whether the lack of centromeric 5' K111 LTR-*gag* (-/-K111 genotype) is prevalent in this disease. DNA from CTCL patients was screened for the K111 provirus using a forward primer P1, which binds to the flanking CER:D22Z3 region, and a reverse primer P4, which binds in K111 *gag* (Fig. 1a). This set of primers amplifies centromeric K111 proviruses in 15 human chromosomes [24, 25]. The PCR reaction amplifies multiple products, but the major band, which we call the K111 *beta* band, at a molecular weight of ~ 1614 bp (Fig. 1b), represents multiple centromeric K111 proviral sequences from many chromosomes as determined by cloning and sequencing of this amplification product in our previous study [25]. Several CTCL patients did not show the characteristic broad K111 *beta* band (Fig. 1b and c). When the DNA samples lack the K111 *beta* band, a non-specific amplification product of ~ 2279 bp is detected, which corresponds to the LTR of HK2 3p25.3 inserted in chromosome 3, as determined by sequencing (Fig. 1b and c). Patients with a -/-K111 genotype also do not have the characteristic higher molecular weight gamma and delta bands, which are other K111 proviral sequences flanked by longer centromeric CER elements [25]. A -/- K111 genotype is also seen in the Hut 78 cell line and derivatives of Hut 78, notably H9 and H9/HTLVIII (Fig. 1b). We previously showed that using 4 additional primers that bind to other areas of the 5' flanking region of K111 still did not produce amplification products in -/- K111 cells. This confirmed that the -/- K111 genotype observed with the primers P1 and P4 is not the result of mutations in the 5' flanking area preventing the binding of the P1 primer during PCR, but is due to absence of K111 provirus [26].



The patients with CTCL who were in this study were characteristically much sicker than ordinary CTCL patients having failed UV light treatments and/or topical steroids. They ultimately required more invasive therapies (Additional file 1: Table S1 and Additional file 2: Table S2) to control refractory skin tumors, or they had

developed large cell transformation (LCT), new cancers, or uncontrolled Sz.

Interestingly, the frequency of detection of the –/–K111 genotype in these patients was significantly higher ( $p = 0.0083$ ) in individuals with CTCL (37.14%;  $n = 35$ ) than in those of healthy individuals of the same

ethnicity (15.6%;  $n = 160$ ) (Table 1), suggesting that a  $-/-$ K111 genotype is associated with severe CTCL. The frequency of detection of the  $-/-$ K111 genotype was significantly higher in patients with the progressive stages of the disease such as Sz (45.5%) [ $p = .039$ ;  $n = 11$ ], in contrast to those that have uncomplicated but still severe CTCL (30.7%).

Some CTCL patients also developed multiple other cancers, as shown in Additional file 1: Table S1 Patients who have the  $-/-$ K111 genotype had more secondary cancers (11/13, [84%]) in contrast to those who are  $+/+$ K111, (17/26 [62.9%]), but this was not statistically significant ( $X^2$ : 1.58,  $p = 0.20$ ; Additional file 1: Table S1). Patients who are  $-/-$ K111 were less likely to have formed tumor nodules and/or large plaques that require multiple sessions of radiation therapy (4/13 [30.8%] receiving 18 treatments) than those with a  $+/+$ K111 phenotype (15/26 [52.7%] receiving as many as 59 treatments), but this was not statistically significant ( $X^2$ : 2.51,  $p = 0.1128$ ). During each treatment period of radiation, patients had from 1 to > 5 areas of skin radiated. On the other hand, the mortality of  $-/-$ K111 patients was higher (61.5%  $n = 8$ ) than  $+/+$ K111 (34.6%  $n = 9$ ) and these patients needed whole body radiation (4/13 30.7%) vs. (5/26 19.2%) as well as bone marrow transplantation (2/13 15.3% vs. 3/26 11.5%) to control disease, again all not statistically different.

Having a  $-/-$ K111 is uncommon, and indicates that K111 proviruses are completely absent from the centromeric regions of these CTCL patients, something we also refer to in this paper as *pericentromeric instability*. As we do not know whether or not the K111 genotype might vary among the cell types of the body, we obtained buccal mucosal swabs (BS) to search for the presence of K111 in healthy cells from some CTCL patients. As seen in Fig. 1c, the pattern of detection of the K111 genotype was similar in the buccal mucosal and the peripheral blood lymphocytes (PBLs) DNA, indicating that the lack of K111 or *pericentromeric instability* is found

in the germ line of these patients. Of interest, one patient who had severe Sz, and who received an allogeneic bone marrow transplant, had a  $+/+$ K111 genotype detected in her transplanted peripheral blood lymphocytes (PBLs) but had a  $-/-$ K111 genotype in her buccal mucosal cells (Patient SZ4, Fig. 1b and c). We could deduce then that these  $+/+$ K111 PBLs represent the engrafted cells from her transplant.

#### Detection of $-/-$ K111 genotype in diverse populations and medical diseases

To put the loss of K111 s or *pericentromeric instability* in CTCL patients in perspective, we were interested in knowing the prevalence of the  $-/-$ K111 genotype in diverse populations. To do so, we obtained a panel of DNA from healthy Caucasian individuals from England [26] and DNA samples from people of different ethnicities around the world (Table 2). We also collected PBLs and isolated DNA from healthy Caucasian persons as well as from African patients from Nigeria and from Hispanic individuals as shown in Table 2.

We found that the prevalence of the  $-/-$ K111 genotype in healthy Caucasians individuals was 16.11% ( $n = 180$ ), which stands in stark contrast to healthy African populations ( $n = 127$ ; Table 2), whose prevalence of  $-/-$ K111 was 2.36% ( $p = 0.000169$ ). The prevalence of  $-/-$ K111 was higher in African American individuals (7.6%;  $n = 39$ ) but this was not statistically significantly different from the prevalence in Caucasian HIV patients and or Caucasian Englishmen. We did not find significant differences in the prevalence of the  $-/-$ K111 genotype in populations from Asia or Europe compared to Caucasian Englishmen or Caucasian Americans, except for a slight increase in a population from Siberia (Table 2). A larger sample will be needed in order to determine whether a significant increase in the  $-/-$ K111 genotype exists in the Siberian population.

We further evaluated the K111 genotype in a collection of DNA from Caucasian patients with psoriasis,

**Table 1** Frequency of Detection of K111 in DNA from Cutaneous T-cell Lymphoma (CTCL) patients

Disease	No. tested	$-/-$ K111	Frequency of $-/-$ K111 (%)	$\chi^2$	$p$ value
CTCL Total	35	13	37.14	6.96	0.0083
CTCL	13	4 <sup>b</sup>	30.7	0.32	ns
CTCL + LCT	11	4 <sup>c</sup>	36.4	2.97	0.0423
Sézary Syndrome	11 <sup>a</sup>	5	45.5	4.260	0.0390
Caucasian CTCL	35	13	37.1	8.475	0.003601
African American CTCL	4	0	0	24	ns
Caucasian healthy	160	25	15.6		

The table shows the frequency of detection of the  $-/-$ K111 genotype in patients with CTCL. This proportion is compared to the prevalence of  $-/-$ K111 in a population of healthy Caucasian individuals. <sup>a</sup>Two patients with Sz also developed LCT. Neither of these patients was  $-/-$  K111. <sup>b</sup>One patient developed peripheral T-cell lymphoma and a second patient developed Hodgkin Lymphoma. <sup>c</sup>One patient developed concomitant Hodgkin Lymphoma. LCT Large cell transformation, Sz Sézary Syndrome.  $p$  values were calculated using a  $\chi^2$  test

**Table 2** Frequency of Detection of K111 in DNA from individuals of different ethnic groups

Ethnicity	No. tested	-/- K111	Frequency of -/- K111 (%)	$\chi^2$	<i>p</i> value
Caucasian	180	29	16.11		
English	96	13	13.54		
American	64	12	18.75		
Italian	10	3	30		
Ashkenazi	10	1	10		
African total	127	3	2.36	14.1415	0.000169*
South Sahara	9	0	0	1.65	ns
Pygmy	5	0	0	0.921	ns
Nigerian	113	3	2.65	12.1039	0.000503*
Asian total	54	12	22.22	1.2288	ns
Asian (MACS)	25	6	24	1.0872	ns
Chinese	10	2	20	0.0506	ns
Japanese	10	1	10	0.3408	ns
Southeast Russia	9	3	33.3	1.933	ns
Amerindian (MACS)	21	3	14.3	0.0255	ns
African American	39	3	7.6	1.82	ns

Stars indicate significantly decreased detection of the -/- K111 genotype in Africans, including Nigerian individuals, as compared with the Caucasian population. *p* values were calculated using the  $\chi^2$  test

HIV, lupus erythematosus, Takayasu arteritis, breast cancer and leukemia/lymphoma to determine whether the -/-K111 genotype might be associated with other skin or lymphocytic disorders and/or cancers other than CTCL. To study the prevalence of a -/-K111 genotype in a skin disorder other than CTCL, we screened the DNA samples of Caucasian individuals with psoriasis who lived in the Michigan area (Table 3). The prevalence of -/-K111 in these patients was 15.4% ( $n = 123$ ), similar to the prevalence in healthy Caucasians. We did not find differences in the outcome of the -/-K111 genotype in psoriasis patients compared to the +/+K111 genotype as judged by nail involvement, the need for biologic or systemic treatment, the age of onset and/or any unique HLA type.

We were also interested in studying whether the prevalence of the -/-K111 genotype or *pericentromeric instability* might influence the outcome of another T-cell disorder, notably HIV (Table 3). We screened samples from Caucasian, African American, and Hispanic patients enrolled in the Multicenter AIDS Cohort Study (MACS) and samples from three other institutions using the categories defined above, i.e. LTNPs, Intermediate Progressors, Rapid Progressors, and unclassified.

Overall, the prevalence of the -/-K111 genotype in HIV-infected Caucasians was 15.2% ( $n = 164$ ), similar to healthy Caucasian individuals. Interestingly, the prevalence of the -/-K111 genotype in Caucasian patients with different presentations of HIV infection was 30.55%

( $n = 36$ ) in HIV LTNPs or controllers, 18.75% ( $n = 32$ ) in Intermediate Controllers, and 8.5% ( $n = 47$ ) in Rapid Progressors (Table 3). A statistically significant difference was found in the prevalence of detection of the -/-K111 genotype between the group of HIV LTNPs and all Rapid Progressors ( $p = 0.0097$ ).

We next evaluated the prevalence of the -/-K111 genotype in other cancers as well as in autoimmune disorders. The prevalence of -/-K111 in Caucasian women with breast cancer was 9.4% ( $n = 53$ ) (Table 3) while in African women with breast cancer it was 1.44% ( $n = 139$ ) (Table 4). This difference is likely linked to the ethnicity rather than the disease, as the prevalence of detection of the -/-K111 genotype in African patients with breast cancer is similar to that found in modern healthy African populations of 2.36% ( $n = 127$ ), suggesting again that the -/-K111 genotype originated out of Africa. In Caucasian patients with systemic lupus erythematosus and Takayasu arteritis, the prevalence of the -/-K111 genotype was 18.4% ( $n = 38$ ) and 10% ( $n = 30$ ), respectively (Table 3), similar to the prevalence in healthy Caucasian individuals.

#### Mapping of centromeric proviruses in CTCL patients

We have shown that K111 is present in 15 different human centromeres [25], with chromosomes 21 and 22 having the highest number of these proviruses. Recently, we described the discovery of a second centromeric HK2 we call K222 that resides in many



**Table 3** Frequency of Detection of  $-/-$  K111 in DNA in Caucasian patients with different diseases

Disease	Origin	No. Tested	$-/-$ K111	Frequency of $-/-$ K111 (%)	$\chi^2$	P value
<i>Skin Diseases</i>						
Psoriasis	Michigan	123	19	15.4	0.0017	ns
<i>HIV Infection</i>						
HIV LTNP	MACS	25	5	20	0.3046	ns
	Michigan, New York	11	5	45.5	6.3308	0.0118
Total LTNP		36	11	30.55	4.39 6.691	0.0366 0.0097*
Intermediate	MACS	29	6	20.7	0.4593	ns
	Michigan, New York	3	0	0	0.554	ns
Total Intermediate		32	6	18.75	0.192	ns
Rapid AIDS	MACS	30	3	10	0.6362	ns
	New York, Michigan	17	1	5.9	1.16	ns
Total Rapid AIDS		47	4	8.5	1.526	ns
Unclassified	New York, Michigan, MACS	49	5	10.2	0.896	ns
Total		164	25	15.2	0.009	ns
<i>Solid tumors</i>						
Breast Cancer	New York, Michigan	53	2	9.4	1.26	ns
<i>Lymphomas</i>						
Hodgkin	Michigan	3	0	0	0.554	ns
Mantle cell	Michigan	6	1	16.5	0.048	ns
Marginal zone	Michigan	5	1	20	0.0699	ns
DLBCL	Michigan	17	1	5.9	0.2806	ns
Follicular	Michigan	23	3	13.0	0.1034	ns
CLL	Michigan	11	2	18.2	0.0506	ns
Total		65	8	12.3	0.406	ns
<i>Autoimmune Disorders</i>						
Takayasu	Turkey	30	3	10	0.6362	ns
Lupus	Michigan	38	7	18.4	0.1772	ns

The star indicates a significant difference of the frequency of detection of the  $-/-$ K111 genotype between the LTNP and Rapid AIDS patients

pericentromeric regions [26]. We were first able to detect this set of K222 pericentromeric proviruses in DNA samples from CTCL cells that have the  $-/-$ K111 genotype. K222 has no 5'LTR and has a deletion of *gag* and much of *pro* (Fig. 1a). K222 shares a high homology to K111 in the 3' end of the provirus. The 3' end of K222 has an LTR, but it lacks the 6 base pair target site duplication specific to K111 and is inserted in a pCER element instead of being inserted in CER:D22Z3 [25, 26].

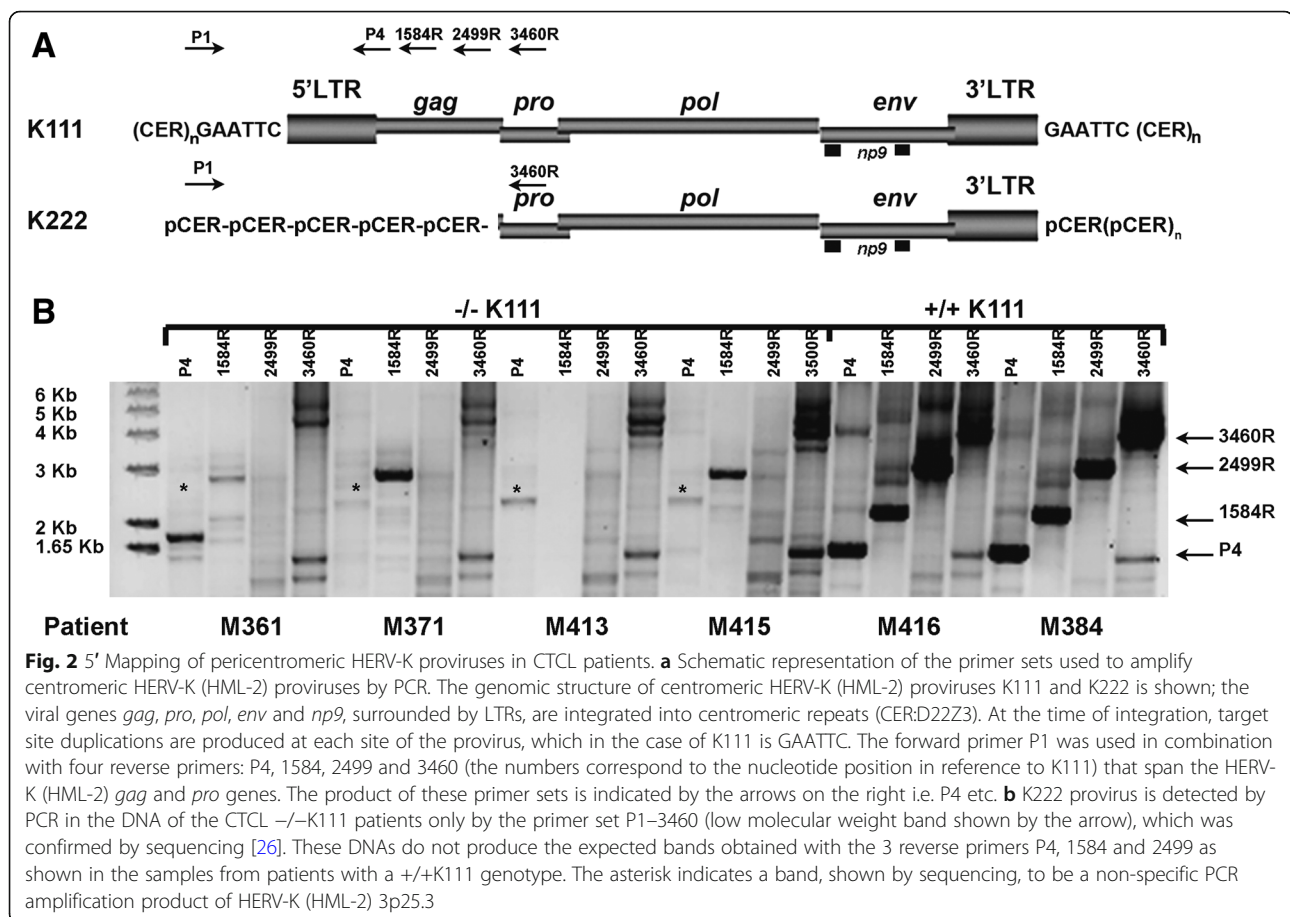
We wanted to assure ourselves that the centromeric proviruses that remain in the DNA of CTCL individuals with the  $-/-$ K111 genotype have a provirus pattern similar to the pattern we previously characterized in healthy people. To determine the structure of centromere proviruses, we mapped the DNA of individuals, using a set of primers that specifically amplify centromere K111 and

K222 [25, 26]. We found that in the DNA of CTCL patients with the  $-/-$ K111 genotype, no K111 beta band (mw 1614 bp) was seen by PCR (Fig. 2) using the primer pair P1 and P4 as also shown in Fig. 1. Similarly, no amplification products were seen in the same  $-/-$ K111 patients when the DNA was amplified with reverse primers that bind at bases 1584 in *gag* and 2499 in *pro*, areas which are absent in the K222 genome (Fig. 2). Amplification products with these primers were readily seen in an individual with a  $+/+$  K111 genotype. Using the primer that binds to 3460 of the *pro* region of both K111 and K222, a low molecular weight band was barely visible in the  $-/-$ K111 patients, while a strong band at 4240 bp was seen in  $+/+$ K111 normal patients. The low molecular weight band is from K222, which can be detected in both  $-/-$  and  $+/+$ K111 individuals. These results indicate that the pattern of genomic composition

**Table 4** Frequency of Detection of  $-/-$ K111 in DNA from African and African American patients with different diseases

Disease	Origin	No. Tested	$-/-$ K111	Frequency of $-/-$ K111 (%)	$X^2$	<i>P</i> value
African						
Healthy	Nigeria	127	3	2.36		
					14.583	0.00001#
African Breast Cancer	Nigeria	139	2	1.44	0.307	ns
					18.2223	0.00002#
<i>Total African</i>		266	5	1.8		0.00001#
HIV/AIDS	Michigan, New York	6	1	16.6	0.216	ns
unclassified	Michigan, New York	5	0	0	0.564	ns
HIV LTNP	Baltimore	18	1	5.6	0.340	ns
DLBCL	Michigan	3	1	33	1.41	ns
CTCL	Michigan	4	0	0	0.452	ns
Breast Cancer	Michigan	3	0	0	0.574	ns
Total African American all diseases		39	3	7.6	1.821	ns

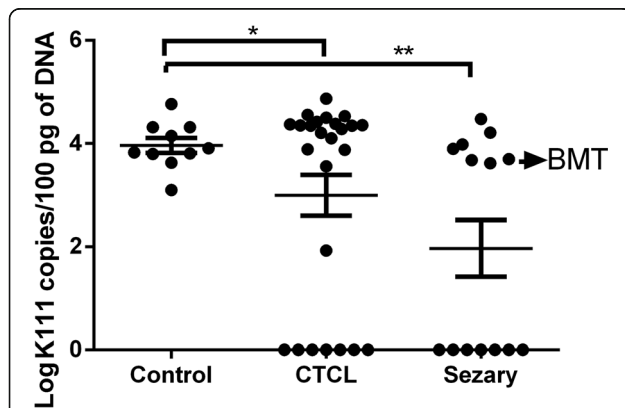
Statistical analysis was determined between healthy individuals and patients of the same ethnicity. # indicates significant differences between healthy Caucasian individuals and healthy African individuals or African patients with breast cancer



of centromere proviruses in  $-/-$ K111 is the same in all the CTCL patients studied, as we had previously determined in CTCL cell lines [25].

### Quantitation of K111 and K222 in humans

We were interested in trying to gauge how many copies of the K111 and K222 proviruses might be in a person's genome. This was important in order to elucidate if  $+/+$ K111 patients with CTCL may have reduced numbers of K111. We developed an assay using primers modified with locked nucleic acids (LNA), which bind with enhanced affinity to the target K111 sequence in the *gag* region of K111, but not to K222 or other HK2 proviruses [25]. Using this assay, we screened DNA from all of our CTCL patients, which yielded the results shown in Fig. 3. Patients with CTCL and, the severe form Sz, had statistically significantly fewer copy numbers of K111 than did control subjects. All  $-/-$ K111 patients detected with the P1 and P4 primers (Figs. 1 and 2, Table 1) had an undetectable proviral load in their genome using the LNA primer assay system. The inability to detect K111 with the LNA primer set in the *gag* region of  $-/-$  K111 individuals also indicates that the failure to detect K111 is not due to some mutation in the CER elements flanking the virus, but is due to the absence of K111. The K111 LNA assay detected numbers of K111 copies in the human population that varied within a range of 1230 to 58,000 copies/100 pg of DNA (average 11,220 copies/100 pg of DNA), which corresponds to 78 to 3480 copies/genome in somatic cells.

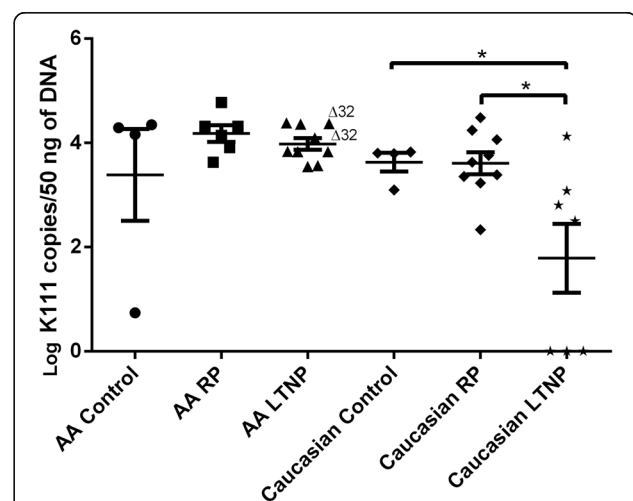


**Fig. 3** Copy number of centromeric K111 in different stages of CTCL. Centromeric K111 was quantitated in the DNA of PBLs by qPCR using a LNA primer that is highly specific for K111 *gag*. The copy number was determined in the same run using a standard curve generated by dilution of DNA chosen from a CTCL patient who had the highest levels of K111 measured in our earliest assays. Note that one CTCL patient had a relatively low copy number while all the rest of the patients have either high copy numbers of K111 or have a  $-/-$ K111 genotype. Note that the DNA from the PBLs of a patient who had a bone marrow transplant (BMT) had a high copy number of K111 while her buccal swab (Fig. 1c) is  $-/-$  K111. *p* values were calculated using a t-test. Statistical significance  $p < 0.05$  (\*),  $p < 0.01$  (\*\*)

These numbers of proviruses are similar to the ones previously estimated in our laboratory [25]. We did not find statistically significant differences in the number of copies of K111 in  $+/+$ K111 individuals who are healthy or in those who have CTCL. Only one CTCL patient had 85 copies of K111 proviruses/100 pg of DNA (Fig. 3). Nonetheless, the  $-/-$ K111 genotype was readily evaluated with the LNA technology and gave no amplification products. This suggests that any association of K111 copies with severe CTCL disease and Sz is determined by having a  $-/-$ K111 genotype rather than fewer copies of K111.

We then determined the numbers of K111 proviruses in HIV-1 infected people in order to investigate whether HIV LTNPs have fewer copies of K111 proviruses than HIV Intermediate and/or HIV Rapid Progressors (Fig. 4). Quantitation of K111 copies using our K111 LNA assay revealed that in contrast to HIV negative subjects and to all Rapid Progressors of Caucasian ethnicity, the copy number of K111 decreases significantly in HIV LTNPs. However, we did not observe this difference between HIV LTNPs of African American ethnicity as compared to their ethnic controls or African American HIV Rapid Progressors, confirming the observations noted above (Fig. 4).

Having determined that human populations of different ethnicities have two different genotypes of centromeric K111 proviruses,  $+/+$ K111 or  $-/-$ K111, we looked at centromeric K222 proviruses, which also have multiple copy numbers in the pericentromeric regions, to



**Fig. 4** Copy number of centromeric K111 in different stages of HIV progression. K111 proviral copy number was quantitated in DNA from patients with HIV using qPCR with the LNA primer specific for K111. AA RP or Caucasian RP represents African American or Caucasian patients with HIV who are HIV Rapid Progressors (RP). Note that the copy number of the Caucasian Long-Term Non-Progressors (LTNPs) was significantly lower than in Caucasian controls and RPs. *p* values were calculated using a t-test. Statistical significance  $p < 0.05$  (\*),  $p < 0.01$  (\*\*)

evaluate whether there is any difference in the numbers of K222 proviruses in CTCL or HIV populations. We estimated the numbers of K222 centromeric proviruses that exist in 9 human chromosomes to be approximately 8 to 80 copies using an assay we previously described [26]. Quantitation of K222 revealed that the number of K222 proviruses does not change significantly between control subjects, patients with CTCL or other cancers, or HIV (Fig. 5). While there was variation in the copy number of K222 in diverse human populations in this study, no significant changes were seen in any of the disease subsets as compared to controls.

#### Genetic variability of K111 proviruses in the human genome

K111s exist in the centromeres in up to 15 different chromosomes, but are mostly found in chromosomes 21 and 22. As K111s appear to have spread via homologous recombination in the centromeres at different times during the evolution of modern humans, one would expect to see genetic variation of K111 in chromosomes 21 and 22. We sequenced the K111 PCR fragments amplified from DNA from human/hamster hybrid cell lines containing human chromosomes 21 or 22 and also DNA from patients with CTCL and HIV using the PacBio platform that can

sequence single DNA fragments up to 15–20,000 bp. This approach assures that all sequences present in a PCR fragment represent one of the many K111 proviruses in that PCR product.

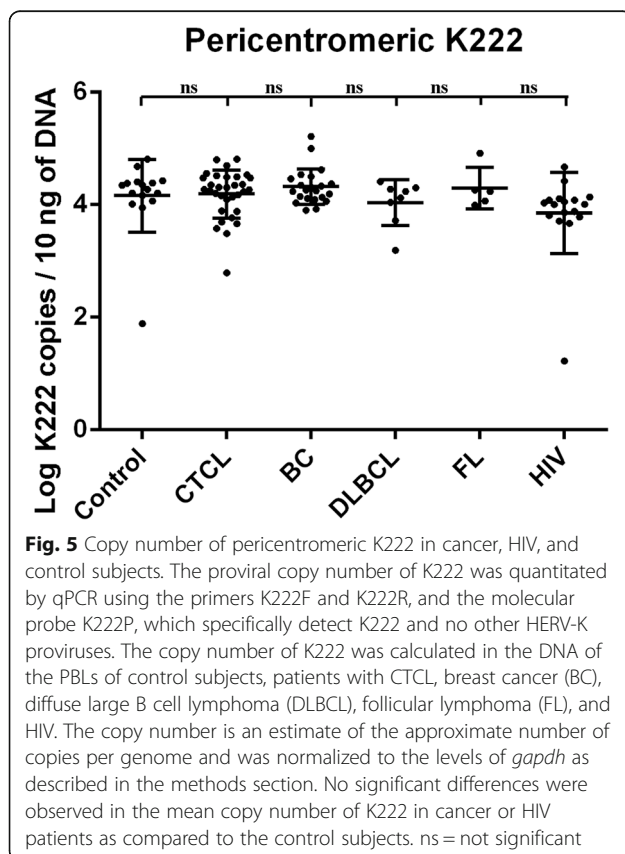
DNA sequence analysis revealed that the K111 proviruses have accumulated many sequence mutations over time (Fig. 6), with many unique to chromosome 21 or 22. The same K111 sequences seen in these hybrid cells were also present in DNA isolated from human DNA samples. This genetic variability is not some type of recombination that arose in the hybridization of these human chromosomes within hamster cells, as the hamster genome lacks HK2 proviruses, and as similar sequences were found in human DNA. The sequences that were seen had many shared indels. The numerous indels variations of K111s appear to have arisen by homologous recombination over evolutionary time, confirming our previous observations [24–26].

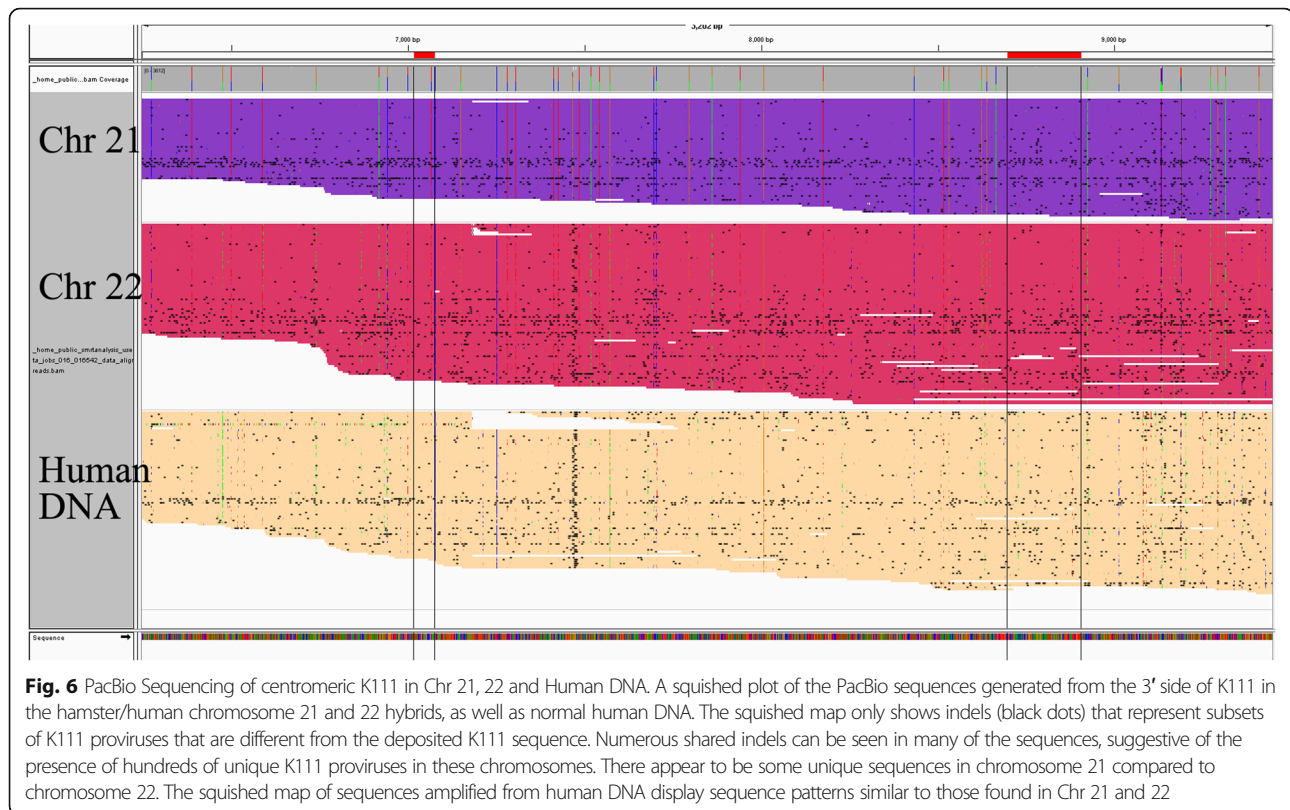
#### Genetic variability of K111 in CTCL and HIV patients

To determine the genetic variability of K111 sequences in CTCL and HIV patients that have either a  $+/+$  or  $-/-$  K111 genotype, we performed PacBio Next-Generation-Sequencing of K111 PCR products generated with primer P2R that binds the 3' integration site of K111 CER:D22Z3, and primer 6353F that binds the *env* region of K111. We also amplified another segment using primer P1 that binds the 5' CER:D22Z3, and primer 2499R that binds downstream of the origin of K111 sequence, in the *pro* region (Fig. 7). Six bp barcodes were added to the 5' end of 6353F and to the 3' end of 2499R in order to distinguish between the sequence results from different patients.

A squished plot alignment of the K111 sequences amplified, displaying only indels, revealed a large degree of sequence variability of K111 proviruses amplified in each human DNA sample (Additional file 3: Figure S1). In patients with a  $+/+$  K111 genotype, about 200 K111 proviruses with unique sequences were detected. DNA sequencing also confirmed the lack of K111 proviral sequences on the 5' side of the K111 provirus in  $-/-$  K111 samples. In these patients, we were able to amplify a fragment of the 3' portion of the centromere provirus, which may be related to the pericentromeric K222 proviruses, or recombinant K111/K222 proviruses, as previously reported [25]. Although, as expected, we did not obtain DNA sequences at the 5' site of the K111 sequence in  $-/-$  K111 samples, we did recover non-specific amplification products of the HK2 provirus 3p25.3 (Additional file 3: Figure S1), validating the observations shown in Fig. 1.

Looking at the sequences of 3p25.3 from 4 CTCL individuals generated by our sequencing technology, we found that only two copies of this provirus existed in





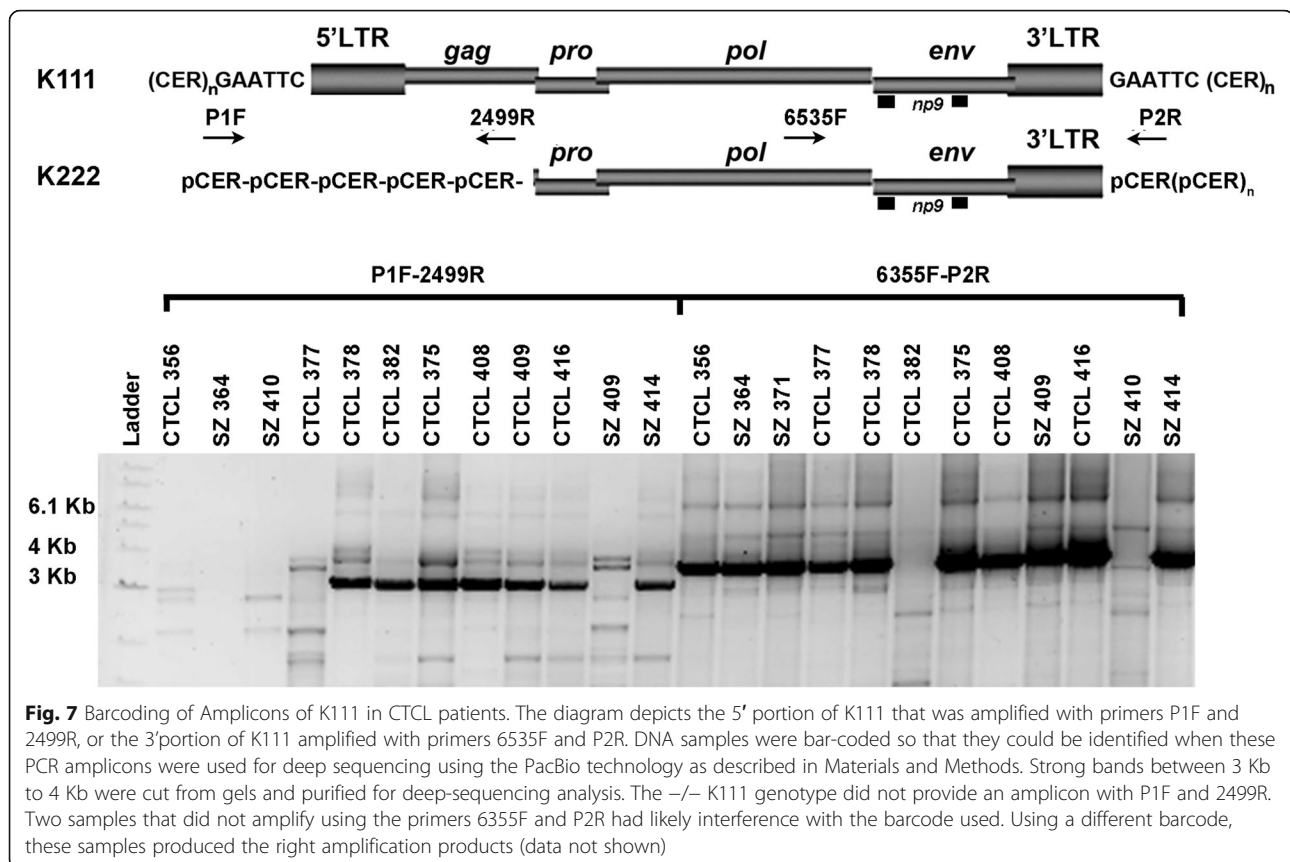
these individuals per diploid genome, validating the fidelity of the PacBio sequencing. In contrast to K111 sequences that have many indels, we did not find indels in HK2 3p25.3 sequences, but only sporadic mutations that appear to be SNPs representative of each individual (Additional file 4: Figure S2). This supports the idea that the multiple mutations in K111s were not due to PCR and sequencing artifact, but rather were due to the evolution and spread of K111 throughout the pericentromeric area by a process of homologous recombination [24–26]. DNA sequencing readily validated the existence of  $-/-$  and  $+/+$  K111 genotypes.

## Discussion

In this study we show that approximately 15% of Caucasian people lack K111 proviruses in their pericentromeric/centromeric region. In contrast, the  $-/-$ K111 genotype was rarely seen in individuals from Nigeria and was not detected in any other sub-Saharan Africans or Pygmy populations, suggesting that the  $-/-$ K111 genotype appeared after ancient humans migrated out of Africa [46]. The finding of a  $-/-$ K111 genotype in 7.6% of African Americans probably arose through the forcible translocation of Africans through the slave trade and then subsequent intermingling with Caucasian and Amerindian populations over many generations. The questions arise: Where did the  $-/-$ K111 genotype

originate, and how can 15% of Caucasians lack several hundreds to thousands of copies of these pericentromeric proviruses while almost, if not all, African individuals have K111 proviruses?

In a retrospective study of DNA samples obtained from Neanderthal and Denisovan, late Pleistocene Hominins that moved out of Africa between 50,000 to 400,000 years ago, we found a few copies of K111 in their genomes, suggesting that they already carried K111 out of Africa [25, 28–31]. Thus, the  $-/-$ K111 genotype we have detected in modern humans does not appear to come from Neanderthals or Denisovan. It is possible that the  $-/-$ K111 genotype was transmitted to early modern humans out of Africa from other primitive hominins that lacked K111. However, it is more likely that some early human who had a  $+/+$ K111 genotype in their genome at some point in evolutionary time may have lost the 5' end of K111 through recombination with K222, a centromere provirus that also lacks the exact 5' end [26]. We have found evidence of recombinant K111/K222 provirus sequences in human genomes using both PCR-sequencing and Pac Bio Next-Generation-Sequencing studies, suggesting the possibility that K222 recombination with K111 can delete the 5' domain of K111. These individuals could have evolved into carrying the  $-/-$ K111 to different parts of the world. The evidence suggests that this event did not take place



in Africa. The worldwide distribution of the  $-/-$ K111 is of great interest, and it is possible that in some isolated populations we might find many individuals that exclusively have the  $-/-$ K111 genotype.

It is of great interest that so many K111 proviruses exist in the centromeric and pericentromeric area of the human genome. These K111s significantly outnumber the proviruses inserted into non-centromeric regions. They appear to have expanded to such numbers by homologous recombination, a mechanism that created many indels in K111 proviruses. In addition, there appear to be uniquely different K111s in chromosome 21 and 22 which have the bulk of the K111s in our genome. How these proviral differences play a role in disorders of chromosome 21 such as trisomy 21 and/or disorders of chromosome 22 is now under investigation. In the genetic disorder trisomy 21, we have shown that K111 is reduced, with some evidence of further recombination events occurring in these patients' cells [47].

It should be pointed out that a large number of centromeric proviruses are not unique to man. In fact, kangaroo species have several endogenous retroviruses (KERV) in centromeres found in tandem, and are major integral parts of active centromeres similar to what is seen with human K111s [48].

We asked the question about diseases that could be associated with the  $-/-$ K111 genotype, which produces *pericentromeric instability*. Our studies show no differences in the prevalence of the  $-/-$ K111 genotype in Caucasian patients with psoriasis, lupus, Takayasu arteritis, breast cancer and/or lymphomas. However, the prevalence of the  $-/-$ K111 genotype was significantly higher in Caucasian individuals with CTCL. These CTCL patients with a  $-/-$ K111 genotype were more severely affected by complications like Sz and LCT and more aggressive forms of disease requiring extensive treatment to control the disease. It is possible then that *pericentromeric instability* in Caucasians in some way contributes to pathogenesis of CTCL in these patients, and that missing pericentromeric K111 makes the disease more severe and difficult to control. Therefore, the  $-/-$ K111 genotype appears to be an independent risk factor in determining treatment outcome for patients with CTCL. This could be assessed by a prospective study of patients with CTCL in order to understand whether there are outcome differences in disease in  $-/-$ K111 vs.  $+/+$ K111 patients. Although African and/or African-American patients can develop severe CTCL disease, as do Caucasians, a  $-/-$ K111 genotype is less prevalent. We have insufficient data to understand the role of  $-/-$ K111 in African American

patients. Only 4 patients in our cohort were African-Americans and 2 had severe disease; both had a +/+K111 genotype. When considering the role of centromeric K111s in human disease, as we have shown, these proviruses may serve as a marker of the health of the pericentromeric genome [47].

It has been reported that in patients with CTCL there is overexpression of HERV-K and HERV-W mRNA compared to healthy individuals, but there was no difference in expression levels in more advanced disease [49]. We did not assay the extent of expression of K111 s' putative gene *np9*, or spliced variants of this gene [32, 33], and/or other genes of K111, but the role of expression of these genes in CTCL is of great interest for future studies.

In some cases, patients with severe CTCL may require bone marrow transplantation. It is possible that donors with a +/+K111 genotype may be more beneficial than a -/-K111 donor. This could be ascertained by retrospective studies of outcome in such transplanted patients where -/-K111 genotyping of donors could be performed. One of our -/-K111 patients received a +/+K111 genotype donor marrow. She had engraftment with good control of her disease. Unfortunately, she subsequently died of cardiac complications from earlier chemotherapy. Screening donors for a +/+K111 genotype might result in better control of disease in these sick patients.

As a group, Caucasian HIV patients have a 15.3% prevalence of the -/-K111 genotype. However, there is a statistically significant higher prevalence of -/-K111 genotype in Caucasian/Hispanic patients who are LTNPs as compared to Rapid Progressors. It is conceivable that the numbers of centromeric K111 copies in the genome increases the risk factor of AIDS development, while missing centromeric K111 can add to better control of HIV. The low prevalence of -/-K111 in one population vs. high prevalence of -/-K111 in another population might also contribute to the differences in rates of spread of HIV in the world with -/-K111 acting as limiting factor in transmission. Larger studies will be necessary to confirm these results.

We should note that K111 was first discovered in the blood of HIV infected individuals, the only type of individuals in whose blood we have found RNA transcripts of these centromeric proviruses [24, 25]. K111 RNA was not detected in the blood of healthy individuals or patients with breast cancer or several types of lymphomas. We previously observed that HIV upregulates K111 via Tat protein transactivation of the K111 promoter. In addition, Tat was found to relax the heterochromatin at centromeric areas, allowing for transcriptional activation of K111 sequences [50, 51]. It would be interesting to study whether the transcriptional activation of K111

affects HIV pathology and disease progression, and whether not having K111 negatively affects the course of HIV infection. It is of interest that the -/-K111 cell lines Hut 78 and H9, which have the CD4 receptor for HIV, support the growth of HIV, producing high titers of virus without going through cell lysis. These cell lines originated from a patient with CTCL and produced sufficient virus to make a successful HIV test for ELISA and Western Blotting [37, 52]. These cell lines do not need to be replenished with feeder cells, but continuously produce significant titers of HIV and do not undergo cell death. How these cells resist cell lysis and death remains unexplained, but perhaps the -/-K111 genotype plays a role in preventing cellular death.

This study adds to our understanding of how HK2 viruses may play a role in control of HIV infection. It has been shown that HK2 *gag* interacts with HIV *gag* reducing HIV-1 release and infectivity, thus interfering with HIV replication [53, 54]. In addition, HIV-1 induces aberrant expression of HK2 Env [55]. This allows human T-cell clones to HK2 and human anti HK2 Env antibodies to eliminate HIV infected cells in vitro [56–59]. In addition, elite controllers were shown to have strong HK2 antibody as well as cellular responses to HK2 providing protection against aggressive viral replication [55].

Finally, K111s are type 1 proviruses that have the ability to make functional Np9 proteins, similar to other HK2 Type 1 proviruses. Recent evidence has shown that K111 can also produce alternative spliced variants of *np9* [32]. The function of Np9 has begun to be elucidated, but the function of Np9 variants is not yet known. Therefore, K111 encoded Np9 protein variants may play some important role in T-cell biology and HIV infection. In the case of CTCL, the absence of K111 *np9* variants may contribute to failure of cells to die and thus to become neoplastic, while in HIV this may provide protection against T-cell lysis from HIV and result in better preservation of T-cells as seen in HIV LTNPs. Clearly, modern humans have hundreds to thousands of copies of K111, each one with the potential of generating an Np9 protein or variants of Np9. Greater understanding of these proteins may be very important for delineating the function of the T-cell in CTCL and/or in HIV infection, leading perhaps to better therapies for T-cell disorders.

## Conclusions

Our data indicate that HK2 proviruses in the pericentromeres of several chromosomes, or at least the 5' portion of their genomes, are deleted in a subset of modern humans. Considering that ~1000 of these retroelements exist in the pericentromeres, having a deletion in the 5' portion of these retroelements indicates that 3400 Kb of pericentromeric sequence is

missing. Further, it appears that *pericentromeric instability* is associated with survival of a subset of CD4+ T-cells, which together contribute to the development of neoplastic transformation with concomitant severe CTCL. In another unknown way, pericentromeric instability allows a different subset of CD4+ T-cells to survive better in HIV infection resulting in long term survival.

## Additional files

**Additional file 1: Table S1.** Clinical Features of Cutaneous T-cell Lymphoma patients: age, gender, mortality, diagnosis, disease transformation, and the development of secondary cancers in patients with CTCL (DOCX 16 kb)

**Additional file 2: Table S2.** Treatment of Patients with Cutaneous T-Cell Lymphoma (DOCX 14 kb)

**Additional file 3: Figure S1.** Squished plot of K111 PacBio sequences from patients with CTCL. The squished maps show several K111 sequences seen among the patients, with as many as 200 different genetic patterns visible. Shown are sequences amplified in individuals with a  $-/-$ K111 or  $+/+$ K111 genotypes (see Y axis). Each color plot represents an individual patient. Differences in nucleotide sequences are indicated by black dots. These black dots only represent indels. Single nucleotide polymorphisms (SNPs) are not displayed for clarity. Note that in  $-/-$ K111 patients a cluster of sequences on the 5' site, but not the 3' site, show the HERV-K HML-2 provirus 3p25.3. Black dots in 3p25.3 sequences only indicate indels relative to the 5' K111 sequence. No particular pattern of K111 could be recognized that distinguished CTCL patients from Sézary patients other than the absence of K111 sequences on the 5' side of K111 (PDF 6634 kb)

**Additional file 4: Figure S2.** Highlight plot of HERV-K HML-2 3p25.3 sequences. The highlight plot displays the differences in nucleotide sequence of the provirus 3p25.3 amplified by PCR and sequenced using the PacBio platform. Sequences are compared to provirus 3p25.3 (Acc No. JN675020.1), the master sequence. The plot indicates nucleotide substitutions T (red ticks), A (green ticks), C (blue ticks), and G (yellow ticks). A neighbor joining phylogenetic tree is indicated at the right side. The highlight plot was generated using Highlighter from Los Alamos HIV Sequence Database [https://www.hiv.lanl.gov/content/sequence/HIGH-LIGHT/highlighter\\_top.html](https://www.hiv.lanl.gov/content/sequence/HIGH-LIGHT/highlighter_top.html). This indicates that variation in fixed endogenous viral sequence occurs mostly by spontaneous mutations over time rather than homologous recombination which is so characteristic of the enormous sequence variation seen in K111 (PDF 3482 kb)

## Abbreviations

CLL: Chronic lymphocytic leukemia; CTCL: Cutaneous T-cell Lymphoma; HK2: Human Endogenous Retroviruses type K HML-2; LCT: Large cell transformation; LNX: Ligand of numb protein X; LTNP: Long-Term Non-Progressors; LTR: Long Terminal Repeat; MACS: Multicenter AIDS Cohort Study; PBLs: Peripheral blood lymphocytes; PBMCs: Peripheral blood mononuclear cells; PLZF: Promyelocytic leukemia zinc finger protein; Sz: Sézary syndrome; TCR: T-cell receptor

## Acknowledgments

The cell lines Hut 78, H9, and H9/HTLVIII were provided by the AIDS Research and Reference Reagent Program. DNA from patients with Takayasu arteritis and Lupus erythematosus were generously provided by Dr. Amr H. Sawalha. Drs. James T. Elder and Trilokraj Tejasvi generously provided DNA from patients with psoriasis. Thanks to Carol Dosik Kaplan for editorial assistance.

## Funding

Data in this manuscript were collected by the Multicenter AIDS Cohort Study (MACS). MACS Principal Investigators: Johns Hopkins University Bloomberg School of Public Health (Joseph Margolick, Todd Brown), grant U01-AI35042;

Northwestern University (Steven Wolinsky), grant U01-AI35039; University of California, Los Angeles (Roger Detels, Otoniel Martinez-Maza, Otto Yang), grant U01-AI35040; University of Pittsburgh (Charles Rinaldo, Lawrence Kingsley, Jeremy Martinson), grant U01-AI35041; the Center for Analysis and Management of MACS, Johns Hopkins University Bloomberg School of Public Health (Lisa Jacobson, Gypsyamber D'Souza), grant UM1-AI35043. The MACS is funded primarily by the National Institute of Allergy and Infectious Diseases (NIAID), with additional co-funding from the National Cancer Institute (NCI), the National Institute on Drug Abuse (NIDA), and the National Institute of Mental Health (NIMH). Targeted supplemental funding for specific projects was also provided by the National Heart, Lung, and Blood Institute (NHLBI), and the National Institute on Deafness and Communication Disorders (NIDCD). MACS data collection is also supported by grant UL1-TR001079 (JHU ICTR) from the National Center for Advancing Translational Sciences (NCATS), a component of the National Institutes of Health (NIH) and the NIH Roadmap for Medical Research. The contents of this publication are solely the responsibility of the authors and do not represent the official views of the National Institutes of Health (NIH), Johns Hopkins ICTR, or NCATS. The MACS website is located at <http://aidscohortstudy.org/>. C.A.A. and S.N.A. were supported by the Fogarty International Center of the NIH; grant number D43TW009106; the National Human Genome Research Institute, Grant number U01HG009784; and the National Cancer Institute, grant number 5P30CA134274. This work was supported by a grant 05-5089 from the Concerned Parents for AIDS Research (CPFA) to M.H.K.; grants 1R01AI110259-01A1 and 1I01BX002358-01A1 to M.S.; and a grant K22 CA177824 from the National Cancer Institute to R.C.G. R.C.G. is the recipient of a New Investigator Award from the Scleroderma Foundation.

## Availability of data and materials

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

## Authors' contributions

MHK and RCG designed experiments; MHK, JZ, and RCG performed experiments; MHK, MK, JME, SDG, JTE, TT, EG, AHS, JPM, MHD, HD, GSD, SNA, CAA, and MS obtained patient's samples, acquired patient's information and draft the manuscript. MHK, JZ and RCG analyzed the data; MHK, MK, JZ, MS, and RCG wrote the manuscript; MHK and RCG oversaw the work. All authors read and approved the final manuscript.

## Ethics approval and consent to participate

All patient samples were acquired with written informed consent and with the approval of the respective Institutional Review Boards. Review boards include the Institutional Review Board of the University of Michigan Medical School (IRBMED), The University of Maryland Institutional Review Board, The MACS Institutional Review board (CAMACS), The Veterans Affairs Central Institutional Review Board, and the Northshore University Health System Institutional Review Board.

## Consent for publication

Not applicable.

## Competing interests

The authors declare that they have no competing interests.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## Author details

<sup>1</sup>Division of Infectious Diseases, Department of Internal Medicine, University of Michigan, Ann Arbor, MI 48109, USA. <sup>2</sup>Division of Hematology/Oncology, Department of Internal Medicine, University of Michigan, Ann Arbor, MI 48109, USA. <sup>3</sup>Division of Dermatology, Department of Internal Medicine, University of Michigan, Ann Arbor, MI 48109, USA. <sup>4</sup>Division of Rheumatology, Department of Internal Medicine, University of Michigan, Ann Arbor, MI 48109, USA. <sup>5</sup>Division of Infectious Diseases, The Feinstein Institute for Medical research, Manhasset, NY 11030, USA. <sup>6</sup>Hematology Oncology Associates, Setauket, NY 11776, USA. <sup>7</sup>Division of Rheumatology, School of Medicine, Marmara University, Istanbul, Turkey. <sup>8</sup>Department of Physiology, Istanbul Faculty of Medicine, Istanbul University, Istanbul, Turkey.



<sup>9</sup>Department of Epidemiology and Public Health, University of Maryland School of Medicine, Baltimore, MD 21201, USA. <sup>10</sup>Greenebaum Comprehensive Cancer Center, University of Maryland School of Medicine, Baltimore, MD 21201, USA. <sup>11</sup>Institute of Human Virology, University of Maryland School of Medicine, Baltimore, MD 21201, USA. <sup>12</sup>Ann Arbor Veterans Affairs Hospital, Ann Arbor, MI 48105, USA. <sup>13</sup>Hormel Institute, University of Minnesota, Austin, MN 55912, USA.

Received: 1 November 2018 Accepted: 15 April 2019

Published online: 02 May 2019

## References

- Lower J, Lower R, Stegmann J, Frank H, Kurth R. Retrovirus particle production in three of four human teratocarcinoma cell lines. *Haematol Blood Transfus.* 1981;26:541–4.
- Ono M. Molecular cloning and long terminal repeat sequences of human endogenous retrovirus genes related to types A and B retrovirus genes. *J Virol.* 1986;58(3):937–44.
- Lower R, Lower 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):5177–84.
- Medstrand P, Mager DL. Human-specific integrations of the HERV-K endogenous retrovirus family. *J Virol.* 1998;72(12):9782–7.
- Barbulescu M, Turner G, Seaman MI, Deinard AS, Kidd KK, Lenz J. Many human endogenous retrovirus K (HERV-K) proviruses are unique to humans. *Curr Biol.* 1999;9(16):861–8.
- Tonjes RR, Czauderna F, Kurth R. Genome-wide screening, cloning, chromosomal assignment, and expression of full-length human endogenous retrovirus type K. *J Virol.* 1999;73(11):9187–95.
- Andersson ML, Lindeskog M, Medstrand P, Westley B, May F, Blomberg J. Diversity of human endogenous retrovirus class II-like sequences. *J Gen Virol.* 1999;80 ( Pt 1):255–60.
- Reus K, Mayer J, Sauter M, Scherer D, Muller-Lantzsch N, Meese E. Genomic organization of the human endogenous retrovirus HERV-K (HML-2.HOM) (ERVk6) on chromosome 7. *Genomics.* 2001;72(3):314–20.
- Reus K, Mayer J, Sauter M, Zischler H, Muller-Lantzsch N, Meese E. HERV-K (OLD): ancestor sequences of the human endogenous retrovirus family HERV-K (HML-2). *J Virol.* 2001;75(19):8917–26.
- Turner G, Barbulescu M, Su M, Jensen-Seaman MI, Kidd KK, Lenz J. Insertional polymorphisms of full-length endogenous retroviruses in humans. *Curr Biol.* 2001;11(19):1531–5.
- Buzdin A, Ustyugova S, Khodosevich K, Mamedov I, Lebedev Y, Hunsmann G, et al. Human-specific subfamilies of HERV-K (HML-2) long terminal repeats: three master genes were active simultaneously during branching of hominoid lineages. *Genomics.* 2003;81(2):149–56.
- Hughes JF, Coffin JM. Human endogenous retrovirus K solo-LTR formation and insertional polymorphisms: implications for human and viral evolution. *Proc Natl Acad Sci U S A.* 2004;101(6):1668–72.
- Belshaw R, Dawson AL, Woolven-Allen J, Redding J, Burt A, Tristem M. Genomewide screening reveals high levels of insertional polymorphism in the human endogenous retrovirus family HERV-K (HML2): implications for present-day activity. *J Virol.* 2005;79(19):12507–14.
- Bannert N, Kurth R. The evolutionary dynamics of human endogenous retroviral families. *Annu Rev Genomics Hum Genet.* 2006;7:149–73.
- Subramanian RP, Wildschutte JH, Russo C, Coffin JM. Identification, characterization, and comparative genomic distribution of the HERV-K (HML-2) group of human endogenous retroviruses. *Retrovirology.* 2011; 8:90.
- Macfarlane C, Simmonds P. Allelic variation of HERV-K (HML-2) endogenous retroviral elements in human populations. *J Mol Evol.* 2004;59(5):642–56.
- Macfarlane CM, Badge RM. Genome-wide amplification of proviral sequences reveals new polymorphic HERV-K (HML-2) proviruses in humans and chimpanzees that are absent from genome assemblies. *Retrovirology.* 2015;12:35.
- Lower R, Boller K, Hasenmaier B, Korbmaier C, Muller-Lantzsch N, Lower J, et al. Identification of human endogenous retroviruses with complex mRNA expression and particle formation. *Proc Natl Acad Sci U S A.* 1993;90(10): 4480–4.
- Marchi E, Kanapin A, Magiorkinis G, Belshaw R. Unfixed endogenous retroviral insertions in the human population. *J Virol.* 2014;88(17):9529–37.
- Wildschutte JH, Williams ZH, Montesion M, Subramanian RP, Kidd JM, Coffin JM. Discovery of unfixed endogenous retrovirus insertions in diverse human populations. *Proc Natl Acad Sci U S A.* 2016;113(16):E2326–34.
- Boller K, Schonfeld K, Lischer S, Fischer N, Hoffmann A, Kurth R, et al. Human endogenous retrovirus HERV-K113 is capable of producing intact viral particles. *J Gen Virol.* 2008;89(Pt 2):567–72.
- Beimforde N, Hanke K, Ammar I, Kurth R, Bannert N. Molecular cloning and functional characterization of the human endogenous retrovirus K113. *Virology.* 2008;371(1):216–25.
- Jha AR, Pillai SK, York VA, Sharp ER, Storm EC, Wachter DJ, et al. Cross-sectional dating of novel haplotypes of HERV-K 113 and HERV-K 115 indicate these proviruses originated in Africa before *Homo sapiens*. *Mol Biol Evol.* 2009;26(11):2617–26.
- Contreras-Galindo R, Kaplan MH, Contreras-Galindo AC, Gonzalez-Hernandez MJ, Ferlenghi I, Giusti F, et al. Characterization of human endogenous retroviral elements in the blood of HIV-1-infected individuals. *J Virol.* 2012; 86(1):262–76.
- Contreras-Galindo R, Kaplan MH, He S, Contreras-Galindo AC, Gonzalez-Hernandez MJ, Kappes F, et al. HIV infection reveals widespread expansion of novel centromeric human endogenous retroviruses. *Genome Res.* 2013; 23(9):1505–13.
- Zahn J, Kaplan MH, Fischer S, Dai M, Meng F, Saha AK, et al. Expansion of a novel endogenous retrovirus throughout the pericentromeres of modern humans. *Genome Biol.* 2015;16:74.
- Metzdorf R, Gottert E, Blin N. A novel centromeric repetitive DNA from human chromosome 22. *Chromosoma.* 1988;97(2):154–8.
- Meyer M, Kircher M, Gansauge MT, Li H, Racimo F, Mallick S, et al. A high-coverage genome sequence from an archaic Denisovan individual. *Science.* 2012;338(6104):222–6.
- Green RE, Krause J, Briggs AW, Maricic T, Stenzel U, Kircher M, et al. A draft sequence of the Neandertal genome. *Science.* 2010;328(5979):710–22.
- Noonan JP, Coop G, Kudravalli S, Smith D, Krause J, Alessi J, et al. Sequencing and analysis of Neanderthal genomic DNA. *Science.* 2006; 314(5802):1113–8.
- Reich D, Green RE, Kircher M, Krause J, Patterson N, Durand EY, et al. Genetic history of an archaic hominin group from Denisova cave in Siberia. *Nature.* 2010;468(7327):1053–60.
- Schmitt K, Heyne K, Roemer K, Meese E, Mayer J. HERV-K (HML-2) rec and np9 transcripts not restricted to disease but present in many normal human tissues. *Mob DNA.* 2015;6:4.
- Heyne K, Kolsch K, Bruand M, Kremmer E, Grasser FA, Mayer J, et al. Np9, a cellular protein of retroviral ancestry restricted to human, chimpanzee and gorilla, binds and regulates ubiquitin ligase MDM2. *Cell Cycle.* 2015;14(16): 2619–33.
- Chen T, Meng Z, Gan Y, Wang X, Xu F, Gu Y, et al. The viral oncogene Np9 acts as a critical molecular switch for co-activating beta-catenin, ERK, Akt and Notch1 and promoting the growth of human leukemia stem/progenitor cells. *Leukemia.* 2013;27(7):1469–78.
- Denne M, Sauter M, Armbruester V, Licht JD, Roemer K, Mueller-Lantzsch N. Physical and functional interactions of human endogenous retrovirus proteins Np9 and rec with the promyelocytic leukemia zinc finger protein. *J Virol.* 2007;81(11):5607–16.
- Armbruester V, Sauter M, Roemer K, Best B, Hahn S, Nty A, et al. Np9 protein of human endogenous retrovirus K interacts with ligand of numb protein X. *J Virol.* 2004;78(19):10310–9.
- Schupbach J, Popovic M, Gilden RV, Gonda MA, Sarngadharan MG, Gallo RC. Serological analysis of a subgroup of human T-lymphotropic retroviruses (HTLV-III) associated with AIDS. *Science.* 1984;224(4648):503–5.
- Jawed SI, Myskowski PL, Horwitz S, Moskowitz A, Querfeld C. Primary cutaneous T-cell lymphoma (mycosis fungoides and Sezary syndrome): part I. diagnosis: clinical and histopathologic features and new molecular and biologic markers. *J Am Acad Dermatol.* 2014;70(2):205 e1–16 quiz 21–2.
- Olsen E, Vonderheid E, Pimpinelli N, Willemze R, Kim Y, Knobler R, et al. Revisions to the staging and classification of mycosis fungoides and Sezary syndrome: a proposal of the International Society for Cutaneous Lymphomas (ISCL) and the cutaneous lymphoma task force of the European Organization of Research and Treatment of Cancer (EORTC). *Blood.* 2007;110(6):1713–22.
- Olsen EA, Rook AH, Zic J, Kim Y, Porcu P, Querfeld C, et al. Sezary syndrome: immunopathogenesis, literature review of therapeutic options, and

- recommendations for therapy by the United States cutaneous lymphoma consortium (USCLC). *J Am Acad Dermatol.* 2011;64(2):352–404.
41. Olsen EA, Whittaker S, Kim YH, Duvic M, Prince HM, Lessin SR, et al. Clinical end points and response criteria in mycosis fungoides and Sezary syndrome: a consensus statement of the International Society for Cutaneous Lymphomas, the United States cutaneous lymphoma consortium, and the cutaneous lymphoma task force of the European Organisation for Research and Treatment of Cancer. *J Clin Oncol.* 2011; 29(18):2598–607.
  42. Batrani M, Bhawan J. Pitfalls in the diagnosis of cutaneous lymphoma. *Am J Dermatopathol.* 2014;36(1):90–100.
  43. Willemze R, Jaffe ES, Burg G, Cerroni L, Berti E, Swerdlow SH, et al. WHO-EORTC classification for cutaneous lymphomas. *Blood.* 2005;105(10):3768–85.
  44. Arulogun SO, Prince HM, Ng J, Lade S, Ryan GF, Blewitt O, et al. Long-term outcomes of patients with advanced-stage cutaneous T-cell lymphoma and large cell transformation. *Blood.* 2008;112(8):3082–7.
  45. Kadin ME, Hughey LC, Wood GS. Large-cell transformation of mycosis fungoides-differential diagnosis with implications for clinical management: a consensus statement of the US cutaneous lymphoma consortium. *J Am Acad Dermatol.* 2014;70(2):374–6.
  46. Pagani L, Lawson DJ, Jagoda E, Morseburg A, Eriksson A, Mitt M, et al. Genomic analyses inform on migration events during the peopling of Eurasia. *Nature.* 2016;538(7624):238–42.
  47. Contreras-Galindo R, Fischer S, Saha AK, Lundy JD, Cervantes PW, Mourad M, et al. Rapid molecular assays to study human centromere genomics. *Genome Res.* 2017;27(12):2040–9.
  48. Ferreri GC, Brown JD, Obergefell C, Jue N, Finn CE, O'Neill MJ, et al. Recent amplification of the kangaroo endogenous retrovirus, KERV, limited to the centromere. *J Virol.* 2011;85(10):4761–71.
  49. 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(1):38–43.
  50. Gonzalez-Hernandez MJ, Cavalcoli JD, Sartor MA, Contreras-Galindo R, Meng F, Dai M, et al. Regulation of the human endogenous retrovirus K (HML-2) transcriptome by the HIV-1 tat protein. *J Virol.* 2014;88(16):8924–35.
  51. Gonzalez-Hernandez MJ, Swanson MD, Contreras-Galindo R, Cookinham S, King SR, Noel RJ Jr, et al. Expression of human endogenous retrovirus type K (HML-2) is activated by the tat protein of HIV-1. *J Virol.* 2012;86(15):7790–805.
  52. Gazdar AF, Carney DN, Bunn PA, Russell EK, Jaffe ES, Schechter GP, et al. Mitogen requirements for the in vitro propagation of cutaneous T-cell lymphomas. *Blood.* 1980;55(3):409–17.
  53. Monde K, Contreras-Galindo R, Kaplan MH, Markovitz DM, Ono A. Human endogenous retrovirus K gag coassembles with HIV-1 gag and reduces the release efficiency and infectivity of HIV-1. *J Virol.* 2012;86(20):11194–208.
  54. Monde K, Terasawa H, Nakano Y, Soheilian F, Nagashima K, Maeda Y, et al. Molecular mechanisms by which HERV-K gag interferes with HIV-1 gag assembly and particle infectivity. *Retrovirology.* 2017;14(1):27.
  55. de Mulder M, SenGupta D, Deeks SG, Martin JN, Pilcher CD, Hecht FM, et al. Anti-HERV-K (HML-2) capsid antibody responses in HIV elite controllers. *Retrovirology.* 2017;14(1):41.
  56. Garrison KE, Jones RB, Meiklejohn DA, Anwar N, Ndhlovu LC, Chapman JM, et al. T cell responses to human endogenous retroviruses in HIV-1 infection. *PLoS Pathog.* 2007;3(11):e165.
  57. Michaud HA, de Mulder M, SenGupta D, Deeks SG, Martin JN, Pilcher CD, et al. Trans-activation, post-transcriptional maturation, and induction of antibodies to HERV-K (HML-2) envelope transmembrane protein in HIV-1 infection. *Retrovirology.* 2014;11:10.
  58. Michaud HA, SenGupta D, de Mulder M, Deeks SG, Martin JN, Kobie JJ, et al. Cutting edge: an antibody recognizing ancestral endogenous virus glycoproteins mediates antibody-dependent cellular cytotoxicity on HIV-1-infected cells. *J Immunol.* 2014;193(4):1544–8.
  59. SenGupta D, Tandon R, Vieira RG, Ndhlovu LC, Lown-Hecht R, Ormsby CE, et al. Strong human endogenous retrovirus-specific T cell responses are associated with control of HIV-1 in chronic infection. *J Virol.* 2011;85(14): 6977–85.

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

**At BMC, research is always in progress.**

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

