NEOPLASIA
www.neoplasia.com

Volume 20 Number 9 September 2018 pp. 883–893 883

Mutational Intratumor Heterogeneity is a Complex and Early Event in the Development of Adult T-cell Leukemia/Lymphoma

Amir Farmanbar[*,†,1], Sanaz Firouzi[*,1], Wojciech Makałowski[‡], Robert Kneller[§], Masako Iwanaga[¶], Atae Utsunomiya[#], Kenta Nakai[*,†] and Toshiki Watanabe[*,]**

[*]Department of Computational Biology and Medical Sciences, Graduate School of Frontier Sciences, The University of Tokyo, Tokyo, Japan; [†]Human Genome Center, The Institute of Medical Science, The University of Tokyo, Tokyo, Japan; [‡]Institute of Bioinformatics, Faculty of Medicine, University of Muenster, Germany; [§]Research Center for Advanced Science and Technology, The University of Tokyo, Tokyo, Japan; [¶]Department of Frontier Life Science, Graduate School of Biomedical Sciences, Nagasaki University, Nagasaki, Japan; [#]Department of Hematology, Imamura General Hospital, Kagoshima, Japan; [**]Department of Advanced Medical Innovation St. Marianna University Graduate School of Medicine, Kanagawa, Japan

Abstract

The clonal architecture of tumors plays a vital role in their pathogenesis and invasiveness; however, it is not yet clear how this clonality contributes to different malignancies. In this study we sought to address mutational intratumor heterogeneity (ITH) in adult T-cell leukemia/lymphoma (ATL). ATL is a malignancy with an incompletely understood molecular pathogenesis caused by infection with human T-cell leukemia virus type-1 (HTLV-1). To determine the clonal structure through tumor genetic diversity profiles, we investigated 142 whole-exome sequencing data of tumor and matched normal samples from 71 ATL patients. Based on SciClone analysis, the ATL samples showed a wide spectrum of modes over clonal/subclonal frequencies ranging from one to nine clusters. The average number of clusters was six across samples, but the number of clusters differed among different samples. Of these ATL samples, 94% had more than two clusters. Aggressive ATL cases had slightly more clonal clusters than indolent types, indicating the presence of ITH during earlier stages of disease. The known significantly mutated genes in ATL were frequently clustered together and possibly coexisted in the same clone. *IRF4*, *CCR4*, *TP53*, and *PLCG1* mutations were almost clustered in subclones with a moderate variant allele frequency (VAF), whereas *HLA-B*, *CARD11*, and *NOTCH1* mutations were clustered in subclones with lower VAFs. Taken together, these results show that ATL displays a high degree of ITH and a complex subclonal structure. Our findings suggest that clonal/subclonal architecture might be a useful measure for prognostic purposes and personalized assessment of the therapeutic response.

Neoplasia (2018) 20, 883–893

Abbreviations: ATL, Adult T-cell Leukemia/Lymphoma; HTLV-1, Human T-cell Leukemia Virus Type-1; PVL, Proviral Load; NGS, Next Generation Sequencing Technology; WES, Whole Exome Sequencing; SNV, Single Nucleotide Variation; CNA, Copy Number Alteration; VAF, Variant Allele Frequency; ITH, Intratumor Heterogeneity.

Address all correspondence to: Prof. Toshiki Watanabe, Department of Advanced Medical Innovation, St. Marianna University Graduate School of Medicine, Miyamae-ku, Kawasaki, Kanagawa Prefecture, 216-8511, Japan, E-mail: tnabe@ims.u-tokyo.ac.jp or Dr. Sanaz, Firouzi 3rd floor, 2nd building, The Institute of Medical Science, Shirokanedai Campus, The, University of Tokyo, 4-6-1 Shirokanedai, Minato-ku, Tokyo 108-8639, Japan. E-mail: firouzisanaz1@gmail.com

Received 14 April 2018; Revised 2 July 2018; Accepted 6 July 2018

1476-5586
https://doi.org/10.1016/j.neo.2018.07.001

Introduction

Intratumor Heterogeneity in Adult T-cell Leukemia/Lymphoma

Diversity of clonal architecture is a common key feature among a broad range of malignancies, and it has been addressed from different perspectives [1–9]. It has been more than 40 years since the concept of clonal evolution in cancer was first proposed [10]; however, many questions about clonality are still unanswered. Recently, the quantitative nature of next-generation sequencing data has allowed for investigating genetic diversity among tumors and elucidating the clonal architecture of cancers with higher resolution [4,11]. Determining the profiles of somatic point mutations and copy number alterations (CNAs) of specific subpopulations in a tumor that exhibits intratumor heterogeneity (ITH) remains as one of the challenging issues in the field [7,12–15].

Clonal heterogeneity within malignancies has been implicated as a driving force of tumor development and progression because a high degree of genetic variability is associated with an increased risk of subclones having a proliferative advantage, thus leading to clonal expansion [16]. The clonal structures of several cancers have been described as widely diverse patterns ranging from simple monoclonal to complex polyclonal structures [17–19]. At the molecular level, clonal genetic diversity seems to be associated with more aggressive disease [20]. However, the association of the levels of ITH with disease outcome and with efficacy of therapeutic intervention depends on the type of malignancy. For example, in chronic lymphocytic leukemia, subclonal mutations are associated with unfavorable outcomes [21]; however, glioblastoma patients with subclonal mutations manifested longer event-free survival than patients with clonal tumors [22]. Thus, achieving a better understanding of the clonal structure of cancer cells is of vital importance for prognostics and targeted therapies [22]. Widely diverse clonal architecture for each patient's tumor indicates a variation in mutational evolution, and thus knowledge about the clonal architecture is crucial for optimizing a patient's treatment [16,17,23,24]. Recently the number, size, and mutational content of clones within a patient's tumor have been explored extensively [5,13,21,23,25–27].

Adult T-cell leukemia/lymphoma (ATL) is an aggressive and complex malignancy that is caused by infection with human T-cell leukemia virus type-1 (HTLV-1) over a long latency period [28–33]. An analysis of the clonality pattern and the number of clones based on the provirus integration sites [34] indicates that the absolute abundance of infected and leukemic clones is a determining factor for ATL development [35,36]. High-throughput longitudinal analysis indicates that infected individuals with small clones and polyclonal patterns remain healthy over time, whereas those with large clones having an oligo- or monoclonal pattern develop ATL [37, 38]. Also, recently a study on multi-organ clonality analysis in Simian T-Lymphotropic Virus type-1 associated leukemia- a simian counterpart of ATL- reported a complex clonality pattern in this disease [39]. However, the molecular mechanism and pathogenesis behind clonal expansion in ATL remain largely unknown, and the clonal composition based on somatic mutations in ATL has not been monitored.

Although most of the studies analyzing genetic abnormalities of ATL have focused on a limited number of candidate genes [40], a recent study comprehensively revealed the genome-wide mutational spectrum of a large number of ATL cases and proposed a list of frequently mutated genes in ATL [41]. To our knowledge, clonality analysis based on tumor mutational diversity has never been investigated in ATL because it requires costly and complex analysis and needs deep multidisciplinary knowledge for data interpretation. ITH in ATL has the potential to be used as a prognostic biomarker as well as a measure for disease pathogenesis and therapeutic response. Thus, the main aim of the current study was to address clonal heterogeneity in ATL based on mutation profiles of cross-sectional whole-exome sequencing (WES) samples.

Methods

ATL Samples

We used 71 samples from different subtypes of ATL (smoldering, $N = 5$; chronic, $N = 22$; acute, $N = 33$; and lymphoma, $N = 11$) and 71 non-tumor DNA from the same patient deposited in the European Genome-phenome Archive (EGA) with accession number EGAD00001001410 [41]. Sequencing libraries were prepared from tumor DNA (peripheral blood mononuclear cells -PBMCs-) and non-tumor DNA (buccal mucosa) from the same patient to identify acquired (somatic) mutations. The information regarding the samples is included in Table S1.

SNV Bioinformatics Pipeline

We analyzed single-nucleotide variants (SNVs) using our bioinformatics pipeline described below and in Figure S1. In brief, raw sequencing data underwent the following steps: alignment, sorting, indexing, PCR duplicate removal, variant calling, report generation, and data visualization. The short sequencing reads were aligned to the human genome sequence hg38 [42] using the Burrows-Wheeler Aligner (BWA) [43]. Sorting, indexing, and removal of PCR duplicates were conducted with Samtools [44]. SNV detection was performed with the Genome Analysis Toolkit (GATK) HaplotypeCaller [45]. Population variants were removed using dbSNP for human sequences (Version 142) [46]. Further analysis and variant classification were conducted by our in-house Perl and Python scripts. Finally, detected SNVs were annotated by ANNOVAR [47]. The processing of tumor/normal pairs was carried out under identical conditions, and variants detected within the normal matched control samples were removed from tumor variants to retrieve only somatic mutations. In this manuscript, we limited our analysis to mutations with a depth of ≥10 supporting tags.

CNA Bioinformatics Pipeline

Somatic CNAs were detected using data from matched tumor–normal pairs. Exome sequencing reads were aligned against the human reference genome (hg38) using BWA [43]. Subsequently, Samtools [44] mpileup files were used as an input for Varscan v2.2.4 [48]. Segmentation was performed using the DNAcopy library from BioConductor [49]. Gistic2 was used for final analysis and visualization [50]. We analyzed CNAs using our bioinformatics pipeline described in Figure S1.

Inference of Genetic ITH Using SciClone

SciClone [51] was used as an approach to infer the clonal composition of each tumor sample. The output consisting of CNAs and SNVs detected within coding regions was used as input for SciClone. Clonal analysis was performed with a > 10% variant allele frequency (VAF) threshold using a variational Bayesian mixture model implemented by SciClone.