論 文 の 内 容 の 要 旨

論文題目 Monitoring the clonal composition of HTLV-1-infected cells（HTLV−1 感染細胞のクローナリティ解析）

氏　　名　フィルジ　サナーズ

<Abstract>

Human T-cell leukemia virus type-I (HTLV-1) mainly survives *in vivo* by persistent proliferation of infected cells. HTLV-1 infection is the initial necessary event of multiple leukemogenic events that lead to adult T-cell leukemia (ATL) onset. As the first generation of studies on ATL risk factors, our Joint Study on Predisposing Factors of ATL Development (JSPFAD) group demonstrated that a proviral load (PVL) >4% is one of the risk factors for progression to ATL; however, PVL alone cannot predict development of the disease. Moreover, how the threshold of 4% and high levels of PVL are maintained and how they contribute to ATL onset remains to be elucidated.

Thus, with revolutionized insights, I have started a next generation of studies on prevention and revealing the molecular mechanisms of ATL development. For this purpose, I developed and validated an original methodology to detect clonality as accurate as qPCR while also taking advantage of the ability of deep sequencing to precisely characterize and distinguish large numbers of infected clones, based on provirus integration sites. This new methodology has been published, and is currently the most reliable method for accurate analysis of HTLV-1 clonality that can be used for clinical applications worldwide [Firouzi *et al. Genome medicine* 2014]. The realization of the potential clinical applications of this methodology will have far-reaching impacts on the diagnosis, prognosis, and treatment of infected individuals. Here I present my original methodology and part of pilot data on the clonal composition of HTLV-1 infected cells.

Summary of results

<Background and the necessity of this study>

ATL is a highly aggressive leukemia of T-cells, with an extremely poor prognosis and a short median survival time due to development of multidrug resistance. Prevention and treatment of ATL remain to be unresolved problems.

In 2002 JSPFAD was established as a nationwide collaborative study group to collect biomaterial samples from individuals infected with human T-cell leukemia virus type-I (HTLV-1) to facilitate research on the mechanisms and risk factors associated with ATL development. In the first generation of studies on ATL risk factors, JSPFAD assessed the correlation between disease outcome and proviral load (PVL). PVL represents the burden of HTLV-1 infection, defined as the percentage of infected cells among the total peripheral blood mononuclear cells (PBMCs), accurately measurable by qPCR. The JSPFAD initiative has currently collected > 9000 samples, all of which have had their PVL measured in our laboratory. PVL levels are different among infected individuals, with patients with malignant ATL having a significantly higher PVL than asymptomatic carriers (ACs). The initial JSPFAD study showed that a PVL >4% is one of the risk factors for progression to ATL [Iwanaga *et al. Blood* 2010]. However, some of ACs have abnormally high PVLs but do not develop ATL, and some infected individuals with low PVLs develop acute ATL. Thus, although an elevated PVL is currently the best-characterized risk factor associated with ATL development, a high PVL alone is not sufficient to predict disease progression and there is a need to discover additional predictive factors. Moreover, the mechanisms behind the maintenance of high levels of PVL and how these high PVLs lead to ATL onset need to be elucidated.

HTLV-1 infection is the initial necessary event among the multiple leukemogenic events that lead to ATL onset. HTLV-1 integrates into the human genome and maintains itself *in vivo* through persistent clonal growth of primarily infected cells. Following a long latency period of 40–60 years, about 5% of infected individuals convert from a polyclonal population of HTLV-1 infected cells into a monoclonal pattern that terminates in ATL onset. The monoclonal proliferation of HTLV-1-infected cells as a hallmark of ATL was

first detected by Southern blotting showing monoclonal bands [Yoshida *et al*. 1984]. Later, PCR-based analyses isolated HTLV-1 provirus integration sites and revealed that in addition to a monoclonal proliferation of infected cells, an oligoclonal or polyclonal proliferation occurs even in nonmalignant HTLV-1 carriers [Wattel *et al.* 1995, Etoh *et al.* 1997]. The overall proliferation levels of infected cells (PVL) are quantifiable by qPCR, and the general patterns of proliferation can be identified by the conventional techniques of Southern blot and inverse PCR. However, in-depth monitoring of the clonal composition of infected cells requires an advanced quantitative method that fulfills the three main criteria:

(1) High throughput isolation of a large numbers of integration sites (2) Detection of low abundance clones with high sensitivity even from the sample with low PVLs (3) Accurate measurement of the number of the infected cells in each clone (clone size).

Recently, a research group from the Imperial College of London devised a method that met only the first two criteria. Their method employed sonication to shear DNA to generate fragments of different lengths as a strategy for making unique fragments prior to PCR for the determination of clone size [Gillet *et al. Blood 2011]*. Owing to the limited variation in DNA fragment size observed with shearing, the probability of generating starting fragments of the same lengths is high, leading to a nonlinear relationship between fragment length and clone size; thus, introducing high error with this method. Therefore, Gillet *et al.* used a calibration curve to statistically correct the shear site data. However, even with these statistical corrections, they had a bias of at least >20% in the prediction of large clones [Berry *et al. Bioinformatics* 2012].

I showed that a major problem with shear site strategy of Gillet *et al*. is that practically shear sites can provide <250 variations [Firouzi *et al. Genome medicine* 2014]. This number of variations is not enough to accurately estimate the size of clones because most of the time, the number of infected cells in each clone exceeds the number of variations of the shear sites. Because the incidence of large clones (clones with >250 infected cells) increases with disease progression from the healthy AC state to the malignant states of smoldering, chronic, or acute ATL, an accurate measurement of clone size, and particularly of large clones, is of great clinical significance. Because the method of Gillet *et al*. leads to an underestimation of the clone sizes, the development of an alternative methodology with a high accuracy is necessary for clinical applications.

<Results and Discussion>

<Approaches of our study to fulfill all three aforementioned criteria >

A novel methodology for accurate quantitative analysis of clonality with a potential far-reaching impact on worldwide clinical applications [Firouzi *et al. Genome medicine* 2014]

I conducted a comprehensive multidisciplinary study combining our expertise in the field of HTLV-1 with genomics and bioinformatics analysis. I took advantage of next-generation sequencing (NGS) technology, using a tag system and an i*n silico* analysis pipeline to develop and internally validate a new high-throughput methodology (Figure 1). Analyzing control samples with already known clone sizes ensured accurate measurement of the size of clones using this method. This high-throughput method enables specific isolation of HTLV-1 integration sites, and allows for accurate quantitative clonality analysis of not only the major clones and high-PVL samples but also low-abundance clones (minor clones) and samples with low PVLs (Figure 2, 3).. An original strategy to remove PCR bias and to measure clone size was developed using a tag system, in which 8-bp random nucleotides are incorporated at the end of DNA fragments. Each tag acts as a molecular barcode, which gives each DNA fragment a unique signature prior to PCR. Information on the frequency of observed tags from the deep-sequencing data can be used to remove PCR duplicates, and thereby more

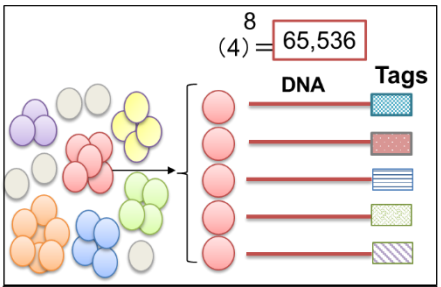

Figure 1. Prior to PCR, we incorporated 8-bp random tags into each DNA fragment to uniquely mark them. Random tags could theoretically provide approximately 65,536 variations. The number of potential variations is expected to amply cover large numbers of infected cells.

accurately estimate the original clonal abundance in the starting sample. Owing to their random design, the tags theoretically provide approximately 65,536 variations (4^8 = 65,536). This degree of potential variation is expected to provide a unique tag for a large number of sister cells in each clone (Figure 1). I proved my methodology to be reliable for isolating large numbers of integration sites and to be accurate for quantifying clone size. To the best of our knowledge, our methodology is the first in which the accurate size of clones is able to be experimentally measured without using any statistical corrections. This new methodology is currently the most reliable method for accurate analysis of HTLV-1 clonality available worldwide (Figure 2, 3) [Firouzi *et al. Genome medicine* 2014].

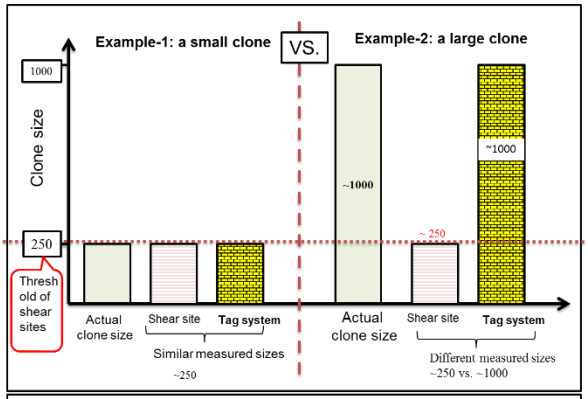

Figure 2. A simple comparison of shear site and tag variations
A simple image representing that the two approaches result in the similar clone sizes when the size of clones are small enough to be covered by shear site approach (Example-1). However, when the size of clones exceeds the shear site variations (~250 based on our experiments), clone sizes get underestimated by the shear site approach. Thus, the size of large clones are only measureable by the tag system (Example-2).

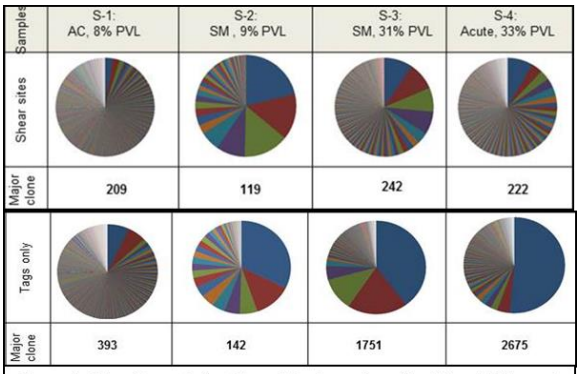

Figure 3. Clonality analysis of four clinical samples with different PVLs and disease states.The clone size of the first major clone (the largest clone) of each sample measured by shear site *vs.* tags system was [S-1: 209 *vs.* 393], [S-2:119 *vs.* 142], [S-3:242 *vs.* 1751],and [S-4:222 *vs.* 2675]. In the case of S-3 and S-4 which harboured large clones, the clone sizes measured by shear sites were significantly different from those of the tag system.

Results and discussion: "Analysis of Clinical samples"

A method to enable accurate quantification of clonality is in the center of this project

In this study, I aimed to connect the clonal composition of HTLV-1 infected cells with the diagnosis, prediction of prognosis, and elucidation of the mechanism underlying the multistep leukemogenesis of ATL. Similar to the project of JSPFAD on PVL, I have planned to study the clonality patterns of HTLV-1 infected individuals by taking advantage of my new methodology to accurately quantify the clonality of HTLV-1 infected cells. In the following sections, I present the results of my pilot data.

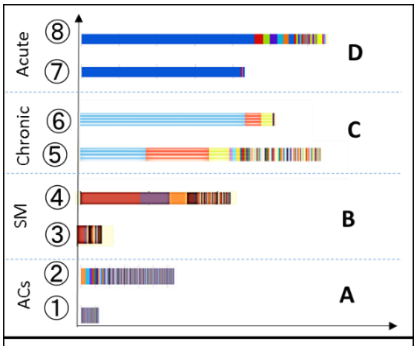

Figure 4. A) Acs showed Polyclonal patterns of infected cells. B)Smoldering patients showed polyclonal or Oligoclonal expansions C) Chronic samples showed Monoclonal patterns or largely expanded Oligoclonal patterns (major clones harbored large numbers of infected cells).D) Acute samples showed Monoclonal patterns.

Although the number of analyzed samples is limited, our initial data suggested different clonality patterns specific to individuals who were AC and those with the different subtypes of ATL (Figure 4). I analyzed samples from ACs, patients with the indolent types of ATL [smoldering (SM) and chronic] and patients with aggressive ATL (acute). Despite similar PVLs, AC vs. SM could be distinguished using clonality patterns (polyclonal vs. a shift towards oligoclonal). The clones of ACs showed a uniform distribution pattern with no large difference in clone size; however, clones of SM types had non-uniform sizes (Figure 4A-B). Chronic subtypes showed expanded oligoclonal patterns with a large shift to monoclonality (Figure 4C). All of the samples from patients with acute ATL harbored a largely expanded clone with a high absolute number of infected cells (Figure 4D). The clonality pattern of the chronic samples was more similar to the acute than the smoldering types (Figure 4C-D).

Due to diverse clinical manifestations and varying prognosis, ATL patients are categorized into distinct subtypes, based on standard clinical criteria: presence of organ involvement, leukemic manifestation, and levels of lactate dehydrogenase (LDH) and calcium. Currently, in clinical practice distinct treatment

strategies are used for the different subtypes of ATL [Tsukasaki *et al.* JCO 2009]. Therefore, classifying ATL patients into distinct subgroups is of high importance for selecting appropriate therapeutic interventions [Tsukasaki *et al.* Hematology 2013]. Considering the intimate link between ATL diagnosis and treatment, a more robust classification of ATL subtypes mediated by HTLV-1 clonal composition would be of fundamental clinical significance. Further examination of clonality patterns with a greater numbers of samples is necessary to validate the relationship between clonality patters and ATL subtypes, and to apply these patterns to diagnosis.

Impact on Prognosis and Prevention: An immediate demand for an effective prognostic indicator of ATL onset In a pilot study, I obtained data from four SM patients over 4 years. Two of the samples showed no progression in disease status (T1 = SM, T2 = SM); the other samples had progression into the chronic stage over the time course (T1 = SM, T2 = Chronic) (Figure 5). I detected a significant difference between the clonality patterns of the two sample sets independent of their PVL. Non-progressed samples manifested a polyclonal or oligoclonal expansion with a low number of infected cells (Figure 5B). However the progressed samples manifested monoclonal or largely expanded oligoclonal patterns (Figure 5A). Moreover, when I analyzed a particular individual over a time course of 6 years who progressed from AC to acute ATL (T1: AC, T2: AC, T3: Chronic, and T4: Acute), the major clone of the T4: acute state (showed with asterisk mark) was found to be dominant in earlier time points. This suggests the potential connection between clone size and the fate of the clone (Figure 6). In addition, I examined the effect of therapy on clonality patterns of patients before and after treatment. I could detect both stable and fluctuating clones from these samples. Most of the samples harbored a stable major clone before and after relapse (Figure 7A). However, in one patient I did find changes in size and order of the clones before and after treatment (Figure 7B).

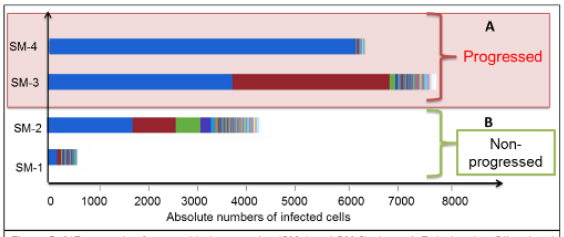

Figure 5. A)Progression free smoldering samples (SM-1 and SM-2) showed Polyclonal or Oligoclonal patterns (with small numbers of infected cells). B) Progressed samples (SM-3 and SM-4) showed largely expanded Oligoclonal and Monoclonal patterns, respectively.

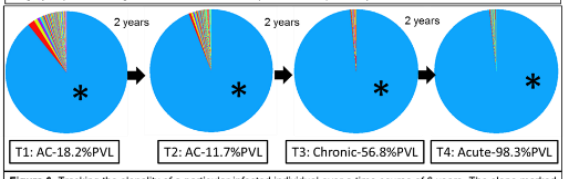

Figure 6. Tracking the clonality of a particular infected individual over a time course of 6 years. The clone marked by asterisks was integrated at chr2:74467952 position, and maintained over time.

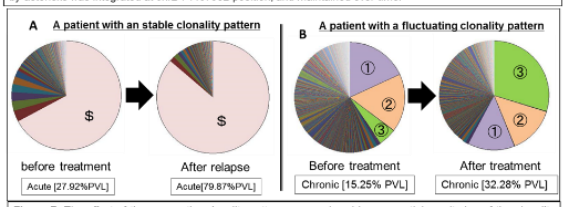

Figure 7. The effect of therapy on the clonality pattern was analyzed by sequential monitoring of the clonality patterns before and after treatment. A) The clone marked by $ was stable and dominant before and after therapy. B) Clones ①②③ were undergo change in size and order following treatment. Clone① and Clone ③ were dominant before and after therapy, respectively.

Although still preliminary, the data suggests that clonality patterns can be of prognostic use to patients. To pursue this, large-scale expansion of the project is recommended. This analysis should be helpful for decision making or developing timely and appropriate therapeutic intervention, based on clonality status of patients. ACs harbor a polyclonal population of HTLV-1 infected cells, whereas ATL patients show monoclonal patterns. Thus, changes in the clonality pattern and onset of a clonal expansion of HTLV-1-infected cells are a risk indicator of progression into ATL. The comparison of clonality patterns in individuals who progress from AC to development of ATL is expected to provide critical information on the clonality alterations that are associated with the transition from the AC to ATL state. Using this information as a prognostic indicator appears to be beneficial for the early detection of ATL onset, and eventually, ATL prevention. Accurate monitoring of the clonality patterns among infected individuals may help us to differentiate progressive and non-progressive patterns as well as assessments of the risk of disease development.

<Conclusion> In this next generation study on ATL risk factors, I have developed an original methodology to accurately monitor clonal composition of HTLV-1-infected cells. Our pilot data is promising and suggests possible applications of this methodology in enabling the molecular-based diagnosis of ATL, as well as predicting ATL development among HTLV-1-infected individuals. For this purpose, a cohort study to evaluate the clonal composition of infected cells is currently in progress. In summary overtime monitoring of clinical samples suggested the importance of our method for generating biologically meaningful information.