A viral clonality evenness score to predict progression to adult T-cell leukaemia in asymptomatic carriers of human T-lymphotropic virus type 1 in Japan: a retrospective longitudinal cohort study

Snehal Dilip Karpe, Maria Artesi, Jérôme Wayet, Keith Durkin, Vincent Hahaut, Kaoru Uchimaru, Junya Makiyama, Masako Iwanaga, Atae Utsunomiya, P Martijn Kolijn, Anton W Langerak, Arsène Burny, Ambroise Marçais, Olivier Hermine, Toshiki Watanabe, Michel Georges, Anne Van den Broeke

Summary

Background Adult T-cell leukaemia/lymphoma (ATL) is a highly aggressive T-cell malignancy that occurs in approximately 2–7% of individuals with human T-lymphotropic virus type 1 (HTLV-1), after decades of asymptomatic infection. To address the urgent need for predictive biomarkers to identify asymptomatic carriers of HTLV-1 at high risk of progression to ATL, we aimed to evaluate viral clonality sequencing as a potential tool for risk stratification.

Methods This retrospective longitudinal cohort study involved HTLV-1 carriers enrolled in the Joint Study on Predisposing Factors of ATL Development, a nationwide cohort study initiated in Japan in 2002. Participants were selected from this cohort on the basis of their baseline proviral load at the time of enrolment as an asymptomatic carrier, length of follow-up, and clinical outcome. The cohort was subdivided into three subgroups: the first comprising HTLV-1 carriers who developed ATL, the second comprising carriers with high proviral load (≥4%) who did not progress to ATL, and the third comprising carriers with low proviral load (<4%) who did not progress to ATL. DNA extracted from peripheral blood mononuclear cells collected at enrolment and at least one follow-up visit was analysed by HTLV-1 clonality sequencing and the proviral load was quantified. We calculated a viral clonality evenness (VCE) score, based on the Shannon Evenness Index, to quantify the uniformity of the clonal distribution of samples, for which 0 represents a perfectly monoclonal architecture and 1 indicates a completely polyclonal landscape. We then estimated the performance of proviral load thresholds and VCE scoring to classify the risk of progression to ATL using the area under the receiver operating characteristic curve (AUC), the accuracy, and Matthews correlation coefficient. VCEs were compared between participant subgroups with the Wilcoxon rank sum test.

Findings 56 participants followed up by JSPFAD between Feb 6, 2003, and July 19, 2022, were included in this study: 17 who progressed to ATL (mean follow-up 8·3 years [SD 4·0]), 18 who had a high proviral load and did not progress to ATL (9·7 years [3·4]), and 21 who had a low proviral load and did not progress to ATL (7·5 years [3·0]). Clonality sequencing of samples from 39 participants who did not progress to ATL revealed hundreds to thousands of HTLV-1 integration sites at both timepoints, corresponding to multiple clones of low and uniform abundance, and these participants had high VCE scores (≥0·694) at baseline. By contrast, most participants (14 of 17) who progressed to ATL had a single predominant clone or two to four predominant clones at both timepoints, and lower VCE scores (<0·694) at baseline than those who did not progress ($p<0·0001$). AUCs were very similar for proviral load thresholds (91 [95% CI 80–98]) and VCE scoring (91 [78–100]), although when using methods that give equal weight to every individual, VCE scoring outperformed proviral load thresholds in predicting progression to ATL (accuracy: proviral load 0·76 [95% CI 0·76–0·77], VCE scoring 1·00 [0·99–1·00]; Matthews correlation coefficient: proviral load 0·23 [95% CI 0·19–0·24], VCE scoring 0·91 [0·80–1·00]). Prediction based on VCE scoring indicated no false positives, compared with 20% when using proviral load, although VCE scoring yields a greater number of false negatives (0·3% *vs* 0·1%).

Interpretation The implementation of VCE scoring in clinical practice could inform early pre-emptive therapeutic interventions, exclusively targeting individuals with HTLV-1 at high risk and aiming to prevent progression to aggressive, treatment-refractory disease. Further validation, including independent confirmation of the performance of VCE scoring in multiple populations and the characterisation of its temporal dynamics, will be crucial to determine its clinical utility and potential integration into care pathways.

Funding Association Jules Bordet, FNRS-Télévie, FCC, WALInnov, FLF, JSPS-KAKENHI, and CoBiA.

Copyright © 2025 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Lancet Microbe 2025; 6: 101198

Published Online November 7, 2025
https://doi.org/10.1016/j.lanmic.2025.101198

Laboratory of Viral Oncogenesis, Institut Jules Bordet—HUB, ULB, Brussels, Belgium (S D Karpe PhD, V Hahaut PhD, A Van den Broeke PhD); **Unit of Animal Genomics, GIGA, ULiège, Liège, Belgium** (S D Karpe, M Artesi PhD, J Wayet MSc, K Durkin PhD, V Hahaut, Prof A Burny PhD, Prof M Georges PhD, A Van den Broeke); **Laboratory of Tumor Cell Biology, The University of Tokyo, Tokyo, Japan** (Prof K Uchimaru MD); **Department of Hematology, Sasebo City General Hospital, Nagasaki, Japan** (J Makiyama MD); **Clinical Epidemiology, Nagasaki University Graduate School of Biomedical Sciences, Nagasaki, Japan** (Prof M Iwanaga MD); **Department of Hematology, Imamura General Hospital, Kagoshima, Japan** (A Utsunomiya MD); **Laboratory of Medical Immunology, Department of Immunology, Erasmus MC, Rotterdam, The Netherlands** (P M Kolijn PhD, Prof A W Langerak PhD); **Service d'hématologie Adultes, Hôpital Universitaire Necker, Université René Descartes, Assistance publique hôpitaux de Paris, Paris, France** (A Marçais MD, Prof O Hermine MD); **Laboratoire d'onco-hématologie, Institut Necker-Enfants Malades, INSERM U1151, Université Paris Descartes, Paris, France** (A Marçais); **Institut Imagine, INSERM U1163, CNRS ERL8654, Paris, France** (Prof O Hermine); **Graduate School of Medicine, St. Marianna University,**

Kanagawa, Japan
(Prof T Watanabe MD)

Correspondence to:
Dr Anne Van den Broeke, Laboratory of Viral Oncogenesis, Institut Jules Bordet—HUB, ULB, 1070 Brussels, Belgium
anne.vandenbroeke@hubruxelles.be; **@AnneVdBroe**

Introduction

Adult T-cell leukaemia/lymphoma (ATL) is a highly aggressive T-cell malignancy that typically occurs after decades of asymptomatic chronic infection with human T-lymphotropic virus type 1 (HTLV-1).[1–3] HTLV-1 is present worldwide, with high endemicity in Japan, the Caribbean, South America, west Africa, and central Africa and foci in Australia, Romania, and Iran.[4] Globalisation and the migration of people living with HTLV-1 from highly endemic areas have resulted in an increase of HTLV-1 infection in non-endemic regions of the world. Raising awareness of HTLV-1 internationally is therefore increasingly important.[5] ATL has an extremely poor prognosis, with a median overall survival of 8–10 months for aggressive subtypes.[2,6] The vast majority of ATLs are resistant to chemotherapy, and allogeneic haematopoietic stem-cell transplantation—the only curative treatment—is dependent on the availability of suitable donors and a sufficiently remitted disease.[7,8]

An asymptomatic individual with HTLV-1 carries several thousand distinct T-cell clones, each uniquely identified by the HTLV-1 proviral integration site in the host genome.[9] Progression to ATL is driven by one of these clones undergoing malignant transformation and clonal

Research in context

Evidence before this study

To identify data on predictive biomarkers of the development of adult T-cell leukaemia/lymphoma (ATL) and the risk stratification of asymptomatic carriers of human T-lymphotropic virus type 1 (HTLV-1), we searched PubMed for reports published in English between database inception and Dec 31, 2024, using the search terms "HTLV-1", "Adult T-cell leukemia", "ATL", "asymptomatic carrier", "risk stratification", "PVL", "clonality", and "predictive biomarker". Our search returned 351 results. Proviral load is the current biomarker used to predict the risk of progression to ATL; however, stratifying HTLV-1 carriers on the basis of proviral load results in a high proportion of false positives (80%). Oligoclonality, a hallmark of premalignancy in HTLV-1 carriers, has been suggested as an alternative biomarker, building on early evidence from low-resolution Southern blot analysis. Oligoclonality has since been assessed through higher-resolution approaches in five studies, including linker-mediated PCR with high-throughput sequencing, rapid amplification of integration sites without interference by genomic DNA, flow cytometry, quantification of T-cell-receptor diversity, and detection of driver mutations. These studies suggested that individuals at high risk of progression to ATL can be identified months or years before clinical presentation; however, the methods used are difficult to translate into clinical applications, both technically and in terms of cost, and few patients with ATL had matching retrospective samples before the onset of disease. Furthermore, these previous studies did not report estimates of the performance of these tests.

Added value of this study

We established viral clonality evenness (VCE) scoring, an HTLV-1 clonality metric derived from viral clonality sequencing, and compared its performance with that of proviral load, the current predictive biomarker for progression to ATL. Performance estimates were obtained after adjusting the composition of the study cohort to better reflect the proportions of HTLV-1 asymptomatic carriers with high proviral load and of individuals who progress to ATL in the general population. VCE scoring outperformed proviral load, including in the risk stratification of HTLV-1 carriers with a high proviral load—recategorising 80% of the individuals who were considered to be at high risk of progression according to current predictive approaches (20% of all individuals with HTLV-1). VCE scoring also has the potential to identify carriers who have a low proviral load but are at high risk of progression—a category of individuals that currently receives insufficient attention. Furthermore, VCE scoring is robust at identifying individuals at high risk of progression even in the presence of an ATL clone with multiple proviral insertion sites. Finally, VCE scores could serve as a more reliable diagnostic classifier for borderline cases of ATL than the combination of clinical characteristics, laboratory features, and history of the disease used in the Shimoyama classification of ATL subtype diagnosis, thereby addressing diagnostic uncertainty. The clonality sequencing method uses high multiplexing and commercially available next-generation sequencing adaptors and does not require viable cells and flow cytometry sorting, reducing cost and enhancing the translational value of this approach. Despite the overall small size of the cohort inherent to the disease, to our knowledge our study involves the largest number of patients with ATL for whom multiple matching retrospective samples are available.

Implications of all the available evidence

Annual assessment of VCE scores in asymptomatic carriers of HTLV-1, regardless of their proviral load, has the potential to identify individuals who could benefit from more intensive follow-up. Furthermore, clonality scoring can also provide some reassurance to a considerable proportion of carriers regarding their risk of progression to ATL. This technique could potentially be used as a molecular diagnostic classifier for borderline cases of ATL, addressing some of the limitations of the current Shimoyama classification, which has remained unchanged since 1991. This study is timely: by proposing a new prognostic and diagnostic biomarker, the findings address the urgent need, emphasised by WHO in 2024, for the development of risk-prediction biomarkers to improve the clinical management of HTLV-1 carriers. Fostering international collaboration to enhance the recruitment of HTLV-1 carriers worldwide will enable larger studies to further evaluate the prognostic value of clonality scoring. Although VCE scoring has yet to be implemented in clinical practice, these findings hold the potential to not only improve patient outcomes but also enhance the wellbeing of individuals with HTLV-1 worldwide.

expansion. At the time of diagnosis, all malignant cells carried by a patient with ATL will belong to a single predominant infected T-cell clone, with typically one proviral copy per cell.[9–12] Linker-mediated PCR with high-throughput sequencing clonality methods can be used to map and quantify HTLV-1 proviral integration sites, enabling the evolution of clonal populations within an individual to be monitored over time.[10,13–15]

The lifetime risk of developing ATL is approximately 2–7% among all asymptomatic individuals carrying HTLV-1.[16] However, this risk is predominantly concentrated within the 25% of carriers with a high proviral load (≥4%; ie, ≥4 copies of HTLV-1 per 100 peripheral blood mononuclear cells [PBMCs]),[17–19] in whom it increases to more than 20%. A limitation of stratifying carriers into high-risk versus low-risk groups on the basis of this proviral load threshold is that it could cause unwarranted anxiety and uncertainty for 20% of all individuals with HTLV-1 (approximately 220 000 people in Japan) as a result of false positives, while 0·1% of carriers (approximately 1100 people in Japan) with a low proviral load who will develop ATL might go unnoticed and not receive the same level of monitoring (false negatives).[16] Molecular tools are therefore urgently needed to improve the risk stratification of individuals with a high proviral load and to identify asymptomatic carriers at high risk of progression to ATL—even those with a low proviral load.

Early observations based on low-resolution Southern blot analysis of HTLV-1-infected cells in the blood of carriers suggested that changes in clonality are an early marker of transformation.[20] In the past 5 years, T-cell-receptor sequencing, Tag-based next-generation sequencing (NGS), rapid amplification of integration sites without interference by genomic DNA, and flow cytometry analysis of PBMCs from individuals with HLTV-1 have shown that, although most carriers have a polyclonal integration-site signature with a uniform distribution of infected clones, some individuals have a more oligoclonal landscape, characterised by a clonally expanded cell population that represents the ATL premalignant clone.[21–25] However, these methods are costly, lack sensitivity, or require viable PBMCs and flow cytometry sorting; as a consequence, they are difficult to translate into clinical practice. We postulated that HTLV-1 clonality sequencing could potentially be used to identify the premalignant stage of ATL in carriers by detecting the presence of a predominant ATL precursor clone before the onset of symptoms, thereby providing a predictive biomarker of transformation. To test this theory, we studied longitudinal samples from a cohort of individuals with HTLV-1 enrolled in the Joint Study on Predisposing Factors of ATL Development (JSPFAD), a nationwide cohort study in Japan that began recruiting in 2002.[17] There are currently around 1 100 000 individuals known to be living with HTLV-1 in Japan and, to our knowledge, JSPFAD is the largest biobank of samples from individuals with this virus worldwide, holding longitudinal samples from those who have progressed to ATL and those who have not progressed, all of whom enrolled at the asymptomatic carrier stage. Annual medical assessments and blood collections from participants over many years have created a resource that, to our knowledge, is unparalleled elsewhere.

We aimed to analyse longitudinal samples obtained from asymptomatic carriers of HTLV-1 who subsequently developed ATL, and compare the data with those from groups of carriers with high proviral load and low proviral load who were followed up over an equivalent period of time and did not progress to ATL.

Methods

Study design and participants

This retrospective longitudinal cohort study recruited asymptomatic carriers of HTLV-1 enrolled in JSPFAD, a nationwide cohort study in Japan that began in 2002 and is ongoing.[17] Individuals enrolled in JSPFAD, who voluntarily present to one of 45 affiliated hospitals on an annual basis, were informed about the study via the JSPFAD website or other participants. Study participants were selected on the basis of their baseline proviral load at the time of enrolment as an asymptomatic carrier, length of follow-up, and clinical outcome (appendix 1 p 42). The initial cohort was divided into three subgroups of equal size: group 1 consisted of carriers who progressed to ATL, group 2 comprised carriers with a high proviral load (≥4%) who did not progress to ATL, and group 3 comprised carriers with a low proviral load (<4%) who did not progress to ATL. The final study population comprised participants who were followed by JSPFAD between Feb 6, 2003, and July 19, 2022.

See Online for appendix 1

Participants gave written informed consent, and the study was approved by the Research Ethics Committee of the University of Tokyo under approval numbers 19–304 and 20–369.

Procedures

Participant follow-up

PBMCs were collected from participants during annual follow-up visits, from which DNA was extracted for HTLV-1 clonality sequencing and proviral load quantification. We sequenced at least two longitudinal samples for each participant, obtained at times T1 and T2. The time of analysis of the earliest available sample for which full clinical and biological data were recorded by JSPFAD was defined as T1; at this timepoint, all participants were asymptomatic carriers. T2 was defined as the time at which the second sample was analysed; for participants who progressed to ATL, T2 was the time at which the first sequenced sample with an ATL diagnosis was obtained. For six participants who progressed to ATL, additional samples were sequenced at later timepoints (15 in total among the six participants) to monitor the ATL subtype over time (appendix 1 pp 35–37, appendix 2 tabs 2, 3).

For more on **JSPFAD** see https://htlv1.jp/jspfad/

See Online for appendix 2

Proviral load quantification

Genomic DNA was extracted from PBMCs using QIAamp Blood DNA kit (Qiagen, Tokyo, Japan) and HTLV-1 proviral

load was measured by quantitative PCR using primers specific to HTLV-1 *tax* and human RNase P genes, as previously described.[26] Proviral loads are reported as percentages and represent the number of HTLV-1 copies per 100 PBMCs, assuming one proviral integration per cell.

HTLV-1 clonality sequencing

HTLV-1 integration sites were mapped and quantified as previously described.[10,14,27] In brief, DNA was sheared, the integration sites were linearly amplified with long terminal repeat (LTR)-specific primers and biotinylated deoxythymidine triphosphate, and the resulting double-stranded DNAs were ligated to Y adaptors (including a unique molecular identifier) and captured with streptavidin-coated magnetic beads. After PCR amplification and purification of the products, a second round of PCR was conducted to add commercially available Illumina Nextera indexes. 150-bp paired-end sequencing reads were acquired on an Illumina MiSeq instrument (Illumina, San Diego, CA, USA; appendix 1 pp 31, 42). Reads that supported either 5′ or 3′ LTR-host junctions were retained and analysed using the clonality analysis pipeline available on GitHub (appendix 1 p 42). HTLV-1 insertion sites were identified, clone abundances quantified, and clonality data plotted as pie charts.

For the **clonality analysis pipeline** see https://github.com/GIGA-AnimalGenomics-BLV/Public/tree/master/PIC

Statistical analysis

Viral clonality evenness as a measure of HTLV-1 clonality

We calculated a viral clonality evenness (VCE) score based on the Shannon Evenness Index, and used this score to quantify the uniformity of the clonal distribution of participant samples (appendix 1 p 42). VCE scores have a value between 0 and 1, with 1 representing complete evenness (all clones with uniform abundance) and 0 representing a perfectly monoclonal architecture. We conducted Wilcoxon rank sum tests to compare VCE scores between subgroups of participants using the wilcox.test() function in R version 4.2.1. p values of less than 0·05 were classed as statically significant.

Estimating the performance of VCE scoring and proviral load thresholds for risk stratification

We first adjusted the cohort composition by repeated sampling with replacement to better reflect the proportions of HTLV-1 asymptomatic carriers with high proviral load and of carriers who progressed to ATL in the general population, according to the latest information from JSPFAD (appendix 1 p 41). The performance of proviral load thresholds and VCE scoring to classify individuals at risk or not at risk of progression to ATL was then estimated using the area under the receiver operating characteristic curve (AUC), the proportion of correctly classified individuals (ie, the accuracy), and Matthews correlation coefficient (appendix 1 p 43).

T-cell-receptor β clonotype sequencing

T-cell-receptor β (TCR-β) sequencing was conducted as previously described,[28] and raw data were analysed on the web-based application ARResT/Interrogate using the previously established EC-NGS.clonality-assessment scenario (appendix 1 p 43). TCR-β clonotypes were plotted on pie charts in a similar manner to that of HTLV-1 clonality data, but with different colours for the 20 topmost clones.

Role of the funding source

The funders of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report.

Results

60 participants were initially recruited to the study, and were divided into three subgroups of 20 participants each: the first comprised carriers who developed ATL, the second comprised carriers with a high proviral load (≥4%) who did not progress to ATL, and the third comprised carriers with a low proviral load (<4%) who did not progress to ATL. Four participants were subsequently excluded owing to an inconsistent diagnosis between timepoints (three from the first subgroup and one from the second) and one participant from the second subgroup was reassigned to the third subgroup. The final study population therefore comprised 56 participants (P1–P56) who were followed up by JSPFAD between Feb 6, 2003, and July 19, 2022: 17 who progressed to ATL (first subgroup: P40–P56) and 39 who did not progress to ATL (18 in the second subgroup: P22–P39; and 21 in the third subgroup: P1–P21). Characteristics of the study participants and the duration of follow-up for each subgroup are reported in the table and appendix 1 (p 42).

We conducted HTLV-1 clonality sequencing[14,27] on PBMCs obtained from all 56 participants (table, appendix 2 tab 3), sequencing at least two longitudinal samples, obtained at timepoints T1 and T2. T1 was defined as the time of analysis of the earliest available sample with full biological and clinical data, at which all participants were asymptomatic carriers. T2 was defined as the time at which the second sample was analysed; for participants who did not progress to ATL (P1–P39), this sample was taken a mean of 3·9 years (SD 1·3; range 0·7–8·3) after T1, whereas for participants who progressed (P40–P56), T2 was the time at which the first sequenced sample with an ATL diagnosis was obtained (3·4 years [2·2; 1·0–7·8] after T1). All participants who did not progress to ATL remained under further observation, for a mean of 7·5 years (3·0; 2·1–12·0) for those with low proviral load (<4%; defined as having a low risk of progression) and 9·7 years (3·4; 2·1–14·3) for those with high proviral load (≥4%; defined as having a high risk of progression), to ensure that they had not progressed to ATL after the last sequenced timepoint (table, appendix 2 tab 2). The 17 participants who progressed to ATL were monitored for a mean of 8·3 years (4·0; 2·0–16·5) after T1, 15 (88%) of whom were monitored beyond T2, which ensured the correct ATL subtype at the latest diagnosis. For six (35%) of the 17 participants who progressed to ATL, further clonality sequencing was conducted at additional timepoints after T2 (15 additional timepoints in total among the six

	Participants with low proviral load and no progression (P1–P21; n=21)	Participants with high proviral load and no progression (P22–P39; n=18)	Participants who progressed to ATL (P40–P56; n=17)
Age at baseline, years	54 (46–60)	50 (42–59)	60 (53–69)
Sex			
Male	6 (29%)	5 (28%)	7 (41%)
Female	15 (71%)	13 (72%)	10 (59%)
Proviral load at T1			
<4% (low risk)	21 (100%)	0	1 (6%)*
≥4% (high risk)	0	18 (100%)	16 (94%)
Diagnosis at T1	Asymptomatic carriers	Asymptomatic carriers	Asymptomatic carriers
Diagnosis at T2	Asymptomatic carriers	Asymptomatic carriers	ATL†
ATL classification‡			
Acute	NA	NA	6 (35%)
Lymphoma§	NA	NA	..
Chronic	NA	NA	2 (12%)
Smouldering	NA	NA	9 (53%)
Time between T1 and T2, years	4·1 (1·4; 1·8–8·3)	3·6 (1·2; 0·7–6·0)	3·4 (2·2; 1·0–7·8)
Time without progression to ATL, years	7·5 (3·0; 2·1–12·0)	9·7 (3·4; 2·1–14·3)	3·6 (2·5; 1·0–7·9)
Time between T1 and latest diagnosis, years	7·5 (3·0; 2·1–12·0)	9·7 (3·4; 2·1–14·3)	8·3 (4·0; 2·0–16·5)

Data are median (IQR), n (%), or mean (SD; range). T1 is the time at which the earliest available sample was analysed by HTLV-1 clonality sequencing. T2 is the time at which the second sample was analysed by HTLV-1 clonality sequencing; for participants who progressed to ATL, this is the time of ATL diagnosis or the time at which the first sequenced sample with an ATL diagnosis was obtained. Time without progression refers to the time between T1 and the latest available diagnosis as an asymptomatic carrier (for participants who did not progress to ATL) or between T1 and the time of ATL diagnosis (for those who did progress to ATL). ATL=adult T-cell leukaemia/lymphoma. HTLV-1=human T-cell lymphotropic virus type 1. NA=not applicable. P=participant. *One of 17 participants who progressed to ATL had a proviral load of 1%. †Of the 17 participants who progressed to ATL, nine had smouldering ATL, two had chronic ATL, and six had acute ATL. ‡The Shimoyama classification has historically subdivided ATL into four clinical subtypes, two aggressive (acute and lymphoma) and two indolent (chronic and smouldering), on the basis of clinical characteristics, laboratory features, and history of the disease.[29] §Biopsy samples from participants with the lymphoma subtype have not been stored in the biobank, preventing the assessment of clonal signatures in the solid lymphoid tumours that are characteristic of this classification.

Table: **Characteristics of study participants**

participants) to monitor ATL subtype dynamics over time (appendix 1 pp 35–37, appendix 2 tabs 2, 3). Demographic, clinical, and biological data for all available longitudinal samples from each participant are provided in appendix 2 (tab 2), including data from T1, T2, and additional non-sequenced timepoints corresponding to annual medical visits during JSPFAD follow-up.

Clonality sequencing of samples from asymptomatic carriers who did not progress to ATL predominantly revealed hundreds to thousands of HTLV-1 integration sites, corresponding to multiple clones of low and uniform abundance, at both T1 and T2 (figure 1A, appendix 1 pp 2–21, 32–34). Among these participants who did not progress, only three individuals had a clone that was relatively more abundant than the others on visual inspection of the pie charts (P21 [25% abundance], P25 [22%], and P34 [23%] at T2). The clonal architecture of participants at high risk who did not progress (ie, proviral load ≥4%) did not visually differ from that of participants at low risk who did not progress (ie, proviral load <4%) at any of the timepoints. By contrast, inspection of the clonal distribution at T1 (ie, the asymptomatic carrier stage) of participants who later progressed to ATL (P40–P56) revealed the presence of a single predominant clone or a few (two to four) predominant clones in 14 (82%) of 17 cases, and these clones were consistently maintained at similar or increased abundance at T2 (ie, the time of ATL diagnosis; figure 1B, appendix 1 pp 21–29, 35–37). The three participants who had a uniform clonal distribution at T1 (P46, P48, P49) had a similar polyclonal pattern at T2 despite their ATL diagnosis.

To more rigorously assess the validity of these observations and establish a metric for HTLV-1 clonality, we quantified the uniformity of clonal distribution in each sample using the Shannon Evenness Index (appendix 1 p 42), referring to the output as the VCE score. This score has a value between 0 and 1, with 1 being complete evenness or a perfectly polyclonal landscape and 0 representing a perfectly monoclonal architecture. As expected from the visual inspection of the pie charts, all participants who did not progress to ATL had high VCE scores (≥0·694) at T1, and those classed as being at high risk on the basis of proviral load were indistinguishable from those at low risk (p=0·49, Wilcoxon rank sum test). By contrast, the majority of participants (14 [82%] of 17) who progressed to ATL, including the participant with a proviral load of 1% (P42), had lower VCE scores (<0·694) at T1, reflecting a more monoclonal architecture (p<0·0001, Wilcoxon rank sum test; figure 2A). The three participants who progressed to ATL with a VCE score higher than 0·694 at T1 (P46, P48, and P49) were three of the nine participants with smouldering ATL, and had a uniform clonal distribution pattern at both T1 and T2 despite their ATL diagnosis

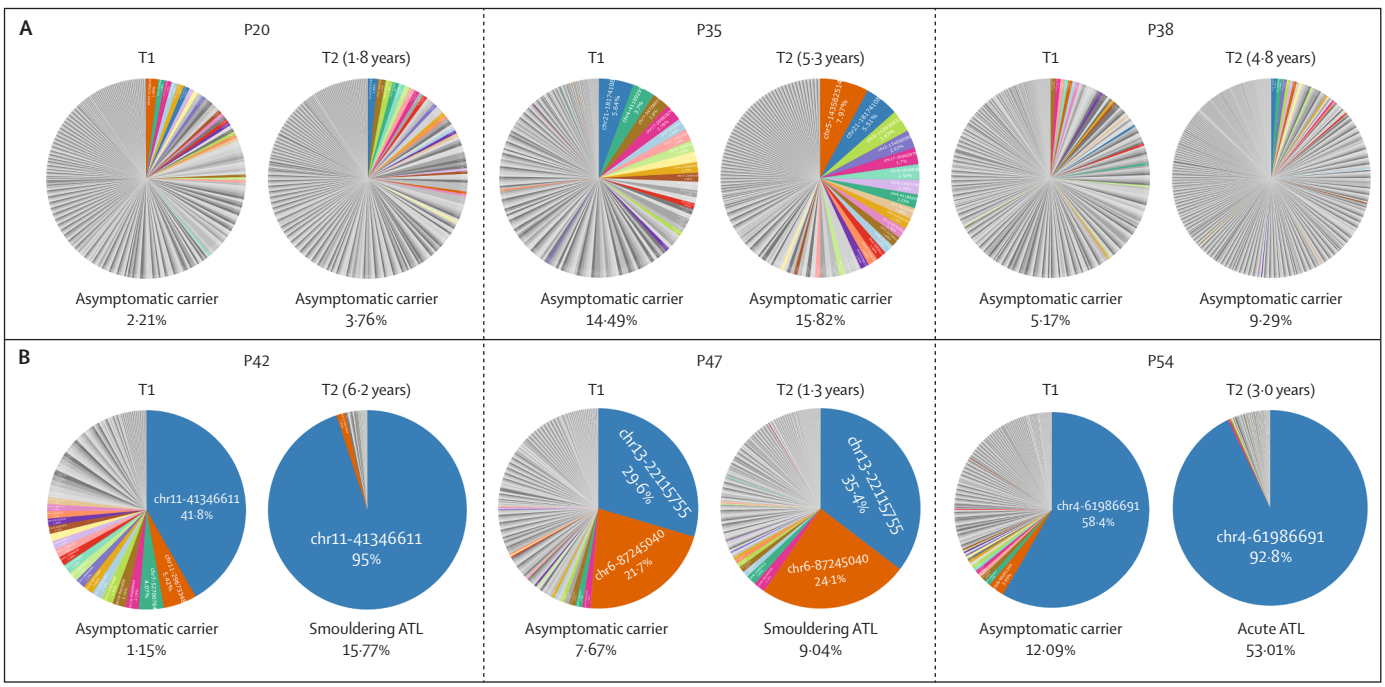

Figure 1: **Use of HTLV-1 clonality to identify asymptomatic carriers at high risk of progression to ATL**
Pie charts showing HTLV-1 clonal landscapes and relative integration site abundances for three representative participants who did not progress to ATL (A; P20, P35, and P38) and three representative participants who did progress to ATL (B; P42, P47, and P54) at T1 and T2. The time between T1 and T2 is shown in parentheses. Proviral loads are shown as percentages. Each slice of a pie represents an independent proviral integration site and its size shows the relative abundance of the corresponding clone among infected PBMCs. The top 20 most abundant clones are shown by consistent unique colours across samples. Pie charts of all sequenced samples for each participant are available in appendix 1 (pp 32–37). ATL=adult T-cell leukaemia/lymphoma. HTLV-1=human T-lymphotropic virus type 1. PBMCs=peripheral blood mononuclear cells.

(appendix 1 p 35). Unfortunately, these participants were lost to follow-up, and no further information on their clonal or ATL subtype dynamics over time is available (appendix 1 p 42).

Using standard classifier performance metrics, we compared proviral load with VCE scoring for classifying the risk of progression to ATL in individuals with HTLV-1.[30] In our sample of 56 participants, both the proportion who progressed to ATL and the proportion with a proviral load of at least 4% were notably higher than has been reported for the population of HTLV-1 carriers both in Japan and worldwide. According to JSPFAD statistics, 2% of carriers progressed to ATL (30% in our sample; 17 participants) and 25% of carriers have a proviral load of at least 4% (59% in our sample; 33 participants; appendix 1 p 41). To address this bias in the composition of our study cohort, we sampled (with replacement) 17 participants who progressed to ATL, approximately 190 participants with a high proviral load who did not progress to ATL, and approximately 640 participants with a low proviral load who did not progress to ATL, to better reflect typical proportions in the population (appendix 1 p 43). We generated 2000 augmented samples, enabling us to compute 95% CIs of our estimates. We first computed AUCs. AUCs were similar for proviral load (91 [95% CI 80–98]) and VCE scoring (91 [78–100]; figure 2B). The AUC is determined by sensitivity (true positive/[true positive + false negative]) and specificity (true negative/[true negative + false positive]), and does not consider the relative proportions of positives (individuals who progress to ATL; ie, true positive + false negative) and negatives (individuals who do not progress to ATL; true negative + false positive) in the population. For example, the method gives an equal weight to the false-positive cases resulting from prediction based on proviral load thresholds (approximately 20% of those who do not progress to ATL) and the false-negative cases resulting from prediction based on VCE scoring (approximately 18% of those who do progress to ATL)—despite the fact that these populations represent, in Japan, approximately 220 000 individuals versus approximately 4000 individuals, respectively. We therefore considered performance metrics that give equal weight to every individual: accuracy (ie, the proportion of correctly classified individuals, or [true positive + true negative]/[true positive + true negative + false positive + false negative]) and Matthews correlation coefficient, which is particularly useful to evaluate the performance of a binary classification model when the classes are imbalanced.[30] Using both methods, VCE scoring outperformed proviral load thresholds in the prediction of progression to ATL (figure 2C). Prediction based on proviral load thresholds has near-perfect sensitivity—ie, it identifies nearly all participants who will progress to ATL—but

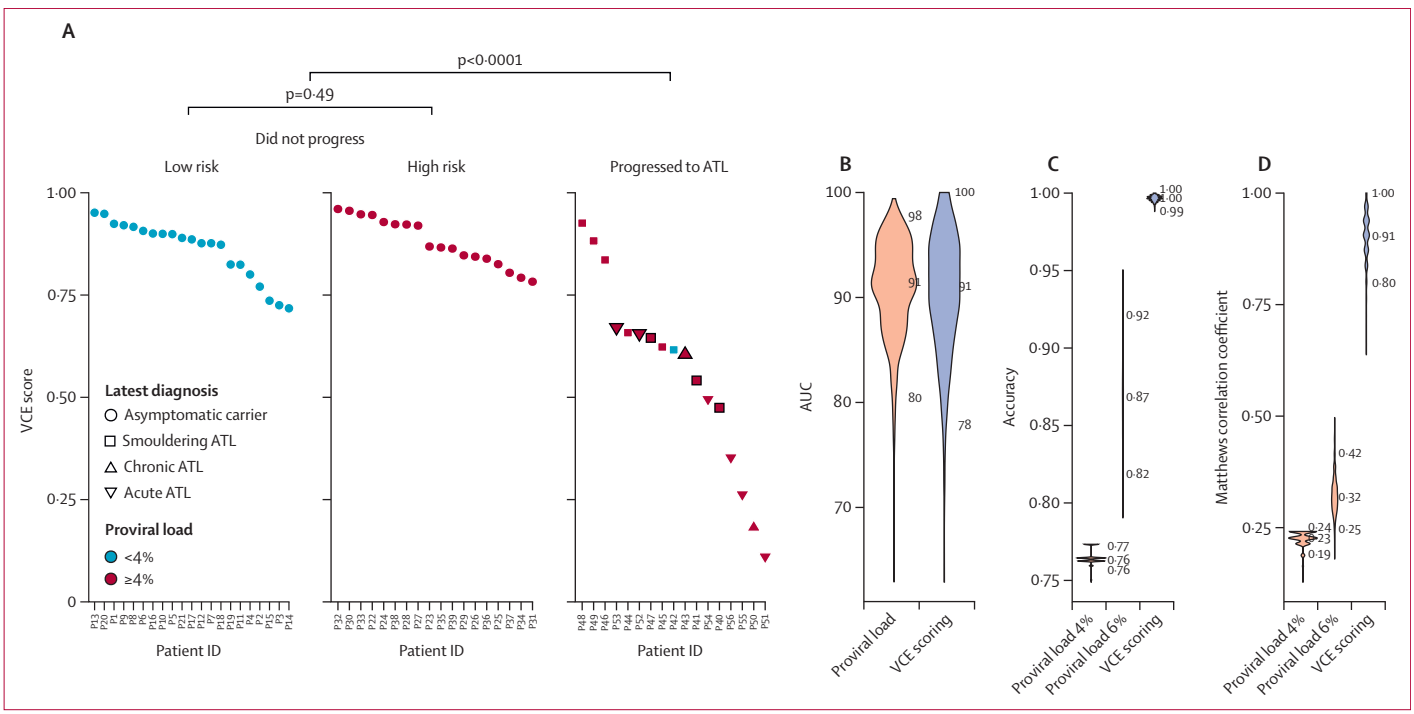

Figure 2: **Performance of VCE scoring and proviral load in measuring HTLV-1 clonality and stratifying risk in asymptomatic carriers**

(A) HTLV-1 clonality measured using VCE scoring at T1 for all participants, grouped by their progression status at T2. Participants who did not progress to ATL are further categorised according to their risk of progression as assigned by proviral load at T1: low risk (<4%), cyan; high risk (≥4%), red. The shapes of the datapoints represent the latest available diagnosis for each participant. Among the 17 participants who progressed to ATL, the shapes with thick borders represent the six participants who had more than one predominant integration site. (B, C) Performance of proviral load and VCE scoring for identifying asymptomatic carriers at high risk of developing ATL, in 2000 augmented samples generated to better reflect typical proportions of HTLV-1 asymptomatic carriers with high proviral load and of carriers who progressed to ATL in the general population (appendix 1 p 42). Data are estimate (95% CI). (B) AUC estimation for proviral load and VCE scoring. (C) Accuracy (left) and Matthews correlation coefficient (right) for proviral load (at the currently used threshold of 4% and the most recurrent best threshold of 6%) and VCE (at the most recurrent best threshold of 0·694; appendix 1 p 43). ATL=adult T-cell leukaemia/lymphoma. AUC=area under the receiver operating characteristic curve. HTLV-1=human T-lymphotropic virus type 1. VCE=viral clonality evenness.

at a cost of a very high number (approximately 220 000 in Japan) of false positives. Prediction based on VCE scoring has near-perfect precision, meaning that the likelihood of a carrier who is predicted to progress to ATL subsequently progressing is nearly 100%, with no false positives. However, this precision comes at the cost of a substantial number of false negatives (around 4000 individuals in Japan).

To assess whether the demographic characteristics of the study participants (appendix 2 tab 2) influenced the findings, we first tested for significant differences in age and sex between groups. HTLV-1 carriers who progressed to ATL were significantly older at enrolment (T1) than those who did not progress (median age 60 [IQR 53–69] years, *vs* 54 [IQR 44–60] years for all carriers [p=0·029] and 50 years [p=0·012] for carriers with a high proviral load). However, no significant differences were observed in sex between groups (appendix 1 p 30). To evaluate the potential confounding effect between age and VCE scoring, we conducted additional tests using the evolution of VCE scores in participants who did not progress to ATL as a proxy (appendix 1 p 38). On the basis of these results, we conclude that the age difference between participants who progressed to ATL and those who did not progress is unlikely to have affected our findings regarding the prognostic value of VCE scoring.

Although the majority (11 [65%] of 17) of participants who progressed to ATL with low VCE scores had a single dominant clone defined by a single integration site, for six (35%) of these 17 participants (P40, P41, P43, P47, P52, and P53), clonality analysis revealed the presence of two or more dominant integration sites of equivalent abundance (illustrated in figure 1B for P47; appendix 1 pp 35–37). This finding could indicate the occurrence of multiple transformed intermediate clones[12] or the integration of multiple proviruses in a single T-cell clone.[10,12] To distinguish between these two scenarios, we conducted TCR-β clonotype sequencing[28] (appendix 1 p 39) and quantified T-cell clonal abundances in a subset of six participants. As expected, carriers with a polyclonal integration profile had polyclonal TCR-β profiles and participants who progressed to ATL with a single dominant integration site had monoclonal TCR-β profiles (figure 3A, B). TCR-β sequencing of three participants carrying two or three abundant HTLV-1 integration sites showed monoclonal TCR-β profiles, indicating the presence of two or three proviral integrations within a single T-cell clone (figure 3C). Despite the presence of these multiple abundant integration sites, which tend to inflate evenness and overestimate proviral load, all six participants had a VCE score of less than 0·694 at T1. This finding suggests that clonality sequencing is robust at

Figure 3: **Viral clonality and TCR-β sequencing to show the integration of multiple HTLV-1 proviruses in a single T-cell clone**
(A) Viral clonality and TCR-β sequencing of asymptomatic carrier P20 at T2, revealing multiple integration sites and TCR-β clones. (B) Viral clonality and TCR-β sequencing of participants P51 and P54, both of whom progressed to ATL, at T2 and T3, characterised by a single predominant clone for both integration and TCR-β. (C) TCR-β sequencing of participants P41, P47, and P52, at T1 and one other timepoint as indicated. Participants P41 and P47, who had two abundant HTLV-1 integration sites, had monoclonal TCR-β profiles, indicating the presence of two integrations within a single T-cell clone. For participant P52, the number of equally highly abundant integration sites changed from four to three over time, whereas the number of TCR-β clonotypes decreased from three to one, indicating that the dominant TCR-β clonotype (red) at the ATL stage harbours three proviral integrations. For all panels, the time from T1 is shown in parentheses and the proviral load is shown as a percentage. ATL=adult T-cell leukaemia/lymphoma. HTLV-1=human T-lymphotropic virus type 1. TCR-β=T-cell-receptor β.

identifying carriers who will develop malignancy years later, even in the case of an ATL clone with multiple proviral insertions.

Altogether, our findings suggest that VCE scoring outperforms proviral load for risk stratification of HTLV-1 carriers with high proviral load and additionally identifies carriers with low proviral load who are at high risk of progression. VCE scoring also reliably detects individuals at high risk even in the presence of ATL clones with multiple proviral insertion sites. Finally, VCE scores could serve as a more reliable diagnostic classifier in borderline ATL cases, thereby reducing diagnostic uncertainty (figure 4).

Discussion

For both HTLV-1 carriers and their clinicians, the probability of developing ATL remains a substantial concern. Although the commonly cited 20% lifetime risk for carriers with a high proviral load (≥4%) is widely recognised, it offers little to these individuals. Many continue to have considerable anxiety despite the fact that the majority

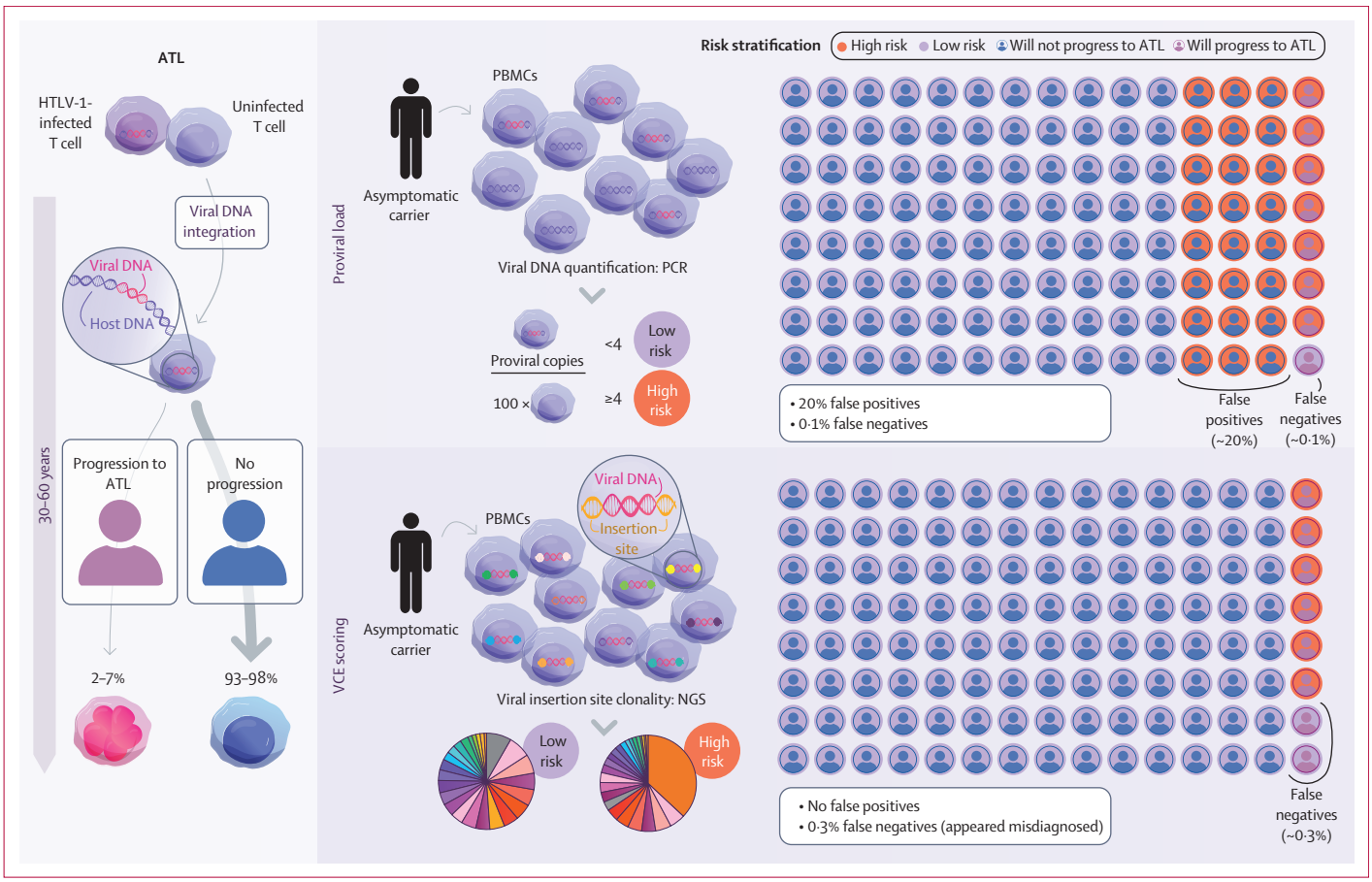

Figure 4: **Risk stratification of asymptomatic HTLV-1 carriers on the basis of proviral load and VCE scoring**
ATL=adult T-cell leukaemia/lymphoma. HTLV-1=human T-lymphotropic virus type 1. NGS=next-generation sequencing. PBMCs=peripheral blood mononuclear cells. VCE=viral clonality evenness.

(around 80%; ie, around 20% of all individuals with HTLV-1) will never progress to malignancy. For clinicians, the absence of more reliable predictors adds to both prognostic and diagnostic uncertainty. In this study, we show that VCE could act as a prognostic biomarker to identify asymptomatic carriers at increased risk of progression to ATL. Considering a combination of appropriate metrics, our estimations show that VCE scoring outperforms proviral load thresholds in predicting the risk of progression by achieving a higher rate of correctly classified individuals and eliminating false positives—a major limitation of predictions based on proviral load. Clonality scoring could help to reassure approximately 80% of carriers who are categorised as being at high risk of progressing to ATL on the basis of proviral load, thereby reducing anxiety and improving their quality of life. Furthermore, VCE scoring could help to identify carriers who have a low proviral load but are at high risk of progression to ATL—a category of individuals who currently do not receive the same level of attention, despite the fact that approximately 5% of patients with ATL in Japan had a baseline proviral load of less than 4% at the carrier stage. Finally, VCE scores are robust at identifying individuals at high risk, even in the case of multiple proviral insertion sites within a single T-cell clone.

The lower sensitivity of VCE scoring than proviral load thresholds was influenced by three participants with smouldering ATL who had high VCE scores and a polyclonal pattern at T1; these cases were classified as false negatives (P46, P48, and P49). Notably, all three participants had a similar uniform clonal distribution and a high VCE score at T2, despite their ATL diagnosis. These participants were classified as having smouldering ATL solely on the basis of the percentage of abnormal lymphocytes (≥5%; the values being 7% for P46, 10% for P48, and 5% for P49), and they had no clinical symptoms.[29] The difficulty in distinguishing between individuals who are carriers and those who have smouldering ATL is apparent, particularly in the absence of clinical symptoms, as the diagnosis relies primarily on examination of morphological blood smears. These cases highlight the shortcomings of the current diagnostic procedures and the limitations of the Shimoyama classification, which has remained unchanged since its original description in 1991.[7,29] Taken together, our findings highlight the need for better diagnostic molecular

classifiers, addressing the urgent need to go beyond the Shimoyama classification for diagnosis and treatment, and strengthen the support for implementing clonality scoring as a more reliable diagnostic index for borderline cases of ATL.[8] Such an approach will require further investigation.

The cohort size in our study might seem small; however, the JSPFAD nationwide cohort study in Japan, which began by enrolling individuals as asymptomatic carriers in 2002 and now holds longitudinal samples from both participants who progressed to ATL and participants who have not progressed, is, to our knowledge, the largest longitudinal biobank of samples from individuals with HTLV-1 worldwide. Participants within JSPFAD are invited to annual medical appointments, during which clinical and biological data are systematically collected, and blood samples are added to the biobank over many years. To our knowledge, this invaluable resource is not available elsewhere. Ideally, our findings would benefit from further validation in an independent retrospective cohort; however, because of the rarity of both ATL itself and matching longitudinal samples from individuals with the condition, such a cohort is currently not available, and several decades would be required for sufficient numbers of participants in a newly established cohort to progress to ATL. Nevertheless, despite these limitations inherent to the disease, the consistency of our findings with previous independent observations[22–24] strengthens our conclusion that clonality measured by VCE scoring can be used as an early biomarker of progression to ATL. Fostering international collaboration to enhance the recruitment of HTLV-1 carriers worldwide will enable larger studies and further evaluation of the prognostic potential of HTLV-1 clonality scoring.

A limitation of our study is the inability to evaluate VCE scores for individuals with the lymphoma subtype of ATL. Although JSPFAD holds blood samples from HTLV-1 carriers and participants with ATL, biopsy samples from participants with the lymphoma subtype have not been stored in the biobank, preventing the assessment of clonal signatures in the solid lymphoid tumours that are characteristic of this subtype. To address this limitation and test whether VCE scoring could reveal the lymphoma precursor clone in PBMCs of individuals with this ATL subtype long before symptoms occur, we analysed clonal signatures of samples from two patients treated at the haematology department of the Necker–Enfants Malades Hospital (Paris, France), data on whom have been previously reported.[14,31] These patients, denoted ATL 43 and ATL 44, had matched blood samples collected at the carrier stage many years before the development of the lymphoma subtype of ATL (appendix 1 p 40). For patient ATL 43, clonality sequencing of PBMCs collected 9 years before diagnosis revealed a high-risk blood profile (premalignant clone 85%, VCE <0·694) despite a low-risk proviral load status (<4% at the carrier stage). By contrast, PBMCs from patient ATL 44 had high VCE and a polyclonal landscape both 6 years before diagnosis (premalignant clone 3%, VCE ≥0·694) and at diagnosis (lymphoma clone 12%, VCE ≥0·694). This analysis suggests that VCE scoring could identify the fraction of carriers at high risk of progression who carry the premalignant lymphoma clone in blood, even if they have a low proviral load. However, a high VCE score will not systematically exclude this risk given that, in approximately two-thirds of patients with the lymphoma subtype of ATL, the malignant clone is not expanded in the blood.[22,23] Altogether, we estimate that the proportion of individuals who progress to ATL that is potentially missed by the clonality sequencing of carriers is approximately 0·2%, considering that the lymphoma subtype accounts for approximately 15% of all cases of ATL.

Our clonality approach analyses DNA, eliminating the need for viable PBMCs and flow cytometry sorting, which makes it more widely accessible than previously reported methods such as the analysis of the proportion of $CD4^+CADM1^+$ T cells by flow cytometry, T-cell-receptor sequencing, or tracking ATL-like driver mutations.[22–24] Furthermore, this method uses commercially available NGS adaptors and libraries can be highly multiplexed, substantially increasing sensitivity and reducing cost compared with previously developed NGS-based clonality protocols.[13,15,25] Altogether, clonality sequencing and VCE scoring are straightforward to implement in routine clinical practice.

In conclusion, our findings have the potential to change clinical practice in HTLV-1 related haematology-oncology, not only improving patient outcomes but also the wellbeing of people living with HTLV-1 worldwide. In response to the increasing awareness of HTLV-1 internationally,[5] WHO published a fact sheet describing the major concerns to be addressed, with a strong focus on prevention and prediction.[32] Our work addresses these unmet needs by proposing a predictive, and potentially diagnostic, biomarker. The annual assessment of clonality scores in asymptomatic carriers, regardless of their proviral load, has the potential to identify individuals with low VCE who could benefit from more intensive follow-up. This approach could help to reassure a considerably larger proportion of individuals than current monitoring processes, thereby improving the quality of life of HTLV-1 carriers, as well as guide prospective trials to evaluate the clinical benefit of pre-emptive therapeutic interventions for asymptomatic carriers at high risk of progression, aiming to prevent progression to aggressive, treatment-refractory disease. Early interventions could include antiretroviral therapy, clearance of dominant clones by combination treatment with zidovudine–interferon alfa-2b, mogamulizumab (anti-CCR4 antibody), and lenalidomide, therapeutic strategies that show good response rates in indolent forms of ATL and might be more efficient in this pre-emptive context. In such trials, increasing the VCE score could be a good endpoint through which to assess efficacy.

Contributors

SDK analysed the data. SDK, AB, MG, and AVdB contributed to the interpretation of the results and wrote the manuscript. MA, JW, and KD did the clonality sequencing experiments. VH contributed to the clonality

pipeline. KU, JM, AU, OH, and AM followed up participants, collected clinical samples, and provided clinical and biological data. MI provided information and statistics on the JSPFAD cohort. AWL and PMK did TCR-β sequencing and provided NGS datasets. TW coordinated the JSPFAD study, provided participant samples, and contributed to cohort analysis. AVdB designed and supervised the study and provided funding. SDK and AVdB accessed and verified the data. All authors had full access to all the data in the study and had final responsibility for the decision to submit for publication.

Declaration of interests

We declare no competing interests.

Data sharing

Sequence data are deposited in the European Nucleotide Archive (hosted by the European Molecular Biology Laboratory's European Bioinformatics Institute) and are accessible through accession number PRJEB55635. All other relevant data are available within the Article and its appendices. The code and a detailed outline of the clonality analysis workflow are available on GitHub (https://github.com/GIGA-AnimalGenomics-BLV/Public/tree/master/PIC).

Acknowledgments

This work was supported by the Association Jules Bordet, the Fonds de la Recherche Scientifique (FNRS), Télévie (including grants to JW and VH), the Fondation Contre le Cancer (Belgium), the Région wallonne WALInnov project CAUSEL (convention number 1710030), the Fonds Leon Frederic, and grants from JSPS-KAKENHI (JP 16H06277) and CoBiA (22H04923) to TW. SDK holds a postdoctoral research fellowship of Télévie (FNRS) and MA is a Télévie Scientific Research Worker. We thank all individuals who participated in this study; Wouter Coppieters, Latifa Karim, Manon Deckers, Emilie Detry, and the GIGA Genomics Platform for sequencing services and technical support; the GIGA Bioinformatics Platform for NGS data analysis; and Illumine for designing a graphical summary (figure 4). Computational resources were provided by GIGA and the Consortium des Équipements de Calcul Intensif.

References

1 Ishitsuka K, Tamura K. Human T-cell leukaemia virus type I and adult T-cell leukaemia-lymphoma. *Lancet Oncol* 2014; **15:** e517–26.
2 Imaizumi Y, Iwanaga M, Nosaka K, et al. Prognosis of patients with adult T-cell leukemia/lymphoma in Japan: a nationwide hospital-based study. *Cancer Sci* 2020; **111:** 4567–80.
3 O'Donnell JS, Hunt SK, Chappell KJ. Integrated molecular and immunological features of human T-lymphotropic virus type 1 infection and disease progression to adult T-cell leukaemia or lymphoma. *Lancet Haematol* 2023; **10:** e539–48.
4 Gessain A, Cassar O. Epidemiological aspects and world distribution of HTLV-1 infection. *Front Microbiol* 2012; **3:** 388.
5 Martin F, Tagaya Y, Gallo R. Time to eradicate HTLV-1: an open letter to WHO. *Lancet* 2018; **391:** 1893–94.
6 Cook LB, Fuji S, Hermine O, et al. Revised adult T-cell leukemia-lymphoma international consensus meeting report. *J Clin Oncol* 2019; **37:** 677–87.
7 Cook LB, Phillips AA. How I treat adult T-cell leukemia/lymphoma. *Blood* 2021; **137:** 459–70.
8 Katsuya H. Current and emerging therapeutic strategies in adult T-cell leukemia–lymphoma. *Int J Hematol* 2023; **117:** 512–22.
9 Bangham CRM, Cook LB, Melamed A. HTLV-1 clonality in adult T-cell leukaemia and non-malignant HTLV-1 infection. *Semin Cancer Biol* 2014; **26:** 89–98.
10 Rosewick N, Durkin K, Artesi M, et al. *Cis*-perturbation of cancer drivers by the HTLV-1/BLV proviruses is an early determinant of leukemogenesis. *Nat Commun* 2017; **8:** 15264.
11 Yoshida M, Seiki M, Yamaguchi K, Takatsuki K. Monoclonal integration of human T-cell leukemia provirus in all primary tumors of adult T-cell leukemia suggests causative role of human T-cell leukemia virus in the disease. *Proc Natl Acad Sci U S A* 1984; **81:** 2534–37.
12 Cook LB, Melamed A, Niederer H, et al. The role of HTLV-1 clonality, proviral structure, and genomic integration site in adult T-cell leukemia/lymphoma. *Blood* 2014; **123:** 3925–31.
13 Gillet NA, Malani N, Melamed A, et al. The host genomic environment of the provirus determines the abundance of HTLV-1-infected T-cell clones. *Blood* 2011; **117:** 3113–22.
14 Artesi M, Marçais A, Durkin K, et al. Monitoring molecular response in adult T-cell leukemia by high-throughput sequencing analysis of HTLV-1 clonality. *Leukemia* 2017; **31:** 2532–35.
15 Firouzi S, López Y, Suzuki Y, et al. Development and validation of a new high-throughput method to investigate the clonality of HTLV-1-infected cells based on provirus integration sites. *Genome Med* 2014; **6:** 46.
16 Iwanaga M. Epidemiology of HTLV-1 infection and ATL in Japan: an update. *Front Microbiol* 2020; **11:** 1124.
17 Iwanaga M, Watanabe T, Utsunomiya A, et al. Human T-cell leukemia virus type I (HTLV-1) proviral load and disease progression in asymptomatic HTLV-1 carriers: a nationwide prospective study in Japan. *Blood* 2010; **116:** 1211–19.
18 Demontis MA, Hilburn S, Taylor GP. Human T cell lymphotropic virus type 1 viral load variability and long-term trends in asymptomatic carriers and in patients with human T cell lymphotropic virus type 1-related diseases. *AIDS Res Hum Retrovir* 2013; **29:** 359–64.
19 Hodson A, Laydon DJ, Bain BJ, Fields PA, Taylor GP. Pre-morbid human T-lymphotropic virus type I proviral load, rather than percentage of abnormal lymphocytes, is associated with an increased risk of aggressive adult T-cell leukemia/lymphoma. *Haematologica* 2013; **98:** 385–88.
20 Imaizumi Y, Iwanaga M, Tsukasaki K, Hata T, Tomonaga M, Ikeda S. Natural course of HTLV-1 carriers with monoclonal proliferation of T lymphocytes ("pre-ATL") in a 20-year follow-up study. *Blood* 2005; **105:** 903–04.
21 Firouzi S, Farmanbar A, Nakai K, et al. Clonality of HTLV-1–infected T cells as a risk indicator for development and progression of adult T-cell leukemia. *Blood Adv* 2017; **1:** 1195–205.
22 Rowan AG, Dillon R, Witkover A, et al. Evolution of retrovirus-infected premalignant T-cell clones prior to adult T-cell leukemia/lymphoma diagnosis. *Blood* 2020; **135:** 2023–32.
23 Wolf SN, Haddow J, Greiller C, Taylor GP, Cook LBM, Rowan AG. Quantification of T cell clonality in human T cell leukaemia virus type-1 carriers can detect the development of adult T cell leukaemia early. *Blood Cancer J* 2021; **11:** 66.
24 Makiyama J, Kobayashi S, Watanabe E, et al. $CD4^+$ $CADM1^+$ cell percentage predicts disease progression in HTLV-1 carriers and indolent adult T-cell leukemia/lymphoma. *Cancer Sci* 2019; **110:** 3746–53.
25 Wada Y, Sato T, Hasegawa H, et al. RAISING is a high-performance method for identifying random transgene integration sites. *Commun Biol* 2022; **5:** 535.
26 Kuramitsu M, Okuma K, Yamochi T, et al. Standardization of quantitative PCR for human T-cell leukemia virus type 1 in Japan: a collaborative study. *J Clin Microbiol* 2015; **53:** 3485–91.
27 Rosewick N, Hahaut V, Durkin K, et al. An improved sequencing-based bioinformatics pipeline to track the distribution and clonal architecture of proviral integration sites. *Front Microbiol* 2020; **11:** 587306.
28 Brüggemann M, Kotrová M, Knecht H, et al. Standardized next-generation sequencing of immunoglobulin and T-cell receptor gene recombinations for MRD marker identification in acute lymphoblastic leukaemia: a EuroClonality-NGS validation study. *Leukemia* 2019; **33:** 2241–53.
29 Lever J, Krzywinski M, Altman N. Classification evaluation. *Nat Methods* 2016; **13:** 603–04.
30 Shimoyama M. Diagnostic criteria and classification of clinical subtypes of adult T-cell leukaemia-lymphoma. A report from the Lymphoma Study Group (1984–87). *Br J Haematol* 1991; **79:** 428–37.
31 Marçais A, Lhermitte L, Artesi M, et al. Targeted deep sequencing reveals clonal and subclonal mutational signatures in adult T-cell leukemia/lymphoma and defines an unfavorable indolent subtype. *Leukemia* 2021; **35:** 764–76.
32 WHO. Human T-lymphotropic virus-1. Dec 4, 2024. https://www.who.int/news-room/fact-sheets/detail/human-t-lymphotropic-virus-type-1 (accessed Dec 15, 2024).